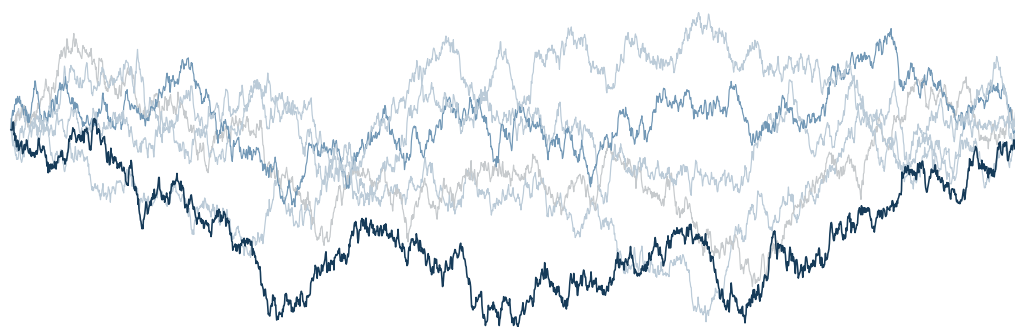


# Limit theorems and applications

## Lecture notes, exercises and solutions



This document brings together the lecture notes and the 91 exercises and problems, together with their solutions, drawn from the tutorials, homework assignments and examinations of the course taught in the M2 in Probability at Paris-Saclay University between 2020 and 2025.

[Part I](#) develops convergence in distribution, Prokhorov's theorem, path spaces, Donsker's theorem, the Brownian bridge and Poisson random measures.

[Part II](#) contains the exercises and their solutions, arranged in seven thematic chapters. The [overview](#) links each topic in the lecture notes to the corresponding exercises.

## Overview

The eight lecture chapters and the seven exercise sheets correspond as shown in the table below.

| Topic                       | Lecture notes                                                          | Exercises                        |
|-----------------------------|------------------------------------------------------------------------|----------------------------------|
| Weak convergence            | <a href="#">Chapter 1</a> , p. 6                                       | <a href="#">Sheet 1</a> , p. 97  |
| Modes of convergence        | <a href="#">Chapter 2</a> , p. 14                                      | <a href="#">Sheet 2</a> , p. 110 |
| Measures on a metric space  | <a href="#">Chapter 3</a> , p. 18                                      | <a href="#">Sheet 3</a> , p. 124 |
| Continuous functions        | <a href="#">Chapter 4</a> , p. 35                                      | <a href="#">Sheet 4</a> , p. 156 |
| Donsker and Brownian bridge | <a href="#">Chapter 5</a> , p. 55<br><a href="#">Chapter 6</a> , p. 62 | <a href="#">Sheet 5</a> , p. 179 |
| Càdlàg paths (Skorokhod)    | <a href="#">Chapter 7</a> , p. 68                                      | <a href="#">Sheet 6</a> , p. 198 |
| Poisson random measures     | <a href="#">Chapter 8</a> , p. 81                                      | <a href="#">Sheet 7</a> , p. 206 |

# Contents

|                                                                      |           |
|----------------------------------------------------------------------|-----------|
| <b>Overview</b>                                                      | <b>2</b>  |
| <b>Part I Lecture notes</b>                                          | <b>5</b>  |
| Notation                                                             | 5         |
| References                                                           | 5         |
| <b>1 Weak convergence in <math>\mathbb{R}^k</math></b>               | <b>6</b>  |
| 1.1 Definition . . . . .                                             | 6         |
| 1.2 Restricting the class of test functions . . . . .                | 7         |
| 1.3 Distribution functions ( $k = 1$ ) and tightness . . . . .       | 8         |
| 1.4 Characteristic functions and Lévy's theorem . . . . .            | 11        |
| <b>Interlude: separability</b>                                       | <b>13</b> |
| <b>2 Other modes of convergence</b>                                  | <b>14</b> |
| 2.1 Almost sure convergence and convergence in probability . . . . . | 14        |
| 2.2 Uniform integrability and convergence in $L^1$ . . . . .         | 15        |
| <b>3 Probability measures on a metric space</b>                      | <b>18</b> |
| 3.1 Regularity of measures . . . . .                                 | 18        |
| 3.2 The portmanteau theorem . . . . .                                | 18        |
| 3.3 Weak convergence in $\mathbb{R}^N$ . . . . .                     | 21        |
| 3.4 Tightness and Prokhorov's theorem . . . . .                      | 24        |
| 3.5 Topological properties of $\mathcal{M}_1(E)$ . . . . .           | 29        |
| 3.6 Further topology . . . . .                                       | 32        |
| <b>4 The space of continuous functions on a compact set</b>          | <b>35</b> |
| 4.1 The Polish space property . . . . .                              | 35        |
| 4.2 Finite-dimensional marginals . . . . .                           | 36        |
| 4.3 Convergence in distribution in this space . . . . .              | 38        |
| 4.4 Compactness in $C$ . . . . .                                     | 39        |
| 4.5 Tightness . . . . .                                              | 41        |
| 4.6 Kolmogorov's tightness criterion . . . . .                       | 43        |
| 4.7 Continuous functions on $\mathbb{R}_+$ . . . . .                 | 46        |
| <b>Interlude: couplings</b>                                          | <b>47</b> |
| C.1 Skorokhod's representation theorem . . . . .                     | 48        |
| C.2 Total variation distance . . . . .                               | 50        |
| C.3 The Lévy–Prokhorov metric . . . . .                              | 52        |
| C.4 Monotonicity and stochastic order . . . . .                      | 53        |
| <b>5 Donsker's theorem and applications</b>                          | <b>55</b> |
| 5.1 Donsker's theorem . . . . .                                      | 55        |
| 5.2 The arcsine law . . . . .                                        | 58        |
| 5.3 Lévy's theorem . . . . .                                         | 60        |
| <b>6 The Brownian bridge</b>                                         | <b>62</b> |
| 6.1 Definition of the Brownian bridge . . . . .                      | 62        |
| 6.2 The local limit theorem . . . . .                                | 63        |
| 6.3 Convergence to the Brownian bridge . . . . .                     | 66        |

|                |                                                                             |            |
|----------------|-----------------------------------------------------------------------------|------------|
| <b>7</b>       | <b>Discontinuous processes: Skorokhod space</b>                             | <b>68</b>  |
| 7.1            | Basic properties . . . . .                                                  | 69         |
| 7.2            | Skorokhod ( $J_1$ ) topology on $\mathbb{D}$ . . . . .                      | 70         |
| 7.3            | Weak convergence in $\mathbb{D}$ . . . . .                                  | 73         |
| 7.4            | Tightness criteria . . . . .                                                | 75         |
| 7.5            | Optional further material . . . . .                                         | 77         |
| <b>8</b>       | <b>Poisson random measures</b>                                              | <b>81</b>  |
| 8.1            | Review of Poisson distributions . . . . .                                   | 81         |
| 8.2            | Poisson random measures . . . . .                                           | 82         |
| 8.3            | Multiple points . . . . .                                                   | 87         |
| 8.4            | Campbell's formulas (additive and exponential) . . . . .                    | 89         |
| 8.5            | Transformations of Poisson random measures . . . . .                        | 91         |
| 8.6            | Poisson random measures with a time parameter . . . . .                     | 92         |
| 8.7            | Palm's formula . . . . .                                                    | 94         |
| <b>Part II</b> | <b>Exercises</b>                                                            | <b>96</b>  |
| <b>1</b>       | <b>Weak convergence in <math>\mathbb{R}^k</math></b>                        | <b>97</b>  |
| 1.1            | Exercises . . . . .                                                         | 97         |
| 1.2            | Solutions . . . . .                                                         | 100        |
| <b>2</b>       | <b>Modes of convergence (almost sure, in probability, <math>L^1</math>)</b> | <b>110</b> |
| 2.1            | Exercises . . . . .                                                         | 110        |
| 2.2            | Solutions . . . . .                                                         | 113        |
| <b>3</b>       | <b>Probability measures on a metric space</b>                               | <b>124</b> |
| 3.1            | Exercises . . . . .                                                         | 124        |
| 3.2            | Solutions . . . . .                                                         | 133        |
| <b>4</b>       | <b>The space of continuous functions on a compact set</b>                   | <b>156</b> |
| 4.1            | Exercises . . . . .                                                         | 156        |
| 4.2            | Solutions . . . . .                                                         | 162        |
| <b>5</b>       | <b>Donsker's theorem, Brownian bridge</b>                                   | <b>179</b> |
| 5.1            | Exercises . . . . .                                                         | 179        |
| 5.2            | Solutions . . . . .                                                         | 185        |
| <b>6</b>       | <b>The space of càdlàg functions</b>                                        | <b>198</b> |
| 6.1            | Exercises . . . . .                                                         | 198        |
| 6.2            | Solutions . . . . .                                                         | 200        |
| <b>7</b>       | <b>Poisson random measures</b>                                              | <b>206</b> |
| 7.1            | Exercises . . . . .                                                         | 206        |
| 7.2            | Solutions . . . . .                                                         | 208        |

## Part I

# Lecture notes

## Notation

- $(E, d)$ : metric space.
- $\mathcal{B}_E$  or  $\mathcal{B}$ : Borel sigma-algebra (the smallest sigma-algebra containing the open sets).
- $\mathcal{M}_1(E)$ : probability measures on  $E$ .
- If  $f : E \rightarrow \mathbb{R}$  is nonnegative measurable or  $\mu$ -integrable, we write

$$\mu(f) = \langle \mu, f \rangle = \int_E f \, d\mu.$$

- $\mathcal{C}_b(E)$ : bounded continuous functions  $E \rightarrow \mathbb{R}$ .
- $\mathcal{C}_c(E)$ : compactly supported continuous functions  $E \rightarrow \mathbb{R}$ .
- $\mathcal{C}(I, E)$  (also denoted by  $C(I, E)$ ): continuous functions from  $I$  to  $E$ .
- $\mathcal{D}(I, E)$ : càdlàg functions from  $I$  to  $E$ ; the domain is omitted when fixed.
- $X_n \xrightarrow{d} X$  or  $X_n \Rightarrow X$ : convergence in distribution. For measures,  $\mu_n \Rightarrow \mu$  denotes weak convergence.

## References

- P. Billingsley, *Convergence of Probability Measures* (2nd ed.).
- O. Kallenberg, *Foundations of Modern Probability* (2nd ed.).
- A. Klenke, *Probability Theory: A Comprehensive Course*.
- J. Kingman, *Poisson Processes*.
- G. Last, M. Penrose, *Lectures on the Poisson Process*.
- P.-L. Méliot, [Convergence of Measures, Poisson Processes and Lévy Processes](#) (lecture notes; title translated)
- G. Miermont, [Limit Theorems and Poisson Processes](#) (lecture notes; title translated)
- T. Lévy, [Convergence of Measures, Large Deviations, Percolation](#). (lecture notes; title translated)

# 1 Weak convergence in $\mathbb{R}^k$

Related exercises: Sheet 1, p. 97.

## In this chapter

|     |                                                            |    |
|-----|------------------------------------------------------------|----|
| 1.1 | Definition . . . . .                                       | 6  |
| 1.2 | Restricting the class of test functions . . . . .          | 7  |
| 1.3 | Distribution functions ( $k = 1$ ) and tightness . . . . . | 8  |
| 1.4 | Characteristic functions and Lévy's theorem . . . . .      | 11 |

## 1.1 Definition

**Definition 1.1.** Let  $(E, d)$  be a metric space and  $\mu_n, \mu \in \mathcal{M}_1(E)$ . We say that  $(\mu_n)$  converges *weakly* to  $\mu$  if

$$\forall f \in \mathcal{C}_b(E), \quad \mu_n(f) \longrightarrow \mu(f).$$

We then write  $\mu_n \Rightarrow \mu$ .

**Example 1.2.** • If  $x_n \rightarrow x \in \mathbb{R}^k$ , then  $\delta_{x_n} \Rightarrow \delta_x$ .

$$\bullet \mu_n = \frac{1}{n+1} \sum_{i=0}^n \delta_{i/n} \Rightarrow \text{Leb}_{[0,1]}.$$

**Probabilistic formulation.** If  $X_n, X$  are random variables with respective distributions  $\mu_n, \mu$ , we say that  $X_n \xrightarrow{d} X$  if  $\mu_n \Rightarrow \mu$ , that is,

$$\forall f \in \mathcal{C}_b(E), \quad \mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)].$$

**Proposition 1.3.** Let  $(E, d)$  be a metric space and  $\mu_n, \mu, \nu \in \mathcal{M}_1(E)$ . If  $\mu_n \Rightarrow \mu$  and  $\mu_n \Rightarrow \nu$ , then  $\mu = \nu$ .

*Proof.* Then  $\mu(f) = \nu(f)$  for every  $f \in \mathcal{C}_b(E)$ . Let  $F$  be closed and  $K > 0$ . Set

$$g_{F,K}(x) = \max(1 - Kd(x, F), 0).$$

The function equals 1 on  $F$ , then decreases linearly as a function of the distance to  $F$  until it vanishes at distance  $1/K$ . Since  $g_{F,K}$  converges pointwise to  $\mathbb{1}_F$  as  $K \rightarrow \infty$ , dominated convergence gives  $\mu(F) = \nu(F)$ , and hence  $\mu = \nu$  by the monotone class theorem.  $\square$

In the rest of this chapter, we restrict ourselves to  $\mathbb{R}^k$ , equipped with the Euclidean metric.

## 1.2 Restricting the class of test functions

**Theorem 1.4** (A basic portmanteau theorem). Let  $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R}^k)$ . Then  $\mu_n \Rightarrow \mu$  if and only if, for every bounded measurable function  $f : \mathbb{R}^k \rightarrow \mathbb{R}_+$  that is continuous  $\mu$ -almost everywhere (that is,  $\mu(\{x \in \mathbb{R}^k : f \text{ is continuous at } x\}) = 1$ ), we have  $\mu_n(f) \rightarrow \mu(f)$ .

See the portmanteau theorem 3.2 for a proof in a more general setting.

**Proposition 1.5.** Let  $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R}^k)$ . Let  $H$  be a set of bounded measurable functions  $\mathbb{R}^k \rightarrow \mathbb{R}$  such that  $\mathcal{C}_c(\mathbb{R}^k) \subset \overline{H}$ , where the closure is taken in the uniform norm. Then

$$[\forall f \in H, \mu_n(f) \rightarrow \mu(f)] \implies \mu_n \Rightarrow \mu.$$

*Proof. Step 1.* We show that  $\mu_n(f) \rightarrow \mu(f)$  for every  $f \in \mathcal{C}_c(\mathbb{R}^k)$  implies  $\mu_n \Rightarrow \mu$ . The idea is a truncation argument. Let  $\chi_r : \mathbb{R}^k \rightarrow [0, 1]$  be continuous, with  $\chi_r(x) = 1$  for  $\|x\| \leq r$  and  $\chi_r(x) = 0$  for  $\|x\| \geq r + 1$ . Let  $f \in \mathcal{C}_b(\mathbb{R}^k)$ . Then

$$|\mu_n(f) - \mu(f)| \leq |\mu_n(f) - \mu_n(f\chi_r)| + |\mu_n(f\chi_r) - \mu(f\chi_r)| + |\mu(f\chi_r) - \mu(f)|.$$

As  $n \rightarrow \infty$ , the limit superior of the first term is bounded above by  $\|f\|_\infty(1 - \mu(\chi_r))$ , that of the second term is zero, and the third term is bounded above by  $\|f\|_\infty(1 - \mu(\chi_r))$ . Thus

$$\limsup_{n \rightarrow \infty} |\mu_n(f) - \mu(f)| \leq 2\|f\|_\infty(1 - \mu(\chi_r)).$$

Since  $\mu(\chi_r) \rightarrow 1$  as  $r \rightarrow \infty$  by dominated convergence, this completes step 1.

*Step 2.* Let  $f \in \mathcal{C}_c(\mathbb{R}^k)$ . Let  $g \in H$  satisfy  $\|f - g\|_\infty \leq \varepsilon$ . Then

$$|\mu_n(f) - \mu(f)| \leq \underbrace{|\mu_n(f) - \mu_n(g)|}_{\leq \varepsilon} + \underbrace{|\mu_n(g) - \mu(g)|}_{\rightarrow 0 \text{ since } g \in H} + \underbrace{|\mu(g) - \mu(f)|}_{\leq \varepsilon}.$$

Therefore  $\mu_n(f) \rightarrow \mu(f)$ . Step 1 completes the proof.  $\square$

**Application.** Let  $\mu \in \mathcal{M}_1(\mathbb{R})$  and let  $(X_n)_{n \geq 1}$  be i.i.d. random variables with distribution  $\mu$ . Let  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ . Then, almost surely,  $\mu_n \Rightarrow \mu$ . In particular, every probability measure in  $\mathcal{M}_1(\mathbb{R})$  is a weak limit of atomic measures.

*Proof.* Let  $(f_k)_{k \geq 1}$  be a sequence in  $\mathcal{C}_c(\mathbb{R})$  that is dense in  $\mathcal{C}_c(\mathbb{R})$ : for example, take polynomials with rational coefficients multiplied by cutoff functions  $T_{K,r}$ , with  $K, r \in \mathbb{N}^*$ , equal to 1 on  $[-K, K]$ , zero outside  $[-K - 1/r, K + 1/r]$ , and affine on the two intervening intervals. Set  $H = \{f_k : k \geq 1\}$ . Since

$$\mu_n(f_k) = \frac{1}{n} \sum_{i=1}^n f_k(X_i) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \mathbb{E}[f_k(X_1)] = \mu(f_k)$$

by the law of large numbers, almost surely  $\mu_n(f) \rightarrow \mu(f)$  for every  $f \in H$ , so the result

follows from the proposition. □

### 1.3 Distribution functions ( $k = 1$ ) and tightness

**Definition 1.6.** For  $\mu \in \mathcal{M}_1(\mathbb{R})$ , we denote its *distribution function* by  $F_\mu(x) = \mu((-\infty, x])$ , for  $x \in \mathbb{R}$ .

**Remark 1.7.** •  $\lim_{x \rightarrow +\infty} F_\mu = 1$ ,  $\lim_{x \rightarrow -\infty} F_\mu = 0$ , and  $F_\mu$  is nondecreasing and càdlàg (right-continuous with left limits at every point).

- Conversely, if  $F : \mathbb{R} \rightarrow [0, 1]$  is nondecreasing, right-continuous, and satisfies  $\lim_{x \rightarrow -\infty} F = 0$ ,  $\lim_{x \rightarrow +\infty} F = 1$ , then there exists  $\mu \in \mathcal{M}_1(\mathbb{R})$  such that  $F = F_\mu$ . Indeed, we can take  $\mu$  to be the pushforward of Lebesgue measure on  $]0, 1[$  under

$$x \mapsto \inf\{y \in \mathbb{R} : F(y) \geq x\}.$$

This is the Lebesgue–Stieltjes measure associated with  $F$ ; see the exercises.

- $\mu$  has an atom at  $x$  if and only if  $\Delta F_\mu(x) = F_\mu(x) - F_\mu(x-) = \mu(\{x\}) > 0$ .
- $\mu$  has at most countably many atoms, since

$$\forall k \geq 1, \quad \#\{x \in \mathbb{R} : \mu(\{x\}) \geq 1/k\} \leq k.$$

- $\mu$  has no atoms if and only if  $F_\mu$  is continuous.
- If  $\mu$  has a density  $g$  with respect to Lebesgue measure, then  $F_\mu$  is differentiable almost everywhere, with derivative  $g$ , and  $F_\mu(x) = \int_{-\infty}^x g(t) dt$ .

*Caution:*  $F_\mu$  is always differentiable almost everywhere, but is not always the integral of its derivative (in fact, this holds if and only if  $\mu \ll \text{Leb}$ ). Consider the devil's staircase, or Cantor function, whose derivative is zero almost everywhere: it is a continuous nondecreasing function with plateaus at every scale.

**Proposition 1.8.** Let  $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R})$ . Then  $\mu_n \Rightarrow \mu$  if and only if

$$F_{\mu_n}(x) \longrightarrow F_\mu(x) \quad \text{for every } x \in \mathbb{R} \text{ such that } \Delta F_\mu(x) = 0.$$

*Proof.* If  $\mu_n \Rightarrow \mu$  and  $\Delta F_\mu(x) = 0$ , then  $f(z) = \mathbb{1}_{\{z \leq x\}}$  is continuous  $\mu$ -almost everywhere, so the basic portmanteau theorem gives  $\mu_n(f) \rightarrow \mu(f)$ .

Conversely, we then have  $\mu_n(g) \rightarrow \mu(g)$  for every function  $g$  of the form

$$g = \sum_{i=1}^r \alpha_i \mathbb{1}_{]x_i, y_i]},$$

where  $x_i$  and  $y_i$  are not atoms of  $\mu$ . Every function in  $\mathcal{C}_c(\mathbb{R})$  is a uniform limit of such functions. The proposition in subsection 2 completes the proof. □

**Definition 1.9.** A family  $(\mu_i)_{i \in I}$  of measures in  $\mathcal{M}_1(\mathbb{R})$  is said to be *tight* if

$$\forall \varepsilon > 0, \quad \exists A > 0, \quad \sup_{i \in I} \mu_i(\mathbb{R} \setminus [-A, A]) \leq \varepsilon.$$

**Remark 1.10.** A sequence  $(\mu_n)_{n \geq 1}$  in  $\mathcal{M}_1(\mathbb{R})$  is tight if and only if

$$\forall \varepsilon > 0, \quad \exists A > 0, \quad \limsup_{n \rightarrow \infty} \mu_n(\mathbb{R} \setminus [-A, A]) \leq \varepsilon.$$

This follows from the fact that a measure  $\mu \in \mathcal{M}_1(\mathbb{R})$  is tight, since  $\lim_{A \rightarrow \infty} \mu(\mathbb{R} \setminus [-A, A]) = 0$ .

The following result is a compactness result and is a special case of Prokhorov's theorem 3.24, presented in Chapter 3.

**Theorem 1.11.** Let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(\mathbb{R})$ . Then  $(\mu_n)_{n \geq 1}$  is tight if and only if, for every subsequence map  $\varphi$ , there exists a subsequence map  $\psi$  such that  $(\mu_{\varphi \circ \psi(n)})_{n \geq 1}$  converges weakly.

By a subsequence map, we mean a map selecting a subsequence or, more formally, an increasing injection  $\mathbb{N}^* \rightarrow \mathbb{N}^*$ .

*Proof. Reverse implication.* Arguing by contradiction, suppose that there exists  $\varepsilon > 0$  such that, for every  $n \geq 1$ ,

$$\mu_{\varphi(n)}(\mathbb{R} \setminus [-n, n]) > \varepsilon.$$

We can choose  $\varphi$  increasing since, for fixed  $i \geq 1$ ,  $\lim_{n \rightarrow \infty} \mu_i(\mathbb{R} \setminus [-n, n]) = 0$ . By assumption, let  $\psi$  be a subsequence map such that  $(\mu_{\varphi \circ \psi(n)})_{n \geq 1}$  converges weakly to  $\mu \in \mathcal{M}_1(\mathbb{R})$ . Then, for every  $x > 0$  such that  $\mu(\{x\}) = \mu(\{-x\}) = 0$ , we have  $[-x, x] \subset [-\psi(n), \psi(n)]$  for all sufficiently large  $n$ . Hence

$$\begin{aligned} \mu([-x, x]) &= \lim_{n \rightarrow \infty} \mu_{\varphi \circ \psi(n)}([-x, x]) \\ &\leq \liminf_{n \rightarrow \infty} \mu_{\varphi \circ \psi(n)}([-\psi(n), \psi(n)]) \leq 1 - \varepsilon. \end{aligned}$$

Letting  $x \rightarrow \infty$  gives  $\mu(\mathbb{R}) < 1$ , a contradiction.

*Forward implication.* Working with  $\mu_{\varphi(n)}$  if necessary, we may assume  $\varphi = \text{id}$ .

*Step 1: identifying a candidate limit.* Set  $F_n = F_{\mu_n}$ . For every  $q \in \mathbb{Q}$ , the sequence  $(F_n(q))_{n \geq 1}$  is bounded. By a diagonal argument, there exists a subsequence map  $\psi$  such that, for every  $q \in \mathbb{Q}$ ,  $(F_{\psi(n)}(q))_{n \geq 1}$  converges to a limit denoted by  $F_\infty(q)$ . We then set

$$F(x) = \inf\{F_\infty(q) : q > x, q \in \mathbb{Q}\}.$$

One checks that  $F$  is nondecreasing and right-continuous. Let  $\varepsilon > 0$  and let  $A > 0$  satisfy

$$\sup_{n \geq 1} \mu_n(\mathbb{R} \setminus [-A, A]) \leq \varepsilon.$$

Let  $q \in \mathbb{Q}$ ,  $q > A$ . Then  $F_n(q) \geq 1 - \varepsilon$ . Letting  $n \rightarrow \infty$  gives  $F_\infty(q) \geq 1 - \varepsilon$ , hence  $F(A) \geq 1 - \varepsilon$ . Similarly,  $F(x) \leq \varepsilon$  for  $x < -A$ . Thus  $\lim_{-\infty} F = 0$ ,  $\lim_{+\infty} F = 1$ . There therefore exists  $\mu \in \mathcal{M}_1(\mathbb{R})$  such that  $F = F_\mu$ : this is the Lebesgue–Stieltjes measure associated with  $F$ ; see the preceding remark.

*Step 2.* We check that  $\mu_{\psi(n)} \Rightarrow \mu$ . Let  $x \in \mathbb{R}$  satisfy  $\mu(\{x\}) = 0$ . Then

$$F(x) = \sup\{F_\infty(q) : q < x, q \in \mathbb{Q}\} = \inf\{F_\infty(q) : q > x, q \in \mathbb{Q}\}.$$

For the first equality, we can use  $F(x) = \lim_{u \uparrow x} F(u)$ . Thus, if  $q < x < q'$  with  $q, q' \in \mathbb{Q}$ ,

$$\begin{aligned} F_\infty(q) &= \lim_{n \rightarrow \infty} F_{\psi(n)}(q) \leq \liminf_{n \rightarrow \infty} F_{\psi(n)}(x) \\ &\leq \limsup_{n \rightarrow \infty} F_{\psi(n)}(x) \leq \limsup_{n \rightarrow \infty} F_{\psi(n)}(q') = F_\infty(q'). \end{aligned}$$

Letting  $q \uparrow x$ ,  $q' \downarrow x$ , we conclude that  $F_{\psi(n)}(x) \rightarrow F(x)$ .

□

**Remark 1.12.** • It is tempting to reformulate this by saying that  $\{\mu_n : n \geq 1\}$  has compact closure in  $\mathcal{M}_1(\mathbb{R})$  equipped with a metric that metrizes weak convergence. This point will be made precise in Section 3.5.

- The extension to  $\mathbb{R}^k$  is similar (see Proposition 5.21 in Kallenberg).

**Definition 1.13.** A family  $(X_i)_{i \in I}$  of real-valued random variables is said to be *tight* if the family of their distributions is tight.

**Example 1.14.** If  $(X_i)_{i \in I}$  is bounded in  $L^1$ , that is, if there exists  $C > 0$  such that  $\sup_{i \in I} \mathbb{E}[|X_i|] \leq C$ , then  $(X_i)_{i \in I}$  is tight. Indeed,

$$\mathbb{P}(|X_i| \geq A) \leq \frac{\mathbb{E}[|X_i|]}{A} \leq \frac{C}{A}.$$

**Lemma 1.15** (Subsequences of subsequences for measures). Let  $\mu_n, \mu \in \mathcal{M}_1(\mathbb{R})$ . Then  $\mu_n \Rightarrow \mu$  if and only if, for every subsequence map  $\varphi$ , there exists a subsequence map  $\psi$  such that  $\mu_{\varphi \circ \psi(n)} \Rightarrow \mu$ .

*Proof.* The forward implication is immediate. For the converse, arguing by contradiction, let  $\varepsilon > 0$ ,  $f \in \mathcal{C}_b(\mathbb{R})$ , and a subsequence map  $\varphi$  satisfy  $|\mu_{\varphi(n)}(f) - \mu(f)| \geq \varepsilon$  for every  $n$ . By assumption, there exists a subsequence map  $\psi$  such that  $\mu_{\varphi \circ \psi(n)} \Rightarrow \mu$ . Then  $\mu_{\varphi \circ \psi(n)}(f) \rightarrow \mu(f)$ , a contradiction. □

**Corollary 1.16.** Let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(\mathbb{R})$  and  $\mu \in \mathcal{M}_1(\mathbb{R})$ . Then  $\mu_n \Rightarrow \mu$  if and only if

- $(\mu_n)_{n \geq 1}$  is tight;

- if  $\mu_{\varphi(n)} \Rightarrow \nu$  for a subsequence map  $\varphi$  and  $\nu \in \mathcal{M}_1(\mathbb{R})$ , then  $\mu = \nu$ .

*Proof.* The forward implication is clear. For the converse, we argue by contradiction. By the preceding lemma, there exists a subsequence map  $\varphi$  such that, for every subsequence map  $\psi$ ,  $\mu_{\varphi \circ \psi(n)} \not\Rightarrow \mu$ . But  $(\mu_{\varphi(n)})$  is tight, so there exist a subsequence map  $\psi$  and  $\nu \in \mathcal{M}_1(\mathbb{R})$  such that  $\mu_{\varphi \circ \psi(n)} \Rightarrow \nu$ . Thus  $\nu = \mu$  by assumption, a contradiction.  $\square$

In practice, the preceding corollary is *very useful* for proving that  $\mu_n \Rightarrow \mu$ , as follows:

- Show that  $(\mu_n)$  is tight: *tightness*.
- Take a subsequence map  $\varphi$  and  $\nu$  such that  $\mu_{\varphi(n)} \Rightarrow \nu$ . Show that  $\nu = \mu$  using a property that characterizes a measure: *uniqueness of the limit*.

## 1.4 Characteristic functions and Lévy's theorem

**Definition 1.17.** For  $\mu \in \mathcal{M}_1(\mathbb{R}^k)$ , we write

$$\phi_\mu(u) = \int_{\mathbb{R}^k} e^{i\langle u, x \rangle} \mu(dx), \quad u \in \mathbb{R}^k,$$

for the *characteristic function* of  $\mu$ ;  $\langle u, x \rangle$  denotes the inner product.

Clearly,  $\mu_n \Rightarrow \mu$  implies  $\phi_{\mu_n} \rightarrow \phi_\mu$  pointwise. For the converse, we have the following more precise result.

**Theorem 1.18** (Lévy). Let  $\mu_n \in \mathcal{M}_1(\mathbb{R}^k)$  satisfy

- for every  $u \in \mathbb{R}^k$ ,  $\phi_{\mu_n}(u)$  converges to a limit denoted by  $\phi(u)$ ;
- $u \mapsto \phi(u)$  is continuous at 0.

Then there exists  $\mu \in \mathcal{M}_1(\mathbb{R}^k)$  such that  $\phi = \phi_\mu$  and  $\mu_n \Rightarrow \mu$ .

For simplicity, we restrict the proof to  $k = 1$ .

**Lemma 1.19** (Injectivity of the characteristic function). Let  $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$  satisfy  $\phi_\mu = \phi_\nu$ . Then  $\mu = \nu$ .

*Proof idea.* We convolve with a Gaussian. Let

$$g_\varepsilon(x) = \frac{1}{\sqrt{2\pi\varepsilon}} e^{-x^2/(2\varepsilon)}, \quad (\mu * g_\varepsilon)(y) = \int \mu(dx) g_\varepsilon(x - y).$$

This is a  $\mathcal{C}^\infty$  probability density; one checks that  $(\mu * g_\varepsilon)(y) dy \Rightarrow \mu$  as  $\varepsilon \rightarrow 0$ . Now

$$\phi_{g_\varepsilon}(z) = e^{-\varepsilon z^2/2} = g_{1/\varepsilon}(z) \sqrt{\frac{2\pi}{\varepsilon}}, \quad g_\varepsilon(z) = \frac{1}{\sqrt{2\pi\varepsilon}} \phi_{g_{1/\varepsilon}}(z).$$

Thus,

$$\begin{aligned} (\mu * g_\varepsilon)(y) &= \int \mu(dx) \frac{1}{\sqrt{2\pi\varepsilon}} \phi_{g_{1/\varepsilon}}(x-y) \\ &= \frac{1}{\sqrt{2\pi\varepsilon}} \int \mu(dx) \int dz e^{iz(x-y)} g_{1/\varepsilon}(z) \\ &\stackrel{\text{Fubini}}{=} \frac{1}{\sqrt{2\pi\varepsilon}} \int dz g_{1/\varepsilon}(z) e^{-izy} \phi_\mu(z). \end{aligned}$$

Thus  $\mu * g_\varepsilon$  can be expressed in terms of  $\phi_\mu$  and  $\varepsilon$ . Letting  $\varepsilon \rightarrow 0$  completes the proof.  $\square$

**Lemma 1.20** (Relation between tails and characteristic functions). There exists  $c > 0$  such that, for every  $\mu \in \mathcal{M}_1(\mathbb{R})$  and every  $K > 0$ ,

$$\mu(\mathbb{R} \setminus [-K, K]) \leq cK \int_{-1/K}^{1/K} (1 - \operatorname{Re} \phi_\mu(u)) du.$$

*Proof.* We have

$$\begin{aligned} K \int_{-1/K}^{1/K} (1 - \operatorname{Re} \phi_\mu(u)) du &= K \int_{-1/K}^{1/K} \left( 1 - \int_{\mathbb{R}} \cos(ux) \mu(dx) \right) du \\ &\stackrel{\text{Fubini}}{=} \int_{\mathbb{R}} \mu(dx) K \int_{-1/K}^{1/K} (1 - \cos(ux)) du \\ &= 2 \int_{\mathbb{R}} \mu(dx) \left( 1 - \frac{\sin(x/K)}{x/K} \right) \\ &\geq 2 \int_{\mathbb{R} \setminus [-K, K]} \mu(dx) \left( 1 - \frac{\sin(x/K)}{x/K} \right) \\ &\geq 2(1 - \sin(1)) \mu(\mathbb{R} \setminus [-K, K]), \end{aligned}$$

since  $\sin(y)/y \leq \sin(1)$  for  $|y| \geq 1$ .  $\square$

*Proof of Lévy's theorem. Step 1: tightness.* We have

$$\mu_n(\mathbb{R} \setminus [-A, A]) \leq cA \int_{-1/A}^{1/A} (1 - \operatorname{Re} \phi_{\mu_n}(u)) du \rightarrow cA \int_{-1/A}^{1/A} (1 - \operatorname{Re} \phi(u)) du.$$

The last quantity tends to 0 as  $A \rightarrow \infty$ , since  $\phi$  is continuous at 0 and  $\phi(0) = 1$ . It follows that

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} \mu_n(\mathbb{R} \setminus [-A, A]) = 0.$$

Thus  $(\mu_n)_{n \geq 1}$  is tight. In particular, there exist a subsequence map  $\varphi_0$  and  $\mu \in \mathcal{M}_1(\mathbb{R})$  such that  $\mu_{\varphi_0(n)} \Rightarrow \mu$ . Then  $\phi_{\mu_{\varphi_0(n)}} \rightarrow \phi_\mu$  pointwise. Hence  $\phi = \phi_\mu$ , which already shows that  $\phi$  is the characteristic function of a probability measure.

*Step 2: uniqueness of the limit.* Let  $\varphi$  be a subsequence map and  $\nu \in \mathcal{M}_1(\mathbb{R})$  satisfy  $\mu_{\varphi(n)} \Rightarrow \nu$ . Then  $\phi_{\mu_{\varphi(n)}} \rightarrow \phi_\nu$  pointwise. But  $\phi_{\mu_n} \rightarrow \phi = \phi_\mu$  pointwise. Thus  $\phi_\mu = \phi_\nu$ , and  $\mu = \nu$  by injectivity of the characteristic function.  $\square$

We again encounter a two-step argument, which is the guiding principle of this course. There are essentially three ingredients:

- (i) Tightness.
- (ii) A certain property  $\mathcal{P}$  passes to the weak limit (for Lévy's theorem,  $\mu_n \Rightarrow \mu$  implies  $\phi_{\mu_n} \rightarrow \phi_\mu$  pointwise).
- (iii) This property  $\mathcal{P}$  characterizes probability measures.

Points 2 and 3 are often called “identification of the limit” or “uniqueness of the limit”.

For these reasons, in general spaces it is often important to have a “good” characterization of compact sets (for tightness) and a “good” characterization of probability measures (for uniqueness of the limit).

## Interlude: separability

Throughout this course, the notion of *separability* is crucial. A metric space  $E$  (or, more generally, a topological space) is said to be *separable* if it has a countable dense sequence.

For example,  $(\mathcal{C}_c(\mathbb{R}), \|\cdot\|_\infty)$  and every compact metric space are separable (use the finite subcover property), whereas  $(\mathcal{C}_b(\mathbb{R}), \|\cdot\|_\infty)$  is not (exercise).

Separability is a very useful property when considering products of spaces. If  $E, F$  are separable metric spaces,

$$\mathcal{B}(E \times F) = \mathcal{B}(E) \otimes \mathcal{B}(F),$$

which is false in general. In particular, if  $X, Y$  are random variables with values in a separable metric space  $(E, d)$ , then  $d(X, Y)$  is measurable as the composition of

$$\begin{aligned} (\Omega, \mathcal{F}) &\longrightarrow (E \times E, \mathcal{B}(E) \otimes \mathcal{B}(E)), \\ \omega &\longmapsto (X(\omega), Y(\omega)), \end{aligned}$$

and

$$\begin{aligned} (E \times E, \mathcal{B}(E \times E)) &\longrightarrow \mathbb{R}, \\ (x, y) &\longmapsto d(x, y), \end{aligned}$$

the second map being measurable because it is continuous, and the two intermediate sigma-algebras being equal.

Separability is a topological property, in the sense that, if  $E$  and  $F$  are homeomorphic, then  $E$  is separable if and only if  $F$  is. N.B.: completeness is not a topological property.

## 2 Other modes of convergence

Related exercises: Sheet 2, p. 110.

In this chapter

|     |                                                                  |    |
|-----|------------------------------------------------------------------|----|
| 2.1 | Almost sure convergence and convergence in probability . . . . . | 14 |
| 2.2 | Uniform integrability and convergence in $L^1$ . . . . .         | 15 |

### 2.1 Almost sure convergence and convergence in probability

Let  $(X_n)_{n \geq 1}$  and  $X$  be random variables defined on the same probability space, with values in a separable metric space  $(E, d)$ .

**Definition 2.1.** • We say that  $X_n$  converges *almost surely* (a.s.) to  $X$ , and write  $X_n \xrightarrow{\text{a.s.}} X$ , if

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

• We say that  $X_n$  converges *in probability* to  $X$ , and write  $X_n \xrightarrow{\mathbb{P}} X$ , if

$$\forall \varepsilon > 0, \quad \mathbb{P}(d(X_n, X) > \varepsilon) \rightarrow 0.$$

**Remark 2.2.** •  $\mathbb{P}(X_n \rightarrow X)$  involves the joint distribution of  $(X, X_1, X_2, X_3, \dots)$ , whereas  $\mathbb{P}(d(X_n, X) > \varepsilon)$  involves only the joint distribution of  $(X_n, X)$ .

• We will see later that almost sure convergence and convergence in probability imply convergence in distribution, and that convergence in distribution to a constant is equivalent to convergence in probability to that constant.

**Proposition 2.3.** We have  $X_n \xrightarrow{\mathbb{P}} X$  if and only if every subsequence  $(X_{\varphi(n)})$  has a further subsequence  $(X_{\varphi \circ \psi(n)})$  that converges almost surely to  $X$ . In particular,  $X_n \xrightarrow{\text{a.s.}} X$  implies  $X_n \xrightarrow{\mathbb{P}} X$ .

*Proof.* First, let us check that

$$X_n \xrightarrow{\mathbb{P}} X \iff \mathbb{E}[d(X_n, X) \wedge 1] \rightarrow 0,$$

where  $\wedge$  denotes the minimum. For the reverse implication, for  $0 < \varepsilon < 1$  we have

$$\mathbb{P}(d(X_n, X) > \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}[d(X_n, X) \wedge 1] \rightarrow 0.$$

For the forward implication, for  $\varepsilon > 0$  we write

$$\begin{aligned} \mathbb{E}[d(X_n, X) \wedge 1] &= \mathbb{E}[(d(X_n, X) \wedge 1) \mathbb{1}_{\{d(X_n, X) \leq \varepsilon\}}] \\ &\quad + \mathbb{E}[(d(X_n, X) \wedge 1) \mathbb{1}_{\{d(X_n, X) > \varepsilon\}}] \\ &\leq \varepsilon + \mathbb{P}(d(X_n, X) > \varepsilon). \end{aligned}$$

Thus  $\limsup_{n \rightarrow \infty} \mathbb{E}[d(X_n, X) \wedge 1] = 0$ . Let us return to the proof of the proposition.

*Forward implication.* Let  $\varphi$  be a subsequence map. There then exists a subsequence map  $\psi$  such that, for every  $n \geq 1$ ,

$$\mathbb{E}[d(X_{\varphi \circ \psi(n)}, X) \wedge 1] \leq 2^{-n}.$$

We have

$$\mathbb{E} \left[ \sum_{n=1}^{\infty} d(X_{\varphi \circ \psi(n)}, X) \wedge 1 \right] = \sum_{n=1}^{\infty} \mathbb{E}[d(X_{\varphi \circ \psi(n)}, X) \wedge 1] < \infty.$$

Thus, almost surely,  $\sum_{n=1}^{\infty} d(X_{\varphi \circ \psi(n)}, X) \wedge 1 < \infty$ , and hence  $X_{\varphi \circ \psi(n)} \xrightarrow{\text{a.s.}} X$ .

*Reverse implication.* Arguing by contradiction, if  $X_n \not\xrightarrow{\mathbb{P}} X$ , we find  $\varepsilon > 0$  and a subsequence map  $\varphi$  such that, for every  $n$ ,  $\mathbb{E}[d(X_{\varphi(n)}, X) \wedge 1] \geq \varepsilon$ . Let  $\psi$  be a subsequence map such that  $X_{\varphi \circ \psi(n)} \xrightarrow{\text{a.s.}} X$ . By dominated convergence,

$$\mathbb{E}[d(X_{\varphi \circ \psi(n)}, X) \wedge 1] \longrightarrow 0,$$

which is a contradiction. □

## 2.2 Uniform integrability and convergence in $L^1$

In this section,  $(X_n)_{n \geq 1}$  and  $X$  are integrable real-valued random variables defined on the same probability space.

**Definition 2.4.** We say that  $(X_n)$  converges to  $X$  in  $L^1$ , and write  $X_n \xrightarrow{L^1} X$ , if  $\mathbb{E}[|X_n - X|] \rightarrow 0$ .

The dominated convergence theorem allows us to deduce  $L^1$  convergence from almost sure convergence under a domination assumption. Scheffé's lemma (see Exercise Sheet 1) also states that

$$X_n \geq 0, X \geq 0, X_n \xrightarrow{\text{a.s.}} X, \mathbb{E}[X_n] \rightarrow \mathbb{E}[X] \implies X_n \xrightarrow{L^1} X.$$

The concept of uniform integrability links convergence in probability with  $L^1$  convergence.

If  $X \in L^1$ , then  $\mathbb{E}[|X| \mathbb{1}_{\{|X| \geq x\}}] \rightarrow 0$  as  $x \rightarrow \infty$ , by dominated convergence. Uniform integrability extends this to a family of random variables.

**Theorem 2.5.** Let  $(X_i)_{i \in I}$  be integrable real-valued random variables. The following three conditions are equivalent:

- (1)  $\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x\}}] \rightarrow 0$  as  $x \rightarrow \infty$ .
- (2)  $\sup_{i \in I} \mathbb{E}[|X_i|] < \infty$  and, for every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that, for every measurable set  $A$ ,

$$\mathbb{P}(A) \leq \delta \implies \sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_A] \leq \varepsilon.$$

(3) There exists a nondecreasing function  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that

$$\frac{\varphi(x)}{x} \rightarrow +\infty \quad (x \rightarrow \infty), \quad \sup_{i \in I} \mathbb{E}[\varphi(|X_i|)] < \infty.$$

This is the *de la Vallée Poussin criterion*.

When these conditions hold,  $(X_i)_{i \in I}$  is said to be *uniformly integrable*, or UI for short.

Before proving this, let us mention a useful corollary of (2).

**Corollary 2.6.** The sum of two UI families is UI.

*Proof of the theorem.* (1)  $\Rightarrow$  (2). For boundedness in  $L^1$ , for sufficiently large  $x$  we write

$$\mathbb{E}[|X_i|] \leq \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| < x\}}] + \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x\}}] \leq x + \sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x\}}] < \infty.$$

Next, fix  $\varepsilon > 0$ . By assumption, there exists  $x_\varepsilon > 0$  such that  $\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x_\varepsilon\}}] \leq \varepsilon$ . Then, if  $\mathbb{P}(A) \leq \varepsilon/x_\varepsilon$ , for  $i \in I$  we have

$$\begin{aligned} \mathbb{E}[|X_i| \mathbb{1}_A] &= \mathbb{E}[|X_i| \mathbb{1}_A \mathbb{1}_{\{|X_i| < x_\varepsilon\}}] + \mathbb{E}[|X_i| \mathbb{1}_A \mathbb{1}_{\{|X_i| \geq x_\varepsilon\}}] \\ &\leq x_\varepsilon \mathbb{P}(A) + \varepsilon \leq 2\varepsilon. \end{aligned}$$

(2)  $\Rightarrow$  (1). By Markov's inequality, if  $x^{-1} \sup_{i \in I} \mathbb{E}[|X_i|] \leq \delta$ , then, for every  $i \in I$ ,  $\mathbb{P}(|X_i| \geq x) \leq \delta$  and hence  $\mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x\}}] \leq \varepsilon$ .

(3)  $\Rightarrow$  (1). Let  $\varepsilon > 0$  and let  $x_\varepsilon > 0$  satisfy  $x \leq \varepsilon \varphi(x)$  for every  $x \geq x_\varepsilon$ . Then

$$\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x_\varepsilon\}}] \leq \varepsilon \sup_{i \in I} \mathbb{E}[\varphi(|X_i|) \mathbb{1}_{\{|X_i| \geq x_\varepsilon\}}] \leq \varepsilon \sup_{i \in I} \mathbb{E}[\varphi(|X_i|)].$$

(1)  $\Rightarrow$  (3). Let  $(x_n)$  be a strictly increasing sequence tending to  $+\infty$  such that

$$\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x_n\}}] \leq 2^{-n}.$$

We then set  $\varphi(x) = \sum_{n=1}^{\infty} (x - x_n)_+$ , for  $x \geq 0$ , which is nondecreasing. Moreover,

$$\frac{\varphi(x)}{x} = \sum_{n=1}^{\infty} \left(1 - \frac{x_n}{x}\right)_+ \rightarrow +\infty$$

by monotone convergence, and

$$\mathbb{E}[\varphi(|X_i|)] \leq \sum_{n=1}^{\infty} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x_n\}}] \leq 1.$$

□

Criterion (3) is often applied with  $\varphi(x) = x^2$ .

**Remark 2.7.** We have seen (see Exercise Sheet 1) that  $(X_i)_{i \in I}$  is tight if and only if there exists  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $\lim_{+\infty} \varphi = +\infty$  and  $\sup_{i \in I} \mathbb{E}[\varphi(|X_i|)] < \infty$ . Tightness is therefore weaker than uniform integrability, and

$$\boxed{\text{UI} \implies \text{tightness.}}$$

**Proposition 2.8.** Let  $(X_n)_{n \geq 1}$  be integrable real-valued random variables and let  $X$  be a real-valued random variable. Then  $X \in L^1$  and  $X_n \xrightarrow{L^1} X$  if and only if  $X_n \xrightarrow{\mathbb{P}} X$  and  $(X_n)_{n \geq 1}$  is UI.

*Proof. Forward implication.* Suppose that  $X_n \xrightarrow{L^1} X$ . Then  $X_n \xrightarrow{\mathbb{P}} X$  by Markov's inequality. To show that  $(X_n)_{n \geq 1}$  is UI, it suffices to show that  $(X_n - X)_{n \geq 1}$  is UI, since the sum of two UI families is UI by characterization (2). To this end, given  $\varepsilon > 0$ , choose  $n_0$  such that  $n \geq n_0$  implies  $\mathbb{E}[|X_n - X|] \leq \varepsilon$ . For these  $n$ , every truncated expectation is therefore bounded above by  $\varepsilon$ . Next, choose  $x_0$  such that, for  $x \geq x_0$ ,

$$\max_{1 \leq i < n_0} \mathbb{E}[|X_i - X| \mathbb{1}_{\{|X_i - X| \geq x\}}] \leq \varepsilon.$$

Then  $\sup_{i \geq 1} \mathbb{E}[|X_i - X| \mathbb{1}_{\{|X_i - X| \geq x\}}] \leq \varepsilon$ .

*Reverse implication.* First, let us check that  $X \in L^1$ . Let  $\varphi$  be a subsequence map such that  $X_{\varphi(n)} \xrightarrow{\text{a.s.}} X$ . Then, by Fatou's lemma,

$$\mathbb{E}[|X|] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[|X_{\varphi(n)}|] < \infty$$

by UI. Thus, as above,  $(X_n - X)_{n \geq 1}$  is UI. Choosing  $\delta$  as in characterization (2), if  $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \delta$ , we deduce that

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n - X| \mathbb{1}_{\{|X_n - X| \geq \varepsilon\}}] + \mathbb{E}[|X_n - X| \mathbb{1}_{\{|X_n - X| < \varepsilon\}}] \leq 2\varepsilon.$$

□

### 3 Probability measures on a metric space

Related exercises: Sheet 3, p. 124.

#### In this chapter

|     |                                               |    |
|-----|-----------------------------------------------|----|
| 3.1 | Regularity of measures                        | 18 |
| 3.2 | The portmanteau theorem                       | 18 |
| 3.3 | Weak convergence in $\mathbb{R}^{\mathbb{N}}$ | 21 |
| 3.4 | Tightness and Prokhorov's theorem             | 24 |
| 3.5 | Topological properties of $\mathcal{M}_1(E)$  | 29 |
| 3.6 | Further topology                              | 32 |

#### 3.1 Regularity of measures

**Theorem 3.1.** Let  $(E, d)$  be a metric space and  $\mu \in \mathcal{M}_1(E)$ . For every  $A \in \mathcal{B}(E)$ ,

$$\mu(A) = \inf\{\mu(O) : O \text{ open, } A \subset O\} = \sup\{\mu(F) : F \text{ closed, } F \subset A\}. \quad (1)$$

*Proof.* This is a measure-theoretic proof. Write

$$\mathcal{A} = \{A \in \mathcal{B}(E) : A \text{ satisfies (1)}\}.$$

If  $O$  is open, then  $O \in \mathcal{A}$ : the infimum is immediate; for the supremum, let  $F_n = \{x \in E : d(x, O^c) \geq 1/n\}$ , which is closed. Since  $O = \bigcup_{n \geq 1} F_n$ , we have  $\mu(O) = \lim_{n \rightarrow \infty} \mu(F_n)$ .

Let us show that  $\mathcal{A}$  is a sigma-algebra. We have  $\emptyset \in \mathcal{A}$ ; closure under complementation is immediate. Let  $(A_n)_{n \geq 1}$  be a sequence in  $\mathcal{A}$ , and let  $F_n$  be closed and  $O_n$  open such that  $F_n \subset A_n \subset O_n$  and  $\mu(O_n \setminus F_n) \leq \varepsilon/2^n$ . Then  $O = \bigcup_{n \geq 1} O_n$  is an open set containing  $\bigcup_{n \geq 1} A_n$  such that

$$\mu\left(O \setminus \bigcup_{n \geq 1} A_n\right) \leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} \leq \varepsilon.$$

However,  $F' = \bigcup_{n \geq 1} F_n$  need not be closed. To get around this, note that  $\bigcup_{n=1}^N F_n \uparrow F'$  as  $N \rightarrow \infty$ , so there exists  $N \geq 1$  such that  $\mu(F' \setminus \bigcup_{n=1}^N F_n) \leq \varepsilon$ . Set  $F = \bigcup_{n=1}^N F_n$ . We then have

$$\mu(O \setminus F) \leq \mu(O \setminus F') + \mu(F' \setminus F) \leq 2\varepsilon.$$

□

In particular, a probability measure is characterized by its values on open sets (or closed sets), which can also be seen directly from the monotone class lemma.

#### 3.2 The portmanteau theorem

Here,  $(E, d)$  is a metric space.

**Theorem 3.2** (Portmanteau / Alexandrov). Let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(E)$  and  $\mu \in \mathcal{M}_1(E)$ . The following properties are equivalent:

- (1)  $\mu_n \Rightarrow \mu$ .
- (2) For every bounded Lipschitz function  $f : E \rightarrow \mathbb{R}$ ,  $\mu_n(f) \rightarrow \mu(f)$ .
- (3) For every closed set  $F$ ,  $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$ .
- (4) For every open set  $O$ ,  $\liminf_{n \rightarrow \infty} \mu_n(O) \geq \mu(O)$ .
- (5) For every  $A \in \mathcal{B}(E)$  such that  $\mu(\partial A) = 0$ , we have  $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ .
- (6) For every bounded measurable function  $f : E \rightarrow \mathbb{R}$  that is continuous at  $\mu$ -almost every point, we have  $\mu_n(f) \rightarrow \mu(f)$ .

The intuition behind (3) is that mass cannot escape from closed sets.

In probabilistic terms, if  $\mathcal{L}(X_n) = \mu_n$  and  $\mathcal{L}(X) = \mu$ ,

$$\begin{aligned} X_n \xrightarrow{d} X &\iff \mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)] \text{ for every bounded Lipschitz } f : E \rightarrow \mathbb{R} \\ &\iff \limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F) \text{ for every closed set } F, \end{aligned}$$

and so on. Before the proof, let us look at a few applications.

**Proposition 3.3.** Let  $X_n, X$  be random variables with values in  $(E, d)$  such that  $X_n \xrightarrow{d} X$ . Let  $f : E \rightarrow \mathbb{R}$  be measurable and almost surely continuous at  $X$ . Then  $f(X_n) \xrightarrow{d} f(X)$ .

*Proof.* Let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be bounded and continuous. Then  $g \circ f : E \rightarrow \mathbb{R}$  is bounded and measurable, and almost surely continuous at  $X$ . Thus, by (6) of the portmanteau theorem,  $\mathbb{E}[g \circ f(X_n)] \rightarrow \mathbb{E}[g \circ f(X)]$ , which is precisely the definition of  $f(X_n) \xrightarrow{d} f(X)$ .  $\square$

**Application.** Let  $(E, d)$  be a separable metric space, and let  $X_n, Y_n$  be random variables with values in  $E$  such that  $X_n \Rightarrow X$  and  $d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$ . Then  $Y_n \Rightarrow X$ .

*Proof.* We apply the portmanteau theorem. Let  $F$  be closed, and write  $F_\varepsilon = \{x \in E : d(x, F) \leq \varepsilon\}$ , which is closed. Then

$$\mathbb{P}(Y_n \in F) \leq \mathbb{P}(X_n \in F_\varepsilon) + \mathbb{P}(d(X_n, Y_n) \geq \varepsilon).$$

Hence

$$\limsup_{n \rightarrow \infty} \mathbb{P}(Y_n \in F) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F_\varepsilon) \leq \mathbb{P}(X \in F_\varepsilon).$$

Since  $F$  is closed,  $\mathbb{P}(X \in F_\varepsilon) \rightarrow \mathbb{P}(X \in F)$  as  $\varepsilon \rightarrow 0$ , which completes the proof.  $\square$

**Corollary 3.4** (Slutsky's lemma). Let  $(E, d_E)$  and  $(E', d_{E'})$  be separable. If  $X_n \Rightarrow X$  in  $E$  and  $X'_n \xrightarrow{\mathbb{P}} x'$  in  $E'$ , with fixed  $x' \in E'$ , then

$$(X_n, X'_n) \Rightarrow (X, x').$$

*Proof.* Equip  $E \times E'$  with the metric

$$d((a, a'), (b, b')) = d_E(a, b) + d_{E'}(a', b').$$

Thus  $(X_n, x') \Rightarrow (X, x')$ , by composing  $X_n \xrightarrow{d} X$  with the continuous map  $x \mapsto (x, x')$ . Moreover,

$$d((X_n, x'), (X_n, X'_n)) = d_{E'}(X'_n, x') \xrightarrow{\mathbb{P}} 0.$$

The result follows. □

**Application.** In  $\mathbb{R}$ , if  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{\mathbb{P}} 1$ , then  $X_n/Y_n \xrightarrow{d} X$ , defining the quotient arbitrarily when  $Y_n = 0$ . Indeed,  $(X_n, Y_n) \xrightarrow{d} (X, 1)$  by Slutsky's lemma, and then  $g(X_n, Y_n) \xrightarrow{d} g(X, 1)$  with

$$g(x, y) = \begin{cases} x/y & \text{if } y \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

which is almost surely continuous at  $(X, 1)$ .

*Proof of the portmanteau theorem.* (1)  $\Rightarrow$  (2). Immediate.

(2)  $\Rightarrow$  (3). Let  $F$  be closed and  $K > 0$ . Set

$$g_{F,K}(x) = \max(1 - Kd(x, F), 0).$$

This function has a plateau at height 1 on  $F$  and vanishes at distance  $1/K$  from  $F$ . Thus  $g_{F,K}$  is  $K$ -Lipschitz and satisfies  $\mathbb{1}_F \leq g_{F,K} \leq 1$ . Hence  $\mu_n(F) \leq \mu_n(g_{F,K})$ , and

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(g_{F,K}).$$

By dominated convergence,  $\mu(g_{F,K}) \rightarrow \mu(F)$  as  $K \rightarrow \infty$ , giving the result.

(3)  $\Rightarrow$  (4). Immediate by taking complements.

(3) and (4)  $\Rightarrow$  (5). Let  $A \in \mathcal{B}(E)$  satisfy  $\mu(\partial A) = 0$ , so that  $\mu(A^\circ) = \mu(A) = \mu(\overline{A})$ . Then

$$\begin{aligned} \mu(A) = \mu(A^\circ) &\leq \liminf_{n \rightarrow \infty} \mu_n(A^\circ) \leq \limsup_{n \rightarrow \infty} \mu_n(A) \\ &\leq \limsup_{n \rightarrow \infty} \mu_n(\overline{A}) \leq \mu(\overline{A}) = \mu(A). \end{aligned}$$

Also inserting  $\liminf_n \mu_n(A)$  between the two bounds, we see that all the inequalities are equalities, giving (5).

(5)  $\Rightarrow$  (6). Let  $f$  be continuous  $\mu$ -almost everywhere and bounded by  $M$ . Writing  $f = f_+ - f_-$  if necessary, we may assume  $f \geq 0$ . Then, by Fubini's theorem,

$$\begin{aligned}\mu_n(f) &= \int_E f(x) \mu_n(dx) = \int_E \int_0^\infty \mathbb{1}_{\{y \leq f(x)\}} dy \mu_n(dx) \\ &= \int_0^\infty dy \int_E \mathbb{1}_{\{y \leq f(x)\}} \mu_n(dx) = \int_0^\infty dy \mu_n(\{f \geq y\}) = \int_0^M \mu_n(A_y) dy,\end{aligned}$$

where  $A_y = \{f \geq y\}$ .

To apply (5), consider  $\partial A_y$ . Let  $D$  be the set of discontinuity points of  $f$ . Let us check that

$$\partial A_y \subset \{x \in E : f(x) = y\} \cup D.$$

*Example for  $E = \mathbb{R}$ .* In the handwritten figure, the graph of  $f$  is drawn in blue, together with a horizontal line at height  $y$ . The set  $A_y = \{x : f(x) \geq y\}$  is marked in red on the horizontal axis, and the points of  $\partial A_y$  in green. Its boundary points occur where the graph crosses level  $y$  or at discontinuities; an isolated point of  $A_y$  can also be a boundary point.

Indeed, let  $x_n \in A_y$  satisfy  $x_n \rightarrow x \in \partial A_y$ . If  $x \in D$ , there is nothing to prove. If  $x \notin D$ , since  $f(x_n) \geq y$  for every  $n$ , we have  $f(x) \geq y$ . Then either  $f(x) = y$ , or  $f(x) > y$ , in which case  $x \in A_y^\circ$ , which is impossible.

There are at most countably many  $y$  such that  $\mu(\{f = y\}) > 0$ : for an integer  $r \geq 1$ , there are at most  $r$  values of  $y$  such that  $\mu(\{f = y\}) \geq 1/r$ , since  $\mu$  is a probability measure. Thus, for Lebesgue-almost every  $y$ ,  $\mu(\partial A_y) = 0$  and  $\mu_n(A_y) \rightarrow \mu(A_y)$ . By dominated convergence, we conclude that

$$\mu_n(f) \rightarrow \int_0^M \mu(A_y) dy = \mu(f).$$

(6)  $\Rightarrow$  (1). Immediate. □

### 3.3 Weak convergence in $\mathbb{R}^{\mathbb{N}}$

Write  $\mathbb{R}^{\mathbb{N}} = \{(x_1, x_2, \dots) : \forall i \geq 1, x_i \in \mathbb{R}\}$  for the set of real sequences.

#### 3.3.1 Sigma-algebra on $\mathbb{R}^{\mathbb{N}}$

**Definition 3.5.** The *product sigma-algebra* on  $\mathbb{R}^{\mathbb{N}}$ , denoted by  $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}$ , is the smallest sigma-algebra on  $\mathbb{R}^{\mathbb{N}}$  making the maps

$$\begin{aligned}\pi_k : \mathbb{R}^{\mathbb{N}} &\longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})), \\ (x_i)_{i \geq 1} &\longmapsto x_k\end{aligned}$$

measurable for every  $k \geq 1$ . It is sometimes called the *cylinder sigma-algebra*.

**Remark 3.6.** One checks that:

- Sets of the form

$$\{(x_i)_{i \geq 1} : x_{i_1} \in A_1, \dots, x_{i_k} \in A_k\} = \pi_{i_1}^{-1}(A_1) \cap \dots \cap \pi_{i_k}^{-1}(A_k),$$

for  $A_1, \dots, A_k \in \mathcal{B}(\mathbb{R})$  and  $1 \leq i_1 < i_2 < \dots < i_k$ , called *cylinder sets*, form a generating  $\pi$ -system for  $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}$ .

- A function  $(\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}})$ ,  $\omega \mapsto (x_i(\omega))_{i \geq 1}$ , is measurable if and only if, for every  $k \geq 1$ ,  $\omega \mapsto x_k(\omega)$  is measurable. In particular,  $(X_i)_{i \geq 1}$  is an  $\mathbb{R}^{\mathbb{N}}$ -valued random variable if and only if each  $X_i$  is a random variable.

**Definition 3.7.** Let  $\mu \in \mathcal{M}_1(\mathbb{R}^{\mathbb{N}})$ . For  $1 \leq i_1 < \dots < i_k$ , write  $\mu_{i_1, \dots, i_k}$  for the pushforward of  $\mu$  under the map

$$\mathbb{R}^{\mathbb{N}} \longrightarrow \mathbb{R}^k, \quad (x_i)_{i \geq 1} \longmapsto (x_{i_1}, \dots, x_{i_k}).$$

If  $(X_i)_{i \geq 1}$  has distribution  $\mu$ , this is the distribution of  $(X_{i_1}, \dots, X_{i_k})$ . The measures  $\{\mu_{i_1, \dots, i_k} : 1 \leq i_1 < \dots < i_k\}$  are called the *k-dimensional marginals* of  $\mu$ . The family of all *k-dimensional marginals*,  $k \geq 1$ , is called the *family of finite-dimensional marginals* of  $\mu$  (fidi for short).

**Proposition 3.8.** A measure  $\mu \in \mathcal{M}_1(\mathbb{R}^{\mathbb{N}})$  is characterized by its finite-dimensional marginals. In particular, the distribution of a random variable  $(X_i)_{i \geq 1} \in \mathbb{R}^{\mathbb{N}}$  is characterized by its finite-dimensional distributions.

*Proof.* This follows from the fact that cylinder sets form a generating  $\pi$ -system for  $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}$ ; two measures in  $\mathcal{M}_1(\mathbb{R}^{\mathbb{N}})$  that agree on these sets are therefore equal by the monotone class lemma.  $\square$

In probabilistic terms,

$$(X_i)_{i \geq 1} \stackrel{d}{=} (Y_i)_{i \geq 1} \iff \forall k \geq 1, \forall 1 \leq i_1 < \dots < i_k, \quad (X_{i_1}, \dots, X_{i_k}) \stackrel{d}{=} (Y_{i_1}, \dots, Y_{i_k}).$$

### 3.3.2 Metric on $\mathbb{R}^{\mathbb{N}}$

To discuss weak convergence in  $\mathbb{R}^{\mathbb{N}}$ , we define a metric on  $\mathbb{R}^{\mathbb{N}}$ .

**Definition 3.9.** For  $(x_i)_{i \geq 1}, (y_i)_{i \geq 1} \in \mathbb{R}^{\mathbb{N}}$ , set

$$d((x_i)_{i \geq 1}, (y_i)_{i \geq 1}) = \sum_{i=1}^{\infty} \frac{|x_i - y_i|}{1 + |x_i - y_i|} \frac{1}{2^i}.$$

If  $(x_i)_{i \geq 1}$  and  $(y_i)_{i \geq 1}$  agree on their first  $K$  terms (that is,  $x_i = y_i$  for  $i \leq K$ ), then

$$d((x_i)_{i \geq 1}, (y_i)_{i \geq 1}) \leq \frac{1}{2^{K-1}}.$$

In particular, the eventually zero sequences of rational numbers are dense in  $\mathbb{R}^{\mathbb{N}}$ , which is therefore separable.

One also checks (exercise) that, if  $(x_i^n)_{i \geq 1} \in \mathbb{R}^{\mathbb{N}}$  and  $(x_i)_{i \geq 1} \in \mathbb{R}^{\mathbb{N}}$ ,

$$d(x^n, x) \rightarrow 0 \iff \forall i \geq 1, \quad x_i^n \rightarrow x_i.$$

**Remark 3.10** (Optional). The metric  $d$  metrizes the product topology on  $\mathbb{R}^{\mathbb{N}}$ , in the sense that the open sets defined by  $d$  coincide with the smallest topology (that is, collection of open sets) for which all the projections  $\pi_k$  are continuous.

**Lemma 3.11.** The Borel sigma-algebra  $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$  coincides with the product sigma-algebra  $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}$ .

*Proof.* Continuity of the projections  $\pi_k$  gives  $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}} \subset \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ . For the reverse inclusion, fix  $z \in \mathbb{R}^{\mathbb{N}}$ . The function

$$x \mapsto d(x, z) = \lim_{K \rightarrow \infty} \sum_{i=1}^K 2^{-i} \frac{|x_i - z_i|}{1 + |x_i - z_i|}$$

is measurable with respect to the product sigma-algebra, as a limit of measurable functions. Every open ball  $B(z, r)$  therefore belongs to this sigma-algebra. Finally,  $\mathbb{R}^{\mathbb{N}}$  is separable: balls with centers in a countable dense subset and strictly positive rational radii form a countable base of open sets. Every open set is therefore a countable union of such balls, proving  $\mathcal{B}(\mathbb{R}^{\mathbb{N}}) \subset \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}$ .  $\square$

### 3.3.3 Weak convergence in $\mathbb{R}^{\mathbb{N}}$

In  $\mathbb{R}^{\mathbb{N}}$ , convergence in distribution is equivalent to convergence of finite-dimensional marginals.

**Theorem 3.12.** Let  $(X^n)_{n \geq 1}$  and  $X$  be  $\mathbb{R}^{\mathbb{N}}$ -valued random variables. Then

$$X^n \xrightarrow{d} X \iff \forall k \geq 1, \forall 1 \leq i_1 < \dots < i_k, \quad (X_{i_1}^n, \dots, X_{i_k}^n) \xrightarrow{d} (X_{i_1}, \dots, X_{i_k}).$$

*Proof.* The forward implication is clear from the continuity of

$$\pi_{i_1, \dots, i_k} : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^k, \quad (x_i)_{i \geq 1} \mapsto (x_{i_1}, \dots, x_{i_k}).$$

For the converse, let  $f : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$  be bounded and Lipschitz. Let  $\varepsilon > 0$ , and let  $\delta > 0$  be such that  $d(x, y) \leq \delta$  implies  $|f(x) - f(y)| \leq \varepsilon$ , for  $x, y \in \mathbb{R}^{\mathbb{N}}$ . Choose  $K$  such that  $2^{-K} < \delta$ . Writing  $T_K x = (x_1, \dots, x_K, 0, 0, \dots)$ , we have

$$\begin{aligned} & |\mathbb{E}[f(X^n)] - \mathbb{E}[f(X)]| \\ & \leq \underbrace{\mathbb{E}[|f(X^n) - f(T_K X^n)|]}_{\leq \varepsilon} + \underbrace{\mathbb{E}[f(T_K X^n)] - \mathbb{E}[f(T_K X)]}_{\rightarrow 0 \text{ by assumption}} + \underbrace{\mathbb{E}[|f(T_K X) - f(X)|]}_{\leq \varepsilon}. \end{aligned}$$

The portmanteau theorem completes the proof.  $\square$

### 3.4 Tightness and Prokhorov's theorem

In the general metric setting, tightness involves the notion of compactness. We need a few topological results.

**Definition 3.13.** Let  $(E, d)$  be a metric space and  $A \subset E$ . We say that  $A$  is *relatively compact* if  $\bar{A}$  is compact.

**Remark 3.14.**  $A$  is relatively compact if and only if every sequence in  $A$  has a convergent subsequence. In particular,  $\{x_n : n \geq 1\}$  is relatively compact if and only if, for every subsequence map  $\varphi$ , there exists a subsequence map  $\psi$  such that  $(x_{\varphi \circ \psi(n)})_{n \geq 1}$  converges.

**Definition 3.15.** Let  $(E, d)$  be a metric space and  $A \subset E$ . We say that  $A$  is *totally bounded* if, for every  $\varepsilon > 0$ ,  $A$  can be covered by finitely many balls of radius  $\varepsilon$ .

**Lemma 3.16.**  $A$  is totally bounded if and only if  $\bar{A}$  is totally bounded.

*Proof.* The reverse implication is clear. For the forward implication, if  $A \subset \bigcup_{n=1}^N B(x_n, \varepsilon)$ , let us check that  $\bar{A} \subset \bigcup_{n=1}^N B(x_n, 2\varepsilon)$ . Let  $y \in \bar{A}$  and  $y_m \rightarrow y$  with  $y_m \in A$ . Let  $m_0$  satisfy  $d(y, y_{m_0}) < \varepsilon$ , and let  $n_1$  satisfy  $y_{m_0} \in B(x_{n_1}, \varepsilon)$ . Then  $y \in B(x_{n_1}, 2\varepsilon)$  by the triangle inequality.  $\square$

Along with separability, completeness plays an important role in Prokhorov's theorem. Unlike separability, completeness is not a topological property:  $(E, d_1)$  and  $(E, d_2)$  can be homeomorphic, with  $(E, d_1)$  complete and  $(E, d_2)$  incomplete. For example, take  $E = \mathbb{R}$ ,  $d_1(x, y) = |x - y|$  and  $d_2(x, y) = |\arctan(x) - \arctan(y)|$ .

In a complete space, total boundedness characterizes relatively compact subsets.

**Example 3.17.** A compact metric space is totally bounded and complete.

The following result will be very useful; see the exercise for the proof.

**Proposition 3.18.** Let  $(E, d)$  be complete and  $A \subset E$ . Then

$$A \text{ totally bounded} \iff A \text{ relatively compact.}$$

**Definition 3.19.** Let  $(E, d)$  be a metric space.

- A measure  $\mu \in \mathcal{M}_1(E)$  is said to be *tight* if, for every  $\varepsilon > 0$ , there exists a compact set  $K$  such that  $\mu(K) \geq 1 - \varepsilon$ .
- A family  $(\mu_i)_{i \in I}$  of measures is said to be *tight* if

$$\forall \varepsilon > 0, \quad \exists K \text{ compact}, \quad \inf_{i \in I} \mu_i(K) \geq 1 - \varepsilon.$$

A family of random variables is tight if the family of their distributions is tight.

**Proposition 3.20.** Let  $(\mu_i)_{i \in I}$  be a tight family on  $(E, d)$  and let  $f : (E, d) \rightarrow (F, d')$  be continuous. Then  $(f_*\mu_i)_{i \in I}$  is tight, where  $f_*\mu_i$  is the pushforward of  $\mu_i$  under  $f$ .

*Proof.* Let  $\varepsilon > 0$  and a compact set  $K$  satisfy  $\inf_{i \in I} \mu_i(K) > 1 - \varepsilon$ . Since  $f$  is continuous,  $f(K)$  is compact. Moreover,  $f^{-1}(f(K)) \supset K$ , so

$$\inf_{i \in I} f_*\mu_i(f(K)) = \inf_{i \in I} \mu_i(f^{-1}(f(K))) \geq \inf_{i \in I} \mu_i(K) > 1 - \varepsilon.$$

□

**Definition 3.21** (Slightly simplified). We say that  $(E, d)$  is *Polish* if it is separable and complete.

**Remark 3.22.** The “true” definition of a Polish space is a separable topological space that is metrizable by a complete metric (see Part 4). For simplicity, however, in this course Polish means metric, complete, and separable.

**Proposition 3.23.** Let  $(E, d)$  be Polish and  $\mu \in \mathcal{M}_1(E)$ . Then  $\mu$  is tight.

*Proof.* Let  $(x_k)_{k \geq 1}$  be a dense sequence. Let  $\varepsilon > 0$ . For every  $n \geq 1$ , choose  $N_n$  such that

$$\mu \left( \bigcup_{k=1}^{N_n} \bar{B}(x_k, 1/n) \right) \geq 1 - \frac{\varepsilon}{2^n}.$$

Set  $K = \bigcap_{n \geq 1} \bigcup_{k=1}^{N_n} \bar{B}(x_k, 1/n)$ . Then  $K$  is closed and totally bounded, hence compact. Moreover,

$$\mu(K^c) \leq \sum_{n=1}^{\infty} \left( 1 - \mu \left( \bigcup_{k=1}^{N_n} \bar{B}(x_k, 1/n) \right) \right) \leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

□

**Theorem 3.24** (Prokhorov). Let  $(E, d)$  be a metric space and let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(E)$ .

- (i) If  $(\mu_n)_{n \geq 1}$  is tight, then, for every subsequence map  $\varphi$ , there exists a subsequence map  $\psi$  such that  $(\mu_{\varphi \circ \psi(n)})_{n \geq 1}$  converges weakly.
- (ii) If  $(E, d)$  is Polish, the converse holds.

We begin with a lemma.

**Proposition 3.25.** Let  $(E, d_E)$  be a separable metric space. Then  $E$  is homeomorphic to a subset of  $(\mathbb{R}^{\mathbb{N}}, d)$ .

*Proof.* Let  $(z^i)_{i \geq 1}$  be a dense sequence in  $E$ . Set

$$\Phi : E \longrightarrow \mathbb{R}^{\mathbb{N}}, \quad x \longmapsto (d_E(x, z^i))_{i \geq 1}.$$

- $\Phi$  is continuous because convergence in  $\mathbb{R}^{\mathbb{N}}$  is coordinatewise convergence.
- Suppose that  $\Phi(x) = \Phi(y)$ . If  $x \neq y$ , choose  $z^i$  such that  $d_E(x, z^i) < d_E(x, y)/2$ . By the triangle inequality, we cannot have  $d_E(x, z^i) = d_E(y, z^i)$ .
- Suppose that  $\Phi(x^n) \rightarrow \Phi(x)$ . Arguing by contradiction, suppose that  $x^n \not\rightarrow x$ . There then exist  $\varepsilon > 0$  and a subsequence map  $\varphi$  such that  $d_E(x^{\varphi(n)}, x) \geq \varepsilon$  for every  $n$ . Choose  $z^i$  such that  $d_E(z^i, x) < \varepsilon/2$ . Then  $d_E(x^{\varphi(n)}, z^i) \geq \varepsilon/2$ . But  $d_E(x^{\varphi(n)}, z^i) \rightarrow d_E(x, z^i)$ , a contradiction.

Thus  $\Phi$  is a homeomorphism onto its image. □

Let us return to the proof of Prokhorov's theorem.

*Proof of the first assertion.* This is the most useful direction, but also the most difficult. The idea is to handle successively the cases  $\mathbb{R}^d$ ,  $\mathbb{R}^{\mathbb{N}}$ ,  $E \subset \mathbb{R}^{\mathbb{N}}$ , and  $E$ .

- *Case  $E = \mathbb{R}$ .* This was covered in Section 1.3; the case  $\mathbb{R}^d$  is treated similarly (see the exercise sheet or Kallenberg, second edition, Proposition 5.21).
- *Case  $E = \mathbb{R}^{\mathbb{N}}$ .* Write

$$\hat{\Pi}_k : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^k, \quad (x_i)_{i \geq 1} \longmapsto (x_1, \dots, x_k),$$

which is continuous. Since the image of a compact set under a continuous function is compact, for every  $k \geq 1$ ,  $(\hat{\Pi}_k \mu_n)_{n \geq 1}$  is tight in  $\mathbb{R}^k$ . By a diagonal argument, for every subsequence map  $\varphi$ , there exists a subsequence map  $\psi$  such that, writing  $\sigma = \varphi \circ \psi$ , for every  $k \geq 1$ ,  $(\hat{\Pi}_k \mu_{\sigma(n)})_{n \geq 1}$  converges weakly to a measure  $\nu_k \in \mathcal{M}_1(\mathbb{R}^k)$ . Since  $(\nu_k)_{k \geq 1}$  is consistent (that is,  $\hat{\Pi}_k \nu_{k+1} = \nu_k$ ), Kolmogorov's extension theorem gives  $\nu \in \mathcal{M}_1(\mathbb{R}^{\mathbb{N}})$  such that  $\hat{\Pi}_k \nu = \nu_k$  for every  $k \geq 1$ . It follows that  $\mu_{\sigma(n)} \xrightarrow{\text{weak}} \nu$ , since in  $\mathbb{R}^{\mathbb{N}}$  weak convergence is equivalent to weak convergence of finite-dimensional marginals.

- *Case of  $\sigma$ -compact  $E$*  ( $E$  is a countable union of compact sets). Since  $E$  is separable, it is homeomorphic to a subset of  $\mathbb{R}^{\mathbb{N}}$  that is  $\sigma$ -compact (the image of a compact set under a continuous function is compact), and hence measurable in  $\mathbb{R}^{\mathbb{N}}$ .

Since weak convergence and tightness are preserved under continuous maps, we may assume that  $E \subset \mathbb{R}^{\mathbb{N}}$  with  $E$  measurable. One checks that

$$\mathcal{B}(E) = \{A \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) : A \subset E\} \quad (\text{exercise}).$$

Extend  $\mu_n$  to  $\mathbb{R}^{\mathbb{N}}$  by setting  $\hat{\mu}_n(A) = \mu_n(A \cap E)$  for  $A \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ ; that is,  $\hat{\mu}_n$  is the pushforward of  $\mu_n$  under the continuous inclusion map  $I : E \rightarrow \mathbb{R}^{\mathbb{N}}$ ,  $x \mapsto x$ .

- a) Then  $(\hat{\mu}_n)_{n \geq 1}$  is tight as the image of  $(\mu_n)_{n \geq 1}$  under  $I$ . Let  $\varphi$  be a subsequence map. By the preceding step, there exists a subsequence map  $\psi$  such that, writing  $\sigma = \varphi \circ \psi$ ,  $(\hat{\mu}_{\sigma(n)})_{n \geq 1}$  converges weakly to  $\hat{\mu} \in \mathcal{M}_1(\mathbb{R}^N)$ .
- b) Let us check that  $\hat{\mu}(E) = 1$ . Let  $K_i$  be a compact subset of  $E$  such that  $\mu_{\sigma(n)}(K_i) \geq 1 - 1/i$  for every  $n \geq 1$ . Since  $K_i$  is closed in  $\mathbb{R}^N$  (exercise), the portmanteau theorem gives

$$\hat{\mu}(E) \geq \hat{\mu}(K_i) \geq \limsup_{n \rightarrow \infty} \hat{\mu}_{\sigma(n)}(K_i) = \limsup_{n \rightarrow \infty} \mu_{\sigma(n)}(K_i) \geq 1 - \frac{1}{i}.$$

Letting  $i \rightarrow \infty$  gives  $\hat{\mu}(E) = 1$ . We then define  $\mu$  on  $E$  by setting  $\mu(A) = \hat{\mu}(A)$  for  $A \in \mathcal{B}(E)$ .

- c) Let us check that  $\mu_{\sigma(n)} \xrightarrow{\text{weak}} \mu$ . We use the portmanteau theorem. Let  $O$  be an open subset of  $E$ . Then  $O = E \cap \hat{O}$ , where  $\hat{O}$  is open in  $\mathbb{R}^N$  (exercise), and

$$\liminf_{n \rightarrow \infty} \mu_{\sigma(n)}(O) = \liminf_{n \rightarrow \infty} \hat{\mu}_{\sigma(n)}(\hat{O}) \geq \hat{\mu}(\hat{O}) = \hat{\mu}(\hat{O} \cap E) = \mu(O),$$

since  $\hat{\mu}(E) = 1$ .

- *General case.* Let  $K_i$  be compact such that  $\mu_n(K_i) \geq 1 - 1/i$  for every  $n \geq 1$ . Set  $E_0 = \bigcup_{i=1}^{\infty} K_i$ , so that  $\mu_n(E_0) = 1$  for every  $n \geq 1$ . Apply the preceding case to  $\tilde{\mu}_n$ , defined on  $E_0$  by  $\tilde{\mu}_n(A) = \mu_n(A)$ . In particular, since  $\mu_n(E_0) = 1$ ,  $\mu_n(A) = \tilde{\mu}_n(A \cap E_0)$  for  $A \in \mathcal{B}(E)$ . The family  $(\tilde{\mu}_n)_{n \geq 1}$  is indeed tight since, for every  $i$ ,  $n \geq 1$ ,  $\tilde{\mu}_n(K_i) \geq 1 - 1/i$ , with  $K_i$  compact.

Let  $\varphi$  be a subsequence map; there exists a subsequence map  $\psi$  such that, writing  $\sigma = \varphi \circ \psi$ , we have  $\tilde{\mu}_{\sigma(n)} \xrightarrow{\text{weak}} \tilde{\mu} \in \mathcal{M}_1(E_0)$ . Define  $\mu$  by  $\mu(A) = \tilde{\mu}(A \cap E_0)$  for  $A \in \mathcal{B}(E)$ . We conclude that  $\mu_{\sigma(n)} \xrightarrow{\text{weak}} \mu$ , since  $\mu_{\sigma(n)}$  and  $\mu$  are the pushforwards of  $\tilde{\mu}_{\sigma(n)}$  and  $\tilde{\mu}$  under the continuous inclusion map  $I : E_0 \rightarrow E$ ,  $x \mapsto x$ .

□

*Proof of the second assertion.* Suppose that  $(E, d)$  is Polish and that, for every subsequence map  $\varphi$ , there exists a subsequence map  $\psi$  such that  $\mu_{\varphi \circ \psi(n)}$  converges weakly. Fix  $\varepsilon > 0$  and a dense sequence  $(x_i)_{i \geq 1}$ . First, let us check that

$$\forall k \geq 1, \quad \exists N_k \geq 1, \quad \forall n \geq 1, \quad \mu_n \left( \bigcup_{i=1}^{N_k} B(x_i, 1/k) \right) \geq 1 - \frac{\varepsilon}{2^k}.$$

Otherwise, there exists  $k \geq 1$  for which this property fails. For every fixed measure  $\mu_j$ , the mass of  $\bigcup_{i=1}^N B(x_i, 1/k)$  tends to 1 as  $N \rightarrow \infty$ . Set  $\varphi(0) = 0$ . At step  $n$ , choose  $N \geq n$  sufficiently large that all indices  $j \leq \varphi(n-1)$  satisfy the bound with this  $N$ . Since the property fails, some index  $\varphi(n)$  violates this bound at index  $N$ ; it is

necessarily strictly greater than  $\varphi(n-1)$ . By inclusion of the coverings, it also violates the bound at index  $n$ . Thus, for every  $n$ ,

$$\mu_{\varphi(n)} \left( \bigcup_{i=1}^n B(x_i, 1/k) \right) < 1 - \frac{\varepsilon}{2^k}.$$

Let  $\psi$  be a subsequence map such that  $\mu_{\varphi \circ \psi(n)} \xrightarrow{\text{weak}} \mu$ . By the portmanteau theorem, for every  $m \geq 1$ ,

$$\begin{aligned} 1 - \frac{\varepsilon}{2^k} &\geq \liminf_{n \rightarrow \infty} \mu_{\varphi \circ \psi(n)} \left( \bigcup_{i=1}^{\psi(n)} B(x_i, 1/k) \right) \\ &\geq \liminf_{n \rightarrow \infty} \mu_{\varphi \circ \psi(n)} \left( \bigcup_{i=1}^m B(x_i, 1/k) \right) \geq \mu \left( \bigcup_{i=1}^m B(x_i, 1/k) \right). \end{aligned}$$

Thus, letting  $m \rightarrow \infty$ , we obtain  $\mu(E) \leq 1 - \varepsilon/2^k$ , a contradiction. Set

$$K = \bigcap_{k \geq 1} \bigcup_{i=1}^{N_k} \overline{B}(x_i, 1/k).$$

As in the earlier proof that a probability measure on a Polish space is tight, we see that  $K$  is compact and  $\mu_n(K) \geq 1 - \varepsilon$  for every  $n \geq 1$ .  $\square$

**Corollary 3.26.** Let  $(E, d)$  be Polish, let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(E)$ , and let  $\mu \in \mathcal{M}_1(E)$ . Then  $\mu_n \xrightarrow{\text{weak}} \mu$  if and only if

- $(\mu_n)_{n \geq 1}$  is tight;
- if  $\varphi$  is a subsequence map and  $\nu \in \mathcal{M}_1(E)$  satisfies  $\mu_{\varphi(n)} \xrightarrow{\text{weak}} \nu$ , then  $\mu = \nu$ .

The proof is the same as that of Corollary 1.16 for  $E = \mathbb{R}$ .

**Probabilistic reformulation.** Let  $(X_n)_{n \geq 1}$  and  $X$  be random variables with values in a Polish space. Then  $X_n \xrightarrow{d} X$  if and only if

- $(X_n)$  is tight, that is,  $\forall \varepsilon > 0, \exists K_\varepsilon$  compact,  $\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \notin K_\varepsilon) \leq \varepsilon$ ;
- if  $X_{\varphi(n)} \xrightarrow{d} Y$ , then  $X \stackrel{d}{=} Y$ .

**Application.** Let  $(X_n)_{n \geq 1}$  and  $(Y_n)_{n \geq 1}$  be random variables with values in Polish spaces  $(E, d)$  and  $(E', d')$ . Suppose that  $X_n \xrightarrow{d} X$ ,  $Y_n \xrightarrow{d} Y$ , and  $X_n \perp\!\!\!\perp Y_n$  for every  $n$ . Then

$$(X_n, Y_n) \xrightarrow{d} (\tilde{X}, \tilde{Y}), \quad \tilde{X} \stackrel{d}{=} X, \quad \tilde{Y} \stackrel{d}{=} Y, \quad \tilde{X} \perp\!\!\!\perp \tilde{Y}.$$

*Proof. Step 1:*  $((X_n, Y_n))_{n \geq 1}$  is tight. Indeed, since  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$ , for  $\varepsilon > 0$  there exist compact subsets  $K, K'$  of  $E, E'$  such that  $\mathbb{P}(X_n \in K) \geq 1 - \varepsilon$  and  $\mathbb{P}(Y_n \in K') \geq 1 - \varepsilon$ . Then  $\mathbb{P}((X_n, Y_n) \in K \times K') \geq 1 - 2\varepsilon$ , with  $K \times K'$  compact in  $E \times E'$ .

*Step 2: uniqueness of the limit.* If  $(X_{\varphi(n)}, Y_{\varphi(n)}) \xrightarrow{d} (\tilde{X}, \tilde{Y})$ , then, for every  $f \in \mathcal{C}_b(E)$ ,  $g \in \mathcal{C}_b(E')$ ,

$$\mathbb{E}[f(X_{\varphi(n)})g(Y_{\varphi(n)})] = \mathbb{E}[f(X_{\varphi(n)})]\mathbb{E}[g(Y_{\varphi(n)})].$$

Passing to the limit as  $n \rightarrow \infty$ , we obtain

$$\mathbb{E}[f(\tilde{X})g(\tilde{Y})] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

As in the proof of the portmanteau theorem, approximate  $\mathbb{1}_{F \times F'}$ , with  $F, F'$  closed, and conclude by the monotone class theorem.  $\square$

### 3.5 Topological properties of $\mathcal{M}_1(E)$

Let  $(E, d)$  be a metric space.

**Definition 3.27.** For  $\mu, \nu \in \mathcal{M}_1(E)$ , set

$$d_{LP}(\mu, \nu) = \inf\{\varepsilon > 0 : \forall A \in \mathcal{B}_E, \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ and } \nu(A) \leq \mu(A^\varepsilon) + \varepsilon\},$$

where  $A^\varepsilon = \{x \in E : d(x, A) < \varepsilon\}$  is the open  $\varepsilon$ -neighborhood of  $A$ . The metric  $d_{LP}$  is called the *Lévy–Prokhorov metric* on  $\mathcal{M}_1(E)$ .

**Proposition 3.28.**  $d_{LP}$  is a metric.

*Proof.* Symmetry is immediate. The triangle inequality follows from  $(A^\varepsilon)^\eta \subset A^{\varepsilon+\eta}$ . For separation, if  $d_{LP}(\mu, \nu) = 0$ , then, for every closed set  $F$ ,

$$\mu(F) \leq \nu(F^{1/n}) + \frac{1}{n}, \quad \nu(F) \leq \mu(F^{1/n}) + \frac{1}{n}.$$

Since  $F = \bigcap_{n \geq 1} F^{1/n}$ , letting  $n \rightarrow \infty$  gives  $\mu(F) \leq \nu(F)$  and  $\nu(F) \leq \mu(F)$ . Thus  $\mu = \nu$  on closed sets, hence  $\mu = \nu$ .  $\square$

It can sometimes be useful to use an asymmetric definition of  $d_{LP}$ .

**Lemma 3.29.** We have

$$d_{LP}(\mu, \nu) = \inf\{\varepsilon > 0 : \forall A \in \mathcal{B}_E, \mu(A) \leq \nu(A^\varepsilon) + \varepsilon\}.$$

*Proof.* Let  $\varepsilon > 0$  satisfy  $\mu(A) \leq \nu(A^\varepsilon) + \varepsilon$  for every  $A \in \mathcal{B}_E$ . Let  $A \in \mathcal{B}_E$ . One checks that  $((A^\varepsilon)^c)^\varepsilon \subset A^c$ , so that

$$\nu(A) = 1 - \nu(A^c) \leq 1 - \nu(((A^\varepsilon)^c)^\varepsilon) \leq 1 - \mu((A^\varepsilon)^c) + \varepsilon = \mu(A^\varepsilon) + \varepsilon.$$

$\square$

**Lemma 3.30.** If  $E$  is separable, for every probability measure  $\mu \in \mathcal{M}_1(E)$  and every  $\eta > 0$ , there exist finitely many functions  $f_1, \dots, f_m \in \mathcal{C}_b(E)$  such that, for every  $\nu \in \mathcal{M}_1(E)$ ,

$$\max_{1 \leq j \leq m} |\nu(f_j) - \mu(f_j)| < \eta \implies d_{LP}(\mu, \nu) \leq 3\eta.$$

*Proof.* Fix a dense sequence  $(x_i)_{i \geq 1}$  in  $E$ . Choose  $N$  such that

$$\mu \left( \bigcup_{i=1}^N B(x_i, \eta) \right) > 1 - \eta.$$

For nonempty  $I \subset \{1, \dots, N\}$ , set

$$B_I = \bigcup_{i \in I} B(x_i, \eta), \quad f_I(x) = \max \left( 1 - \frac{d(x, B_I)}{\eta}, 0 \right),$$

and set  $B_\emptyset = \emptyset$ ,  $f_\emptyset = 0$ . These  $2^N$  functions are continuous and take values between 0 and 1. Suppose that  $|\nu(f_I) - \mu(f_I)| < \eta$  for every  $I$ . For a Borel set  $A$ , set

$$I_A = \{i \in \{1, \dots, N\} : A \cap B(x_i, \eta) \neq \emptyset\}.$$

Then  $A \setminus B_{I_A}$  is contained in the complement of the covering and  $\{f_{I_A} > 0\} \subset A^{3\eta}$ . Consequently,

$$\mu(A) \leq \mu(B_{I_A}) + \eta \leq \mu(f_{I_A}) + \eta < \nu(f_{I_A}) + 2\eta \leq \nu(A^{3\eta}) + 2\eta.$$

The asymmetric definition of the Lévy–Prokhorov metric, established in the preceding lemma, gives  $d_{LP}(\mu, \nu) \leq 3\eta$ .  $\square$

The Lévy–Prokhorov metric is related to weak convergence.

**Proposition 3.31.** Let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(E)$  and  $\mu \in \mathcal{M}_1(E)$ .

- (1)  $d_{LP}(\mu_n, \mu) \rightarrow 0$  implies  $\mu_n \xrightarrow{\text{weak}} \mu$ .
- (2) If  $E$  is separable,  $\mu_n \xrightarrow{\text{weak}} \mu$  implies  $d_{LP}(\mu_n, \mu) \rightarrow 0$ .

Thus, when  $E$  is separable,  $d_{LP}$  metrizes weak convergence on  $\mathcal{M}_1(E)$ .

*Proof.* (1) Suppose that  $d_{LP}(\mu_n, \mu) \rightarrow 0$ . Let  $F$  be closed and  $\varepsilon > 0$ . Let  $n_0$  be such that  $n \geq n_0$  implies  $\mu_n(F) \leq \mu(F^\varepsilon) + \varepsilon$ . Thus  $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F^\varepsilon) + \varepsilon$ . Letting  $\varepsilon \rightarrow 0$  gives  $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$ . Hence, by the portmanteau theorem,  $\mu_n \xrightarrow{\text{weak}} \mu$ .

(2) Let  $E$  be separable and  $\mu_n \xrightarrow{\text{weak}} \mu$ . Fix  $\eta > 0$  and choose the functions  $f_1, \dots, f_m$  from Lemma 3.30. Since there are finitely many of them,

$$\max_{1 \leq j \leq m} |\mu_n(f_j) - \mu(f_j)| \rightarrow 0.$$

For sufficiently large  $n$ , the lemma therefore gives  $d_{LP}(\mu_n, \mu) \leq 3\eta$ . Since  $\eta > 0$  is arbitrary,  $d_{LP}(\mu_n, \mu) \rightarrow 0$ . □

**Corollary 3.32** (Consequence of Prokhorov's theorem). If  $(E, d)$  is compact, then  $(\mathcal{M}_1(E), d_{LP})$  is compact.

*Proof.* Let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(E)$ . It suffices to show that  $(\mu_n)_{n \geq 1}$  has a subsequential limit. Since  $E$  is compact,  $(\mu_n)_{n \geq 1}$  is tight. There therefore exist a subsequence map  $\varphi$  and  $\mu \in \mathcal{M}_1(E)$  such that  $\mu_{\varphi(n)} \xrightarrow{\text{weak}} \mu$ . Since  $E$  is separable,  $\mu_{\varphi(n)} \rightarrow \mu$  in  $(\mathcal{M}_1(E), d_{LP})$ . □

**Remark 3.33.** When  $(E, d)$  is Polish, Prokhorov's theorem can be reformulated as follows: if  $A \subset \mathcal{M}_1(E)$ , the family  $\{\mu : \mu \in A\}$  is tight if and only if  $A$  is relatively compact in  $(\mathcal{M}_1(E), d_{LP})$ . Under the sole assumption of separability, tightness still implies relative compactness.

**Theorem 3.34.** Let  $(E, d)$  be a metric space.

- (1) If  $E$  is separable, then  $(\mathcal{M}_1(E), d_{LP})$  is separable.
- (2) If  $E$  is complete and separable, then  $(\mathcal{M}_1(E), d_{LP})$  is complete and separable.

*Proof.* (1) Let  $(x_i)_{i \geq 1}$  be a dense sequence in  $E$ . We show that the set of measures of the form  $\nu = \sum_{m \in M} r_m \delta_{x_m}$ , with finite  $M \subset \mathbb{N}^*$ ,  $r_m \in \mathbb{Q}_+$ , and  $\sum_{m \in M} r_m = 1$ , is dense.

Let  $\mu \in \mathcal{M}_1(E)$ . Let  $\varepsilon > 0$  and  $M \subset \mathbb{N}^*$ ,  $\text{Card } M < \infty$ , satisfy  $\mu(\bigcup_{m \in M} B(x_m, \varepsilon)) \geq 1 - \varepsilon$ . Replacing the  $B(x_m, \varepsilon)$  by disjoint subsets of these balls, write

$$\bigcup_{m \in M} B(x_m, \varepsilon) = \bigsqcup_{m \in M} B_m, \quad x_m \in B_m \subset B(x_m, \varepsilon).$$

Set  $\nu = \sum_{m \in M} r_m \delta_{x_m}$ , with the  $(r_m)$  chosen so that  $\sum_{m \in M} |r_m - \mu(B_m)| \leq \varepsilon$ . For  $A \in \mathcal{B}(E)$ , set

$$\tilde{A} = \bigcup_{\substack{m \in M \\ B_m \cap A \neq \emptyset}} B_m.$$

Then  $\mu(A) \leq \mu(\tilde{A}) + \mu(A \setminus \tilde{A}) \leq \mu(\tilde{A}) + \varepsilon$ . But  $|\mu(\tilde{A}) - \nu(\tilde{A})| \leq \varepsilon$  and  $\tilde{A} \subset A^{2\varepsilon}$ , so  $\mu(A) \leq \nu(A^{2\varepsilon}) + 2\varepsilon$ , hence  $d_{LP}(\mu, \nu) \leq 2\varepsilon$ .

(2) Let  $(\mu_n)_{n \geq 1}$  be a Cauchy sequence for  $d_{LP}$ . It suffices to show that  $(\mu_n)_{n \geq 1}$  is tight (it will then converge to a subsequential limit, by the properties of Cauchy sequences). Let  $\varepsilon > 0$  and a subsequence map  $\varphi$  satisfy

$$\forall i \geq \varphi(n), \quad d_{LP}(\mu_i, \mu_{\varphi(n)}) < \frac{\varepsilon}{2^{n+1}}.$$

By tightness of  $\mu_1, \dots, \mu_{\varphi(n)}$ , there exists a compact set  $K_n$  such that  $\mu_i(K_n) \geq 1 - \varepsilon/2^{n+1}$  for  $i = 1, \dots, \varphi(n)$ . Then, for  $i \geq \varphi(n)$ ,

$$\mu_i(K_n^{\varepsilon/2^{n+1}}) \geq \mu_{\varphi(n)}(K_n) - \frac{\varepsilon}{2^{n+1}} \geq 1 - \frac{\varepsilon}{2^n}.$$

This also holds for  $i = 1, 2, \dots, \varphi(n)$ . Setting

$$K = \bigcap_{n \geq 1} \overline{K_n^{\varepsilon/2^{n+1}}},$$

one checks that  $K$  is compact (since it is totally bounded and closed) and has measure  $\geq 1 - \varepsilon$ . This completes the proof.  $\square$

### 3.6 Further topology

A few reminders of topology (see, for example, Frédéric Paulin's lecture notes *Topology, Analysis and Differential Calculus* (title translated), available on his webpage).

- A *topology* on a set  $E$  is a collection  $\mathcal{O}$  of subsets of  $E$ , called *open sets*, such that:

- (i) every finite intersection of elements of  $\mathcal{O}$  belongs to  $\mathcal{O}$ ;
- (ii) every union of elements of  $\mathcal{O}$  belongs to  $\mathcal{O}$ .

A topological space is a set equipped with a topology.

- A *base of open sets* of a topological space  $(E, \mathcal{O})$  is a subset  $\mathcal{B}$  of  $\mathcal{O}$  such that every open subset of  $E$  is a union of elements of  $\mathcal{B}$ , that is,

$$\forall U \in \mathcal{O}, \quad \forall x \in U, \quad \exists V \in \mathcal{B}, \quad x \in V \subset U.$$

One can show that if  $(E, \mathcal{O})$  has a countable base of open sets, then  $E$  is separable. The converse holds in a metric space, but is false in general (for example, take  $E = \mathbb{R}$  equipped with the cofinite topology, whose open sets are  $\emptyset$  and sets of the form  $\mathbb{R} \setminus F$  with finite  $F \subset \mathbb{R}$ ).

- In general, a collection of subsets of a set is not automatically a base of open sets for some topology. However, we have the following criterion: if  $\mathcal{B}$  is a subset of  $\mathcal{P}(E)$  such that  $\bigcup \mathcal{B} = E$  and

$$\forall U, V \in \mathcal{B}, \quad \forall x \in U \cap V, \quad \exists W \in \mathcal{B}, \quad x \in W \subset U \cap V,$$

then the collection  $\mathcal{O}$  of unions of elements of  $\mathcal{B}$  is the topology generated by  $\mathcal{B}$ . In particular,  $\mathcal{B}$  is a base of open sets for the topology it generates (the smallest topology containing  $\mathcal{B}$ ).

- If  $(E, \mathcal{O})$  is a topological space, a sequence  $(x_n)$  of elements of  $E$  is said to converge to  $x \in E$  if, for every  $U \in \mathcal{O}$  such that  $x \in U$ , there exists  $N$  such that  $n \geq N \Rightarrow x_n \in U$ .

*Caution:* two different topologies can have the same convergent sequences (for example, take  $\mathbb{N}$  equipped with the discrete topology, which is metrizable,  $\tau_1$ , and

$$\tau_2 = \left\{ A \subset \mathbb{N} : 1 \notin A \text{ or } \sum_{n \in \mathbb{N} \setminus A} \frac{1}{n} < \infty \right\}.$$

- If  $d$  is a metric on  $E$ , the collection  $\mathcal{O}$  of subsets  $U$  of  $E$  such that  $\forall x \in U, \exists \varepsilon > 0, B(x, \varepsilon) \subset U$ , is a topology on  $E$ , called the *topology induced* by  $d$ . A topological space  $(E, \mathcal{O})$  is said to be *metrizable* if such a metric  $d$  exists.
- The notions of “having a countable base of open sets”, “separability”, “convergence”, and “metrizability” are invariant under homeomorphisms.

It is sometimes more natural to define a topology directly without referring to a metric, which may be unnatural, inconvenient to use, or rarely canonical. Moreover, a topological space can be complete or incomplete under different metrics that metrize it.

This is why the “right” definition of a Polish space is: *a separable topological space that is metrizable by a complete metric*. Let us apply the preceding ideas to a topology on the space of probability measures.

Let  $E$  be a metrizable topological space. In view of the definition of weak convergence, it is natural to consider the smallest topology on  $\mathcal{M}_1(E)$  for which the functions  $\mu \mapsto \mu(f)$  are continuous for  $f \in \mathcal{C}_b(E)$  (that is, the *initial topology* on  $\mathcal{M}_1(E)$  defined by the maps  $\mu \mapsto \mu(f)$  for  $f \in \mathcal{C}_b(E)$ ), called the *weak-star topology* on  $\mathcal{M}_1(E)$  (also abbreviated as *weak-\**).

**Lemma 3.35.** The family of sets

$$B_{(\mu, \varepsilon, f_1, \dots, f_n)} = \{\nu \in \mathcal{M}_1(E) : |\mu(f_i) - \nu(f_i)| < \varepsilon \text{ for every } 1 \leq i \leq n\},$$

for  $\varepsilon > 0, f_1, \dots, f_n \in \mathcal{C}_b(E)$ , and  $\mu \in \mathcal{M}_1(E)$ , forms a base of open sets for the weak-\* topology on  $\mathcal{M}_1(E)$ .

*Proof.* These sets are indeed open in the weak-\* topology. The topology generated by this family makes the functions  $\mu \mapsto \mu(f)$ , for  $f \in \mathcal{C}_b(E)$ , continuous, so it contains the weak-\* topology.  $\square$

**Proposition 3.36.** (1) Let  $\mu_n, \mu \in \mathcal{M}_1(E)$ . Then  $\mu_n \xrightarrow{\text{weak}} \mu$  if and only if  $\mu_n \rightarrow \mu$  in the weak-\* topology.

(2) When  $E$  is separable,  $d_{LP}$  metrizes the weak-\* topology.

*Proof.* (1) Suppose that  $\mu_n \xrightarrow{\text{weak}} \mu$ . Let  $U$  be a weak-\* open set containing  $\mu$ . There then exists an open set of the form  $B_{(\mu, \varepsilon, f_1, \dots, f_k)}$  contained in  $U$ . Since  $\mu_n \xrightarrow{\text{weak}} \mu$ ,  $\max_{1 \leq i \leq k} |\mu_n(f_i) - \mu(f_i)| \rightarrow 0$ , and hence  $\mu_n \in B_{(\mu, \varepsilon, f_1, \dots, f_k)}$  for all sufficiently large  $n$ .

Conversely, suppose that  $\mu_n \rightarrow \mu$  in the weak-\* topology. Then, for every  $f \in \mathcal{C}_b(E)$  and every  $\varepsilon > 0$ ,  $\mu_n \in B_{(\mu, \varepsilon, f)}$  for all sufficiently large  $n$ , that is,  $|\mu_n(f) - \mu(f)| < \varepsilon$ .

(2) Let  $U$  be a weak-\* open set and  $\mu \in U$ . Arguing by contradiction, suppose that, for every  $\varepsilon > 0$ ,  $B_{d_{LP}}(\mu, \varepsilon) \not\subset U$ . There then exists a sequence  $\mu_n \in \mathcal{M}_1(E)$  such that  $\mu_n \rightarrow \mu$  for  $d_{LP}$  and  $\mu_n \notin U$  for every  $n \geq 0$ . But we have seen that  $\mu_n \rightarrow \mu$  for  $d_{LP}$  implies  $\mu_n \xrightarrow{\text{weak}} \mu$ , and by (1) this implies that  $\mu_n \rightarrow \mu$  in the weak-\* topology, a contradiction.

Conversely, fix  $\mu \in \mathcal{M}_1(E)$  and  $\varepsilon > 0$ . Choose  $0 < \eta < \varepsilon/3$  and the functions  $f_1, \dots, f_m$  from Lemma 3.30. The set

$$V = \{\nu \in \mathcal{M}_1(E) : |\nu(f_j) - \mu(f_j)| < \eta \text{ for every } 1 \leq j \leq m\}$$

is a weak-\* neighborhood of  $\mu$ , and the lemma gives

$$V \subset \{\nu : d_{LP}(\mu, \nu) \leq 3\eta\} \subset B_{d_{LP}}(\mu, \varepsilon).$$

Every ball centered at  $\mu$  is therefore a weak-\* neighborhood of  $\mu$ . The two topologies coincide.  $\square$

When  $E$  is a metrizable topological space and  $\mathcal{M}_1(E)$  is equipped with the weak-\* topology, one can show that

$$E \text{ compact} \iff \mathcal{M}_1(E) \text{ compact},$$

$$E \text{ separable} \iff \mathcal{M}_1(E) \text{ separable},$$

$$E \text{ Polish} \iff \mathcal{M}_1(E) \text{ Polish}.$$

## 4 The space of continuous functions on a compact set

Related exercises: Sheet 4, p. 156.

In this chapter

|     |                                           |    |
|-----|-------------------------------------------|----|
| 4.1 | The Polish space property                 | 35 |
| 4.2 | Finite-dimensional marginals              | 36 |
| 4.3 | Convergence in distribution in this space | 38 |
| 4.4 | Compactness in $C$                        | 39 |
| 4.5 | Tightness                                 | 41 |
| 4.6 | Kolmogorov's tightness criterion          | 43 |
| 4.7 | Continuous functions on $\mathbb{R}_+$    | 46 |

**Notation.** If  $K, E$  are metric spaces with  $K$  compact, write  $\mathcal{C}(K, E)$  for the set of continuous functions from  $K$  to  $E$ , equipped with the metric  $d(f, g) = \sup_{x \in K} d_E(f(x), g(x))$ .

### 4.1 The Polish space property

**Proposition 4.1.** Let  $(K, \rho)$  be a compact metric space and  $(E, d_E)$  a Polish space. Then  $(\mathcal{C}(K, E), d)$  is Polish.

*Proof. Completeness.* If  $(f_n)_{n \geq 1}$  is Cauchy in  $\mathcal{C}(K, E)$ , then, for every  $x \in K$ ,  $(f_n(x))_{n \geq 1}$  is Cauchy in  $E$ , hence converges to a limit denoted by  $f(x)$ . One then checks that  $d(f_n, f) \rightarrow 0$  and  $f \in \mathcal{C}(K, E)$ .

*Separability.* For  $m, n \geq 1$ , set

$$C_{m,n} = \{f \in \mathcal{C}(K, E) : \forall x, y \in K, \rho(x, y) < 1/m \Rightarrow d_E(f(x), f(y)) < 1/n\}.$$

Since  $K$  is compact, for every  $m \geq 1$  there exists a finite set  $K_m$  such that  $K = \bigcup_{x \in K_m} B(x, 1/(2m))$ . Write  $K_m = \{x_1^m, \dots, x_{k_m}^m\}$ . Also, by separability of  $E$ , for every  $n \geq 1$  there exists a countable collection  $\mathcal{U}_n = \{U_1^n, U_2^n, \dots\}$  of open sets covering  $E$ , each of diameter  $< 1/n$ . For a  $k_m$ -tuple  $(i_1, \dots, i_{k_m})$  of integers  $\geq 1$ , write

$$F_{(i_1, \dots, i_{k_m})}^{m,n} = \{f \in C_{m,n} : f(x_j^m) \in U_{i_j}^n \text{ for every } 1 \leq j \leq k_m\}.$$

Whenever  $F_{(i_1, \dots, i_{k_m})}^{m,n} \neq \emptyset$ , choose a function  $f_{(i_1, \dots, i_{k_m})}^{m,n} \in F_{(i_1, \dots, i_{k_m})}^{m,n}$ . Let  $D_{m,n}$  be the set of these functions.

Let us check that, for every  $f \in C_{m,n}$ , there exists  $g \in D_{m,n}$  such that  $d_E(f(x_j^m), g(x_j^m)) \leq 1/n$  for every  $1 \leq j \leq k_m$ . Let  $f \in C_{m,n}$ . Observe that

$$C_{m,n} = \bigcup_{i_1, \dots, i_{k_m} \geq 1} F_{(i_1, \dots, i_{k_m})}^{m,n}.$$

Choose  $i_1, \dots, i_{k_m}$  such that  $f(x_j^m) \in U_{i_j}^n$  for  $1 \leq j \leq k_m$ . In particular,  $F_{(i_1, \dots, i_{k_m})}^{m,n} \neq \emptyset$ . We can therefore take  $g = f_{(i_1, \dots, i_{k_m})}^{m,n}$ , which has the required property since  $\text{diam}(U_i^n) < 1/n$  for every  $i$ .

Now let  $D = \bigcup_{m,n} D_{m,n}$ , which is countable. Let us check that  $D$  is dense in  $\mathcal{C}(K, E)$ . Let  $f \in \mathcal{C}(K, E)$  and  $\varepsilon > 0$ . Fix  $n > 3/\varepsilon$ . Since  $f$  is uniformly continuous, there exists  $m \geq 1$  such that  $f \in C_{m,n}$ . Let  $g \in D_{m,n}$  satisfy  $d_E(f(x_j^m), g(x_j^m)) < 1/n$  for every  $1 \leq j \leq k_m$ .

By the triangle inequality, for every  $x \in K$ , with  $x_j^m$  chosen so that  $x \in B(x_j^m, 1/(2m))$ , we deduce

$$d_E(f(x), g(x)) \leq d_E(f(x), f(x_j^m)) + d_E(f(x_j^m), g(x_j^m)) + d_E(g(x_j^m), g(x)) \leq \frac{3}{n} \leq \varepsilon.$$

□

## 4.2 Finite-dimensional marginals

### 4.2.1 The space of functions

Let  $(E, d)$  be a Polish space and  $I$  an arbitrary set. The set  $E^I$  of functions from  $I$  to  $E$  is naturally equipped with the product sigma-algebra (as for  $\mathbb{R}^N$  in Section 3.3).

**Definition 4.2.** The *product sigma-algebra* on  $E^I$ , denoted by  $\mathcal{B}(E)^{\otimes I}$ , is the smallest sigma-algebra on  $E^I$  making the maps

$$\pi_t : E^I \longrightarrow (E, \mathcal{B}(E)), \quad (x_i)_{i \in I} \longmapsto x_t,$$

measurable for every  $t \in I$ . It is sometimes called the *cylinder sigma-algebra*.

**Remark 4.3.** One checks that:

- sets of the form

$$\{(x_i)_{i \in I} : x_{t_1} \in A_1, \dots, x_{t_k} \in A_k\} = \pi_{t_1}^{-1}(A_1) \cap \dots \cap \pi_{t_k}^{-1}(A_k),$$

for  $A_1, \dots, A_k \in \mathcal{B}(E)$  and  $t_1, \dots, t_k \in I$ , called *cylinder sets*, form a generating  $\pi$ -system for  $\mathcal{B}(E)^{\otimes I}$ ;

- a function  $(\Omega, \mathcal{A}) \rightarrow (E^I, \mathcal{B}(E)^{\otimes I})$ ,  $\omega \mapsto (x_i(\omega))_{i \in I}$ , is measurable if and only if, for every  $i \in I$ ,  $\omega \mapsto x_i(\omega)$  is measurable. In particular,  $(X_i)_{i \in I}$  is an  $E^I$ -valued random variable if and only if  $X_i$  is a random variable for every  $i \in I$ .

**Definition 4.4.** Let  $\mu \in \mathcal{M}_1(E^I)$ . For  $i_1, \dots, i_k \in I$ , write  $\mu_{i_1, \dots, i_k}$  for the pushforward of  $\mu$  onto  $(E^k, \mathcal{B}(E)^{\otimes k})$  under the map

$$E^I \longrightarrow E^k, \quad (x_i)_{i \in I} \longmapsto (x_{i_1}, \dots, x_{i_k}).$$

If  $(X_i)_{i \in I}$  has distribution  $\mu$ , this is the distribution of  $(X_{i_1}, \dots, X_{i_k})$ . The measures  $\{\mu_{i_1, \dots, i_k} : i_1, \dots, i_k \in I\}$  are called the *k-dimensional marginals* of  $\mu$ . The family of all  $k$ -dimensional marginals,  $k \geq 1$ , is called the *family of finite-dimensional marginals* of  $\mu$  (fidi for short).

**Proposition 4.5.** A measure  $\mu \in \mathcal{M}_1(E^I)$  is characterized by its finite-dimensional marginals. In particular, the distribution of a random variable  $(X_i)_{i \in I} \in E^I$  is characterized by its finite-dimensional distributions.

*Proof.* This follows from the fact that cylinder sets form a generating  $\pi$ -system for  $\mathcal{B}(E)^{\otimes I}$ , so two measures in  $\mathcal{M}_1(E^I)$  that agree on these sets are equal by the monotone class lemma.  $\square$

The advantage of this point of view is the existence of random variables.

**Theorem 4.6** (Kolmogorov's extension theorem). Let  $E$  be a Polish space and  $I$  an arbitrary set. For each nonempty finite subset  $J \subset I$ , suppose that a probability measure  $\mu_J$  on  $E^J$  is given. Assume that this family is consistent: for all nonempty finite subsets  $J \subset K \subset I$ , the projection  $\pi_{K,J} : E^K \rightarrow E^J$  satisfies

$$(\pi_{K,J})_* \mu_K = \mu_J.$$

Then there exists a unique probability measure  $\mu$  on  $(E^I, \mathcal{B}(E)^{\otimes I})$  whose projection onto  $E^J$  is  $\mu_J$  for every nonempty finite subset  $J \subset I$ .

This allows us, for example, to define (pre-)Brownian motion as an  $\mathbb{R}^{\mathbb{R}_+}$ -valued random variable.

*Problem:* in general,  $E^I$  is not Polish. This setting is not well suited to weak convergence!

#### 4.2.2 The space of continuous functions

Here, write  $C = \mathcal{C}(K, E)$ , with  $K$  compact and  $(E, d)$  a Polish space, and let  $\mathcal{B}_C$  be the Borel sigma-algebra on  $C$ . Write  $\mathcal{P}_C$  for the trace on  $C$  of the product sigma-algebra  $\mathcal{B}(E)^{\otimes K}$ , that is,

$$\mathcal{P}_C = \{A \cap C : A \in \mathcal{B}(E)^{\otimes K}\}.$$

Equivalently, it is the smallest sigma-algebra on  $C$  making all the projections  $\pi_t : C \rightarrow E$ ,  $f \mapsto f(t)$ , measurable (exercise).

**Proposition 4.7.** The sigma-algebras  $\mathcal{B}_C$  and  $\mathcal{P}_C$  coincide.

*Proof.* Since, for every  $x \in K$ ,  $\pi_x : C \rightarrow E$  is continuous, it is measurable with respect to  $\mathcal{B}_C$ . Hence  $\mathcal{P}_C \subset \mathcal{B}_C$ .

There are two steps to proving  $\mathcal{B}_C \subset \mathcal{P}_C$ .

*Step 1: open balls in  $C$  are measurable with respect to  $\mathcal{P}_C$ .* To see this, fix  $f_0 \in C$  and consider  $F : C \rightarrow \mathbb{R}$ ,  $f \mapsto d(f, f_0)$ . Since  $K$  is compact, there exists a countable dense subset  $D \subset K$ . By continuity,

$$d(f, f_0) = \sup_{x \in D} d(f(x), f_0(x)).$$

Since, for every  $x \in D$ , the map  $(C, \mathcal{P}_C) \rightarrow \mathbb{R}, f \mapsto d(f(x), f_0(x))$ , is measurable (it is the composition  $\phi \circ \pi_x$  with  $\phi : E \rightarrow \mathbb{R}, y \mapsto d(y, f_0(x))$ , which is continuous and hence measurable),  $F : (C, \mathcal{P}_C) \rightarrow \mathbb{R}$  is measurable as the supremum of a countable family of measurable maps. Thus, for every  $r > 0$ ,  $F^{-1}([0, r]) \in \mathcal{P}_C$ , that is,  $\{f \in C : d(f, f_0) < r\} \in \mathcal{P}_C$ .

*Step 2.* By separability of  $C$ , every open subset of  $C$  is a countable union of open balls (exercise). This completes the proof.  $\square$

By the preceding results, the finite-dimensional marginals  $\{\mu_{t_1, \dots, t_k} : t_1, \dots, t_k \in K\}$  of a measure  $\mu \in \mathcal{M}_1(C)$  characterize it.

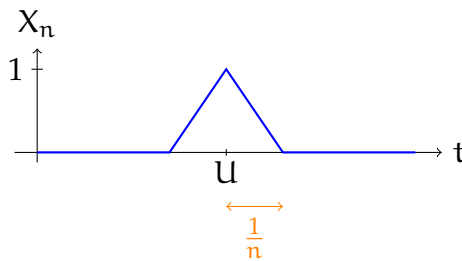
*Caution:* the set  $\{f \in \mathbb{R}^{[0,1]} : f \text{ is continuous}\}$  is not measurable with respect to  $\mathcal{B}(\mathbb{R}^{\otimes [0,1]})$  (exercise).

### 4.3 Convergence in distribution in this space

**Definition 4.8.** Let  $\mu_n, \mu \in \mathcal{M}_1(C)$ . We say that  $\mu_n$  converges to  $\mu$  in *finite-dimensional marginals* if, for every  $t_1, \dots, t_k \in K$ ,  $(\mu_n)_{t_1, \dots, t_k} \xrightarrow{\text{weak}} \mu_{t_1, \dots, t_k}$ . We write  $\mu_n \xrightarrow{\text{f.d.}} \mu$ . Similarly, a sequence  $(X_n)_{n \geq 1}$  of  $C$ -valued random variables converges to  $X$  in finite-dimensional distributions if, for every  $t_1, \dots, t_k \in K$ ,

$$(X_n(t_1), \dots, X_n(t_k)) \xrightarrow{d} (X(t_1), \dots, X(t_k)).$$

*Caution!* By continuity of the projections, weak convergence in  $C$  implies convergence of finite-dimensional marginals. However, unlike in  $\mathbb{R}^{\mathbb{N}}$ , the converse is false: let  $U$  be uniform on  $[0, 1]$  and  $X_n(t) = \max(0, 1 - n|U - t|)$  for  $0 \leq t \leq 1$ .



Then  $X_n \xrightarrow{\text{f.d.}} 0$ , but  $X_n \not\xrightarrow{d} 0$  in  $C$ . Indeed, if this were the case, since  $\sup : C \rightarrow \mathbb{R}$  is continuous, we would have  $1 = \sup X_n \xrightarrow{d} 0$ , a contradiction.

However, we have the following result.

**Theorem 4.9.** Let  $\mu_n, \mu \in \mathcal{M}_1(C)$ .

- (1)  $\mu_n \xrightarrow{\text{weak}} \mu$  if and only if  $\mu_n \xrightarrow{\text{f.d.}} \mu$  and  $(\mu_n)_{n \geq 1}$  is tight.
- (2) If the finite-dimensional marginals of  $(\mu_n)$  converge and  $(\mu_n)_{n \geq 1}$  is tight, then it converges weakly.

*Proof.* (1) The forward implication follows from continuity of the projections and Prokhorov's theorem. Conversely, by Prokhorov's theorem, it suffices to show that if  $\mu_{\varphi(n)} \xrightarrow{\text{weak}} \nu$ , then  $\mu = \nu$ . But then  $\mu_{\varphi(n)} \xrightarrow{\text{f.d.}} \nu$ . Thus  $\mu$  and  $\nu$  have the same finite-dimensional marginals, so they are equal. (2) By tightness, there exist a subsequence map  $\varphi$  and  $\mu \in \mathcal{M}_1(C)$  such that  $\mu_{\varphi(n)} \xrightarrow{\text{weak}} \mu$ . Then  $\mu_{\varphi(n)} \xrightarrow{\text{f.d.}} \mu$ , and hence  $\mu_n \xrightarrow{\text{f.d.}} \mu$ . Thus  $\mu_n \xrightarrow{\text{weak}} \mu$  by (1).  $\square$

Thus, in  $\mathcal{C}([0, 1], E)$ ,  $(X_n(t))_{0 \leq t \leq 1} \xrightarrow{d} (X(t))_{0 \leq t \leq 1}$  if and only if, for every  $0 \leq t_1 < \dots < t_k \leq 1$ , we have

$$(X_n(t_1), \dots, X_n(t_k)) \xrightarrow{d} (X(t_1), \dots, X(t_k)),$$

and  $(X_n)_{n \geq 1}$  is tight.

**Remark 4.10.** Given a consistent family  $\mathcal{F}$  of finite-dimensional marginals, Kolmogorov's extension theorem gives  $\mu \in \mathcal{M}_1(\mathbb{R}^{[0,1]})$  whose finite-dimensional marginals are  $\mathcal{F}$ . But nothing guarantees that such a  $\mu \in \mathcal{M}_1(C)$  exists. By the preceding theorem, this holds if the finite-dimensional marginals  $\mathcal{F}$  are limits of the finite-dimensional marginals of a tight family in  $\mathcal{M}_1(C)$ .

To apply the preceding theorem, we now need to identify the compact subsets of  $C$ .

## 4.4 Compactness in $C$

The notion of equicontinuity is fundamental.

**Definition 4.11.** A family  $\mathcal{F} \subset \mathcal{C}(X, Y)$  is said to be *equicontinuous* if

$$\forall \varepsilon > 0, \quad \exists \eta > 0, \quad \forall x, y \in X, \quad \forall f \in \mathcal{F}, \quad d_X(x, y) \leq \eta \Rightarrow d_Y(f(x), f(y)) \leq \varepsilon.$$

Writing, for  $f \in \mathcal{F}$  and  $\eta > 0$ ,

$$\omega(f, \eta) = \sup\{d_Y(f(x), f(y)) : x, y \in X, d_X(x, y) \leq \eta\},$$

for the *modulus of continuity* of  $f$ , this is equivalent to

$$\forall \varepsilon > 0, \quad \exists \eta > 0, \quad \sup_{f \in \mathcal{F}} \omega(f, \eta) \leq \varepsilon.$$

**Remark 4.12.** For compact  $K$ :

- for  $f \in \mathcal{C}(K, Y)$ ,  $\lim_{h \rightarrow 0} \omega(f, h) = 0$ , since  $f$  is uniformly continuous;
- for every  $h > 0$ , the map  $\mathcal{C}(K, Y) \rightarrow \mathbb{R}_+$ ,  $f \mapsto \omega(f, h)$ , is continuous.

**Theorem 4.13** (Arzelà–Ascoli). Let  $(K, d_K)$  be compact and  $(E, d_E)$  a complete metric space. Let  $\mathcal{F} \subset \mathcal{C}(K, E)$ . Then  $\mathcal{F}$  is relatively compact in  $\mathcal{C}(K, E)$  if and only if

- $\mathcal{F}$  is equicontinuous;
- there exists a countable dense subset  $D \subset K$  such that, for every  $x \in D$ ,  $\mathcal{F}(x) = \{f(x) : f \in \mathcal{F}\}$  is relatively compact in  $E$ .

*Proof.*  $\Leftarrow$  Let  $(f_n)_{n \geq 1}$  be a sequence in  $\mathcal{F}$ . We find a uniformly convergent subsequence. Write  $D = \{a_n : n \geq 1\}$ . Since  $(f_n(a_k))_{n \geq 1}$  lies in  $\overline{\mathcal{F}(a_k)}$ , which is compact in  $E$ , a diagonal argument gives a subsequence map  $\varphi$  such that  $(f_{\varphi(n)}(a_k))_{n \geq 1}$  converges for every  $k \geq 1$ . Set  $g_n(x) = f_{\varphi(n)}(x)$  to simplify the notation.

Let us check that, for every  $x \in K$ ,  $(g_n(x))_{n \geq 1}$  converges. Since  $E$  is complete, it suffices to show that it is Cauchy. To this end, fix  $\varepsilon > 0$  and  $\eta > 0$  such that  $\sup_{f \in \mathcal{F}} \omega(f, \eta) \leq \varepsilon$ . By compactness, there exists  $L \geq 1$  such that  $K = \bigcup_{i=1}^L B(a_i, \eta)$ . For  $1 \leq i \leq L$ , since  $(g_n(a_i))_{n \geq 1}$  converges, it is Cauchy. There therefore exists  $N$  such that

$$p, q \geq N, \quad 1 \leq k \leq L \quad \Rightarrow \quad d_E(g_p(a_k), g_q(a_k)) \leq \varepsilon.$$

Fix  $x \in K$ . Let  $1 \leq k \leq L$  satisfy  $d_K(x, a_k) < \eta$ . Then, for  $p, q \geq N$ ,

$$\begin{aligned} d_E(g_p(x), g_q(x)) &\leq d_E(g_p(x), g_p(a_k)) + d_E(g_p(a_k), g_q(a_k)) \\ &\quad + d_E(g_q(a_k), g_q(x)) \leq 3\varepsilon. \end{aligned}$$

Moreover, writing  $g(x) = \lim_{n \rightarrow \infty} g_n(x)$  and passing to the limit  $q \rightarrow \infty$  in the preceding inequality, we obtain

$$\forall x \in K, \quad \forall p \geq N, \quad d_E(g_p(x), g(x)) \leq 3\varepsilon.$$

Thus convergence is uniform and  $g \in \mathcal{C}(K, E)$ .

$\Rightarrow$  Let  $x \in K$ . Let us show that  $\overline{\mathcal{F}(x)}$  is compact. The map  $\pi_x : \mathcal{C}(K, E) \rightarrow E, f \mapsto f(x)$ , is 1-Lipschitz and hence continuous. Thus  $\pi_x(\overline{\mathcal{F}}) = \overline{\mathcal{F}(x)}$  is compact. But  $\mathcal{F} \subset \overline{\mathcal{F}}$ , so  $\mathcal{F}(x) \subset \overline{\mathcal{F}(x)}$ , hence  $\overline{\mathcal{F}(x)} \subset \overline{\mathcal{F}(x)}$ . Thus  $\overline{\mathcal{F}(x)}$ , being closed in a compact set, is compact.

Let  $\varepsilon > 0$ . By total boundedness of  $\mathcal{F}$ , there exist  $f_1, \dots, f_N \in \mathcal{F}$  such that  $\mathcal{F} \subset \bigcup_{k=1}^N B(f_k, \varepsilon)$ . By uniform continuity, we can find  $\eta > 0$  such that

$$\forall x, y \in K, \quad \forall 1 \leq k \leq N, \quad d_K(x, y) \leq \eta \Rightarrow d_E(f_k(x), f_k(y)) \leq \varepsilon.$$

Then, for every  $f \in \mathcal{F}$  and every  $x, y \in K$  such that  $d_K(x, y) < \eta$ , choose  $1 \leq k \leq N$  such that  $d(f, f_k) < \varepsilon$  in  $\mathcal{C}(K, E)$ , and then

$$d_E(f(x), f(y)) \leq d_E(f(x), f_k(x)) + d_E(f_k(x), f_k(y)) + d_E(f_k(y), f(y)) \leq 3\varepsilon.$$

Thus  $\mathcal{F}$  is equicontinuous. □

**Remark 4.14.** If closed balls in  $E$  are compact (as in  $\mathbb{R}^k$ ) and  $K$  is a compact convex subset of a normed vector space, equipped with the metric induced by the norm, then  $\mathcal{F}$  is relatively compact if and only if it is equicontinuous and  $\mathcal{F}(x_0)$  is relatively compact for some  $x_0 \in K$ .

Indeed, choose  $\eta > 0$  such that  $\sup_{f \in \mathcal{F}} \omega(f, \eta) \leq 1$ . For every  $x \in K$ , divide the segment  $[x_0, x]$  into at most  $1 + \lceil \text{diam}(K)/\eta \rceil$  pieces of length at most  $\eta$ . The triangle inequality gives a uniform bound on  $d_E(f(x), f(x_0))$ , for  $f \in \mathcal{F}$  and  $x \in K$ . The sets  $\mathcal{F}(x)$  are therefore bounded, hence relatively compact in  $E$ , and the Arzelà–Ascoli theorem applies.

A typical application is in  $\mathcal{C}([0, 1], \mathbb{R}^k)$ , taking  $x_0 = 0$ .

## 4.5 Tightness

Let us now look at a probabilistic application.

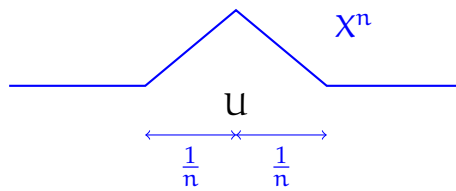
**Theorem 4.15.** Let  $(X^n)_{n \geq 1}$  be a sequence of random variables with values in  $\mathcal{C}([0, 1], E)$ , where  $E$  is Polish. It is tight if and only if

- (1) for every  $\varepsilon > 0, \eta > 0$ , there exists  $\delta > 0$  such that  $\limsup_{n \rightarrow \infty} \mathbb{P}(\omega(X^n, \delta) \geq \eta) \leq \varepsilon$ ;
- (2) for every  $t \in [0, 1] \cap \mathbb{Q}$ ,  $(X^n(t))_{n \geq 1}$  is tight.

If closed bounded subsets of  $E$  are compact, we can replace (2) by (2'):  $(X^n(0))_{n \geq 1}$  is tight.

**Remark 4.16.** If  $(X^n)_{n \geq 1}$  converges in finite-dimensional distributions, condition (2) holds.

**Example 4.17.** If  $U$  is uniform on  $[0, 1]$  and  $X^n(t) = \max(0, 1 - n|U - t|)$  for  $0 \leq t \leq 1$ , the sequence  $(X^n)_{n \geq 1}$  is not tight.



*Proof of the theorem.*  $\Rightarrow$  Let  $\varepsilon > 0$  and let  $K$  be a compact subset of  $C$  such that  $\mathbb{P}(X^n \notin K) \leq \varepsilon$  for every  $n \geq 1$ . Let  $\eta > 0$ .

(1) By Arzelà–Ascoli, we can find  $\delta$  such that  $\sup_{f \in K} \omega(f, \delta) < \eta$ . Then  $\mathbb{P}(\omega(X^n, \delta) \geq \eta) \leq \mathbb{P}(X^n \notin K) \leq \varepsilon$ .

(2) Let  $t \in [0, 1]$ . There exists a compact subset  $A$  of  $E$  such that  $f(t) \in A$  for every  $f \in K$ . Then  $\mathbb{P}(X^n(t) \notin A) \leq \mathbb{P}(X^n \notin K) \leq \varepsilon$ . Thus  $(X^n(t))_{n \geq 1}$  is tight.

$\Leftarrow$  Fix  $\varepsilon > 0$ . Let  $(t_i)_{i \geq 1}$  be an enumeration of the rational numbers in  $[0, 1]$ . We can find compact subsets  $A_i$  of  $E$  such that  $\mathbb{P}(X^n(t_i) \notin A_i) \leq \varepsilon/2^{i+1}$  for every  $n \geq 1$ . For each  $k \geq 1$ , condition (1), applied with  $\eta = 2^{-k}$  and  $\varepsilon/2^{k+2}$ , gives  $\delta'_k > 0$  and  $N_k$  such

that

$$n \geq N_k \implies \mathbb{P}(\omega(X^n, \delta'_k) > 2^{-k}) \leq \frac{\varepsilon}{2^{k+1}}.$$

For each index  $n < N_k$ , continuity of the sample paths implies  $\omega(X^n, \delta) \rightarrow 0$  almost surely as  $\delta \downarrow 0$ . By decreasing  $\delta'_k$ , we therefore obtain  $\delta_k > 0$  such that

$$\forall n \geq 1, \quad \mathbb{P}(\omega(X^n, \delta_k) > 2^{-k}) \leq \frac{\varepsilon}{2^{k+1}}.$$

Setting

$$K_\varepsilon = \{f \in C : \forall i \geq 1, f(t_i) \in A_i; \forall k \geq 1, \omega(f, \delta_k) \leq 2^{-k}\},$$

we see that  $\mathbb{P}(X^n \notin K_\varepsilon) \leq \varepsilon$ , that  $K_\varepsilon$  is closed ( $f \mapsto \omega(f, \delta_k)$  is continuous), and that  $K_\varepsilon$  is compact by Arzelà–Ascoli.

If closed bounded subsets of  $E$  are compact and we assume only (2'), choose a compact set  $A \subset E$  such that  $\sup_n \mathbb{P}(X^n(0) \notin A) \leq \varepsilon/2$ . With the same  $\delta_k$ , the set

$$K'_\varepsilon = \{f \in C : f(0) \in A; \forall k \geq 1, \omega(f, \delta_k) \leq 2^{-k}\}$$

is compact by the remark following Arzelà–Ascoli, and  $\sup_n \mathbb{P}(X^n \notin K'_\varepsilon) \leq \varepsilon$ . This also proves the final assertion.  $\square$

**Remark 4.18.** Condition (1) is equivalent to the following: for every sequence  $\delta_k \rightarrow 0$  and every sequence  $(n_k)$ ,  $\omega(X^{n_k}, \delta_k) \xrightarrow{\mathbb{P}} 0$  (exercise).

**Corollary 4.19.** Let  $(X^n)_{n \geq 1}$  be  $C$ -valued random variables such that

(1) for every  $\varepsilon > 0, \eta > 0$ , there exists  $0 < \delta < 1$  such that, for every  $t \in [0, 1]$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{\delta} \mathbb{P} \left( \sup_{t \leq s \leq (t+\delta) \wedge 1} d(X_s^n, X_t^n) > \eta \right) \leq \varepsilon;$$

(2) for every  $t \in [0, 1] \cap \mathbb{Q}$ ,  $(X_t^n)_{n \geq 1}$  is tight.

Then  $(X^n)_{n \geq 1}$  is tight in  $C$ .

*Proof.* We check condition (1) of the theorem. To this end, divide  $[0, 1]$  into intervals of length  $\delta$ .

$$\begin{array}{c} \text{---} \\ 0 \qquad \qquad \qquad [1/\delta]\delta \qquad 1 \end{array}$$

Let  $f \in C$ . Let us check that

$$\{\omega(f, \delta) > 3\eta\} \subset \bigcup_{i=0}^{[1/\delta]} \left\{ \sup_{i\delta \leq s \leq ((i+1)\delta) \wedge 1} d(f(s), f(i\delta)) > \eta \right\}.$$

Suppose that, for every  $0 \leq i \leq [1/\delta]$ ,  $\sup_{i\delta \leq s \leq ((i+1)\delta) \wedge 1} d(f(s), f(i\delta)) \leq \eta$ . If  $|s - t| \leq \delta$ , there are two cases.

- Case 1:  $s, t$  lie in the same interval  $[i\delta, ((i+1)\delta) \wedge 1]$ . Then

$$d(f(s), f(t)) \leq d(f(t), f(i\delta)) + d(f(s), f(i\delta)) \leq 2\eta.$$

- Case 2:  $s, t$  lie in two consecutive intervals, with  $s \in [i\delta, (i+1)\delta]$  and  $t \in [(i+1)\delta, ((i+2)\delta) \wedge 1]$ . Then

$$\begin{aligned} d(f(s), f(t)) &\leq d(f(s), f(i\delta)) + d(f((i+1)\delta), f(i\delta)) \\ &\quad + d(f(t), f((i+1)\delta)) \leq 3\eta. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{P}(\omega(X^n, \delta) > 3\eta) &\leq (\lfloor 1/\delta \rfloor + 1) \\ &\quad \times \max_{t \in \{0, \delta, 2\delta, \dots, \lfloor 1/\delta \rfloor \delta\}} \mathbb{P} \left( \sup_{t \leq s \leq (t+\delta) \wedge 1} d(X_s^n, X_t^n) > \eta \right). \end{aligned}$$

The limit superior is  $\leq 2\varepsilon$ , since  $\lfloor 1/\delta \rfloor + 1 \leq 1/\delta + 1 < 2/\delta$ . □

## 4.6 Kolmogorov's tightness criterion

The problem with the preceding characterizations of tightness is that they involve random variables that are difficult to study. Effective tightness criteria can involve the distribution of  $X^n$  in the simplest possible way, for example through finite-dimensional marginals. Kolmogorov's criterion involves only marginals of dimension 2 (!).

We first introduce some notation.

**Notation.**

- Let  $D_n = \{i/2^n : 0 \leq i \leq 2^n\}$  be the set of dyadic points of level  $n$  in  $[0, 1]$ , and  $D = \bigcup_{n \geq 0} D_n$ .
- For  $X = (X_t)_{0 \leq t \leq 1} \in E^{[0,1]}$  and  $\gamma > 0$ , write

$$M_\gamma(X) = \sup_{\substack{s, t \in D \\ s \neq t}} \frac{d(X_s, X_t)}{|s - t|^\gamma}.$$

Thus, if  $X$  is  $\gamma$ -Hölder continuous,  $M_\gamma(X) < \infty$ .

**Theorem 4.20** (Kolmogorov's tightness criterion). (1) Let  $X = (X_t)_{0 \leq t \leq 1}$  be an  $E^{[0,1]}$ -valued random variable. Suppose that there exist  $\alpha, \beta, c > 0$  such that

$$\forall s, t \in [0, 1], \quad \mathbb{E}[d(X_s, X_t)^\alpha] \leq c|s - t|^{1+\beta}. \quad (*)$$

Then, for every  $0 < \gamma < \beta/\alpha$ ,  $\mathbb{E}[M_\gamma(X)^\alpha] \leq C(\alpha, \beta, \gamma, c)$ .

- (2) Let  $(X^n)_{n \geq 1}$  be  $C$ -valued random variables. Suppose that

- a) for every  $t \in [0, 1]$ ,  $(X_t^n)_{n \geq 1}$  is tight;  
 b) there exist  $\alpha, \beta, c > 0$  such that, for every  $n \geq 1$  and every  $s, t \in [0, 1]$ ,  
 $\mathbb{E}[d(X_s^n, X_t^n)^\alpha] \leq c|s - t|^{1+\beta}$ .

Then  $(X^n)_{n \geq 1}$  is tight, and every subsequential limit is almost surely  $\gamma$ -Hölder continuous for every  $0 < \gamma < \beta/\alpha$ .

The first part of the proof is deterministic. The idea is to control  $d(X_s, X_t)$  “uniformly” in  $s, t \in D$ , using the triangle inequality along a path of “consecutive” dyadic points.

Write

$$\Delta_k = \max_{0 \leq i < 2^k} d(X_{i2^{-k}}, X_{(i+1)2^{-k}}), \quad K_\gamma = \sup_{k \geq 0} 2^{k\gamma} \Delta_k.$$

**Lemma 4.21.** For every  $\gamma > 0$ ,  $M_\gamma(X) \leq \frac{2^{\gamma+1}}{1-2^{-\gamma}} K_\gamma$ .

*Proof.* We begin by grouping pairs of dyadic points according to their distance:

$$M_\gamma(X) \leq \sup_{m \geq 0} 2^{(m+1)\gamma} \sup_{\substack{s, t \in D \\ |s-t| \leq 2^{-m}}} d(X_s, X_t). \quad (**)$$

Dyadic chaining gives, for every  $m \geq 0$ ,

$$\sup_{\substack{s, t \in D \\ |s-t| \leq 2^{-m}}} d(X_s, X_t) \leq 2 \sum_{k=m}^{\infty} \Delta_k. \quad (***)$$

Indeed, for  $s, t \in D$ ,  $s \leq t$  and  $t - s \leq 2^{-m}$ , set  $s_k = 2^{-k} \lfloor 2^k s \rfloor$  and  $t_k = 2^{-k} \lfloor 2^k t \rfloor$ . The points  $s_m, t_m$  are equal or adjacent in  $D_m$ . Two successive approximations of  $s$  are equal or adjacent in  $D_{k+1}$ , and  $s_k = s$  for all sufficiently large  $k$ . Thus  $d(X_{s_m}, X_s) \leq \sum_{k>m} \Delta_k$ , and similarly for  $t$ . The triangle inequality proves  $(***)$ . If  $K_\gamma = \infty$ , the result is immediate. Otherwise,

$$2 \sum_{k=m}^{\infty} \Delta_k \leq 2K_\gamma \sum_{k=m}^{\infty} 2^{-k\gamma} = \frac{2K_\gamma}{1-2^{-\gamma}} 2^{-m\gamma}.$$

Substituting this bound into  $(**)$  gives the result.  $\square$

*Proof of the theorem.* (1) Assumption  $(*)$  gives

$$\mathbb{E}[\Delta_k^\alpha] \leq \sum_{i=0}^{2^k-1} \mathbb{E}[d(X_{i2^{-k}}, X_{(i+1)2^{-k}})^\alpha] \leq 2^k c 2^{-k(1+\beta)} = c 2^{-k\beta}.$$

For every  $\alpha > 0$ , we deduce

$$\mathbb{E}[K_\gamma^\alpha] \leq \sum_{k \geq 0} 2^{k\alpha\gamma} \mathbb{E}[\Delta_k^\alpha] \leq c \sum_{k \geq 0} 2^{-k(\beta-\alpha\gamma)} < \infty,$$

since  $\gamma < \beta/\alpha$ . The preceding lemma then yields  $\mathbb{E}[M_\gamma(X)^\alpha] \leq C(\alpha, \beta, \gamma, c)$ .

(2) For tightness, it suffices to check that, for every  $\varepsilon > 0, \eta > 0$ , there exists  $\delta > 0$  such that  $\mathbb{P}(\omega(X^n, \delta) \geq \eta) \leq \varepsilon$ .

By definition of  $M_\gamma(X^n)$ , for every  $s, t \in D$  we have  $d(X_s^n, X_t^n) \leq M_\gamma(X^n)|s - t|^\gamma$ . By continuity, this also holds for every  $s, t \in [0, 1]$ . Thus, if  $|s - t| \leq \delta$ , then  $d(X_s^n, X_t^n) \leq M_\gamma(X^n)\delta^\gamma$ , so  $\omega(X^n, \delta) \leq M_\gamma(X^n)\delta^\gamma$ . By Markov's inequality,

$$\mathbb{P}(\omega(X^n, \delta) \geq \eta) \leq \eta^{-\alpha} \delta^{\gamma\alpha} \mathbb{E}[M_\gamma(X^n)^\alpha].$$

The result follows from (1).

Finally, suppose that  $X^{\varphi(n)} \xrightarrow{d} X$  in  $C$ . For  $s, t \in [0, 1]$  and  $R > 0$ , the map  $f \mapsto \min\{d(f(s), f(t))^\alpha, R\}$  is bounded and continuous. By convergence in distribution,

$$\begin{aligned} \mathbb{E}[\min\{d(X_s, X_t)^\alpha, R\}] &= \lim_{n \rightarrow \infty} \mathbb{E}[\min\{d(X_s^{\varphi(n)}, X_t^{\varphi(n)})^\alpha, R\}] \\ &\leq c|s - t|^{1+\beta}. \end{aligned}$$

Letting  $R \uparrow \infty$ , monotone convergence shows that  $X$  satisfies (\*). Assertion (1) and continuity of  $X$  therefore give almost sure  $\gamma$ -Hölder continuity for each  $0 < \gamma < \beta/\alpha$ . Taking a sequence of exponents increasing to  $\beta/\alpha$  and a countable intersection, this regularity holds simultaneously for all these exponents.  $\square$

**Remark 4.22.** Assertion (1) also proves Kolmogorov's continuity criterion: if  $(X_t)_{0 \leq t \leq 1} \in E^{[0,1]}$  satisfies (\*), then, for every  $0 < \gamma < \beta/\alpha$ , there exists a  $\gamma$ -Hölder continuous process  $(\tilde{X}_t)_{0 \leq t \leq 1}$  such that, for every  $t \in [0, 1]$ ,  $\mathbb{P}(X_t = \tilde{X}_t) = 1$ .

Indeed, since  $\mathbb{E}[M_\gamma(X)^\alpha] < \infty$ , there exists an event  $A$  of probability 1 such that, for every  $\omega \in A$ ,  $X$  is  $\gamma$ -Hölder continuous on  $D$ . For  $\omega \in A$ , we can therefore extend  $X$  to a  $\gamma$ -Hölder continuous function, denoted by  $\tilde{X}$ . For  $\omega \notin A$ , set  $\tilde{X} = 0$  (or a fixed point of  $E$ ).

Let us check that, for  $t \in [0, 1]$ ,  $\mathbb{P}(X_t = \tilde{X}_t) = 1$ . By (\*) and Markov's inequality, if  $s_n \rightarrow t$  and  $s_n \in D$ , then  $X_{s_n} \xrightarrow{\mathbb{P}} X_t$ . On the other hand, by construction,  $X_{s_n} \xrightarrow{\text{a.s.}} \tilde{X}_t$ . Hence  $X_{s_n} \xrightarrow{\mathbb{P}} \tilde{X}_t$ . Thus  $\mathbb{P}(X_t = \tilde{X}_t) = 1$ .

**Application.** Let  $(X_i)_{i \geq 1}$  be i.i.d. real-valued random variables such that  $\mathbb{E}[X_1] = 0$  and  $\mathbb{E}[X_1^4] < \infty$ . Set  $S_0 = 0$ ,  $S_n = X_1 + \cdots + X_n$  for  $n \geq 1$  and

$$S_t = (1 - \{t\})S_{[t]} + \{t\}S_{[t]+1}, \quad t \geq 0.$$

Finally, define  $S_t^{(n)} = S_{nt}/\sqrt{n}$  for  $0 \leq t \leq 1$ , so that  $S^{(n)} \in C = \mathcal{C}([0, 1], \mathbb{R})$ . Then  $(S^{(n)})_{n \geq 1}$  is tight in  $C$ .

We will see later that this result remains true under the weaker assumption  $\mathbb{E}[X_1^2] < \infty$ , but the proof is more delicate.

*Proof.* Set  $v = \mathbb{E}[X_1^2]$ ,  $m_4 = \mathbb{E}[X_1^4]$ , and  $C = 1 + m_4 + 3v^2$ . For  $0 \leq s < t \leq 1$  and  $1 \leq i \leq n$ , write

$$a_i = \text{Leb}([ns, nt] \cap [i-1, i]).$$

By definition of the interpolation,  $S_{nt} - S_{ns} = \sum_{i=1}^n a_i X_i$ , with  $0 \leq a_i \leq 1$  and  $\sum_i a_i = n(t-s)$ . Independence and centering give

$$\begin{aligned} \mathbb{E}[|S_{nt} - S_{ns}|^4] &= m_4 \sum_i a_i^4 + 6v^2 \sum_{i < j} a_i^2 a_j^2 \\ &\leq C \left( \sum_i a_i^2 \right)^2 \leq Cn^2(t-s)^2. \end{aligned}$$

In particular, taking  $s = 0$  and  $t = 1$ ,

$$\mathbb{E}[S_n^4] \leq Cn^2. \quad (\dagger)$$

The same calculation gives, without restrictions on the endpoints  $s, t$ ,

$$\mathbb{E}[|S_{nt} - S_{ns}|^4] \leq Cn^2|t-s|^2. \quad (\dagger\dagger)$$

Dividing by  $n^2$ , we obtain  $\mathbb{E}[|S_t^{(n)} - S_s^{(n)}|^4] \leq C|t-s|^2$ . Finally, for  $s = 0$ , independence and centering also give

$$\mathbb{E}[|S_t^{(n)}|^2] = \frac{v}{n} \sum_i a_i^2 \leq vt.$$

The marginals are therefore tight by Markov's inequality. Kolmogorov's criterion applies with  $\alpha = 4$  and  $\beta = 1$ , and proves tightness of  $(S^{(n)})_{n \geq 1}$ .  $\square$

## 4.7 Continuous functions on $\mathbb{R}_+$

In practice, stochastic processes are often indexed by  $\mathbb{R}_+$  rather than  $[0, 1]$ . Since the uniform distance  $D_\infty(f, g) = \sup_{t \geq 0} d(f(t), g(t))$  on  $\mathcal{C}(\mathbb{R}_+, E)$  can equal  $+\infty$ , care is needed. We will always equip  $\mathcal{C}(\mathbb{R}_+, E)$  with a different topology, that of uniform convergence on compact sets, induced by the metric

$$D(f, g) = \sum_{n=1}^{\infty} \frac{\left( \sup_{0 \leq t \leq n} d(f(t), g(t)) \right) \wedge 1}{2^n}.$$

**Remark 4.23** (Optional). In topological terms, this is the compact-open topology on  $\mathcal{C}(\mathbb{R}_+, E)$ , a subbase for which is given by the sets

$$\{f \in \mathcal{C}(\mathbb{R}_+, E) : f(K) \subset O\},$$

with  $K$  compact in  $\mathbb{R}_+$  and  $O$  open in  $E$ .

One readily checks that:

- $f_n \rightarrow f$  in  $\mathcal{C}(\mathbb{R}_+, E)$  if and only if  $f_n \rightarrow f$  uniformly on compact sets;

- $\mathcal{C}(\mathbb{R}_+, E)$  is Polish;
- finite-dimensional marginals characterize measures in  $\mathcal{M}_1(\mathcal{C}(\mathbb{R}_+, E))$ , since the Borel sigma-algebra on  $\mathcal{C}(\mathbb{R}_+, E)$  again coincides with the trace on  $\mathcal{C}(\mathbb{R}_+, E)$  of the product sigma-algebra on  $E^{\mathbb{R}_+}$ .

An important advantage of this space is that relative compactness is a local property; see the following proposition (proof left as an exercise).

**Proposition 4.24.**

- A subset  $\mathcal{F}$  of  $\mathcal{C}(\mathbb{R}_+, E)$  is relatively compact if and only if, for every  $N \geq 1$ , the sets  $\mathcal{F}_N = \{f|_{[0, N]} : f \in \mathcal{F}\}$  are relatively compact in  $\mathcal{C}([0, N], E)$ , where  $f|_{[0, N]}$  is the restriction of  $f$  to  $[0, N]$ .
- A sequence  $(X^n)_{n \geq 1}$  of  $\mathcal{C}(\mathbb{R}_+, E)$ -valued random variables is tight if and only if, for every  $N \geq 1$ , the sequence of restrictions  $(X^n|_{[0, N]})_{n \geq 1}$  is tight in  $\mathcal{C}([0, N], E)$ .

Thus, convergence in distribution in  $\mathcal{C}(\mathbb{R}_+, E)$  reduces, by restriction, to convergence in distribution in the spaces  $\mathcal{C}(I, E)$ , where  $I$  is a compact interval, and the preceding results can be applied.

## Interlude: couplings

A *coupling* of two random variables  $X$  and  $Y$  (not necessarily defined on the same probability space) is a random variable  $(X', Y')$  such that  $X' \stackrel{d}{=} X$  and  $Y' \stackrel{d}{=} Y$ .

**Example C.1.** Let  $n \geq 1$  be an integer and  $0 \leq p \leq q \leq 1$ . Then, for every integer  $0 \leq k \leq n$ ,

$$\mathbb{P}(\text{Bin}(n, p) \geq k) \leq \mathbb{P}(\text{Bin}(n, q) \geq k).$$

*Proof.* We construct a coupling in which  $\text{Bin}(n, p) \leq \text{Bin}(n, q)$  by taking  $(U_i)_{1 \leq i \leq n}$  i.i.d. uniform on  $[0, 1]$  and setting

$$X' = \sum_{i=1}^n \mathbb{1}_{\{U_i \leq p\}}, \quad Y' = \sum_{i=1}^n \mathbb{1}_{\{U_i \leq q\}}.$$

□

**Outline.** Skorokhod's representation theorem; total variation distance; Lévy–Prokhorov metric; monotonicity and stochastic order.

**References for a survey of metrics on  $\mathcal{M}_1(E)$ .**

- Alison L. Gibbs, Francis Edward Su, *On choosing and bounding probability metrics*.
- Svante Janson, *Probability distances*.

## C.1 Skorokhod's representation theorem

**Theorem C.2** (Skorokhod representation). Let  $(E, d)$  be separable, and let  $(X_i)_{i \geq 1}$  and  $X$  be  $E$ -valued random variables such that  $X_n \xrightarrow{d} X$ . We can define random variables  $(X'_i)_{i \geq 1}$  and  $X'$  on a common probability space such that, for every  $n \geq 1$ ,

$$X_n \stackrel{d}{=} X'_n, \quad X \stackrel{d}{=} X', \quad X'_n \xrightarrow{\text{a.s.}} X'.$$

See the exercise sheet for the proof when  $E = \mathbb{R}$ . In practice, this theorem is used to reduce a convergence-in-distribution argument to a “deterministic argument” (we proceed “as if”  $X_n \rightarrow X$  deterministically).

**Caution.** All information about the joint distributions of the  $X_i$  is lost for the  $X'_i$ .

**Proposition C.3** (Application). Let  $X, X_1, X_2, \dots$  be real-valued random variables such that  $X_n \xrightarrow{d} X$ . Suppose that

$$\lim_{a \rightarrow \infty} \sup_{n \geq 1} \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| \geq a\}}] = 0. \quad (*)$$

Then  $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$ .

*Proof.* By Skorokhod's representation theorem, we may assume that  $X_n \xrightarrow{\text{a.s.}} X$  and that  $(*)$  holds (since this is a property of the individual distributions of the  $X_n$ ). Then  $X_n \xrightarrow{\mathbb{P}} X$ , and the conclusion follows from the fact that

$$X_n \xrightarrow{\mathbb{P}} X, \quad (X_n)_{n \geq 1} \text{ uniformly integrable} \implies X_n \xrightarrow{L^1} X.$$

□

For the proof, we assume the existence of independent random variables with prescribed distributions and use the following lemma.

**Lemma C.4.** Let  $(E, d)$  be separable and  $\mu \in \mathcal{M}_1(E)$ . For every  $\varepsilon > 0$ , there exists a partition  $E = \bigsqcup_{i \geq 1} A_i$  into measurable sets such that

$$\forall i \geq 1, \quad \text{diam}(A_i) \leq \varepsilon, \quad \mu(\partial A_i) = 0.$$

*Proof.* Let  $(x_i)_{i \geq 1}$  be a countable dense sequence. For every  $i \geq 1$ ,

$$\{r \in [\varepsilon/4, \varepsilon/2] : \mu(\partial B(x_i, r)) > 0\}$$

is countable. Choose  $\varepsilon/4 \leq r_i \leq \varepsilon/2$  such that  $\mu(\partial B(x_i, r_i)) = 0$ , and write  $B_i = B(x_i, r_i)$ . Define recursively

$$A_1 = B_1, \quad A_i = B_1^c \cap \dots \cap B_{i-1}^c \cap B_i,$$

and check that  $\mu(\partial A_i) = 0$ . Indeed,  $\partial A = \partial(A^c)$  and  $\partial(A \cup B) \subset \partial A \cup \partial B$ , so

$$\mu(\partial A) = 0 \implies \mu(\partial A^c) = 0, \quad \mu(\partial A) = \mu(\partial B) = 0 \implies \mu(\partial(A \cup B)) = 0.$$

□

*Proof of Skorokhod's representation theorem.* Write  $\mu_n$  for the distribution of  $X_n$  and  $\mu$  for the distribution of  $X$ .

**Step 1.** Let  $p \geq 1$  and let  $(A_i^p)_{i \geq 1}$  be a partition given by the lemma for  $\varepsilon_p = 2^{-p}$  and  $\mu$ . Let  $k_p$  satisfy

$$\mu \left( \bigcup_{i=k_p+1}^{\infty} A_i^p \right) < \varepsilon_p.$$

Set  $A_0^p = \bigcup_{i=k_p+1}^{\infty} A_i^p$ , so that  $\mu(A_0^p) < \varepsilon_p$  and  $\mu(\partial A_0^p) = 0$ . In particular,  $A_0^p, A_1^p, \dots, A_{k_p}^p$  is a partition of  $E$ . By convergence in distribution, we can find  $\alpha_p$  such that

$$n \geq \alpha_p \implies \mu_n(A_i^p) \geq (1 - \varepsilon_p)\mu(A_i^p) \quad \text{for } 0 \leq i \leq k_p. \quad (\star)$$

In particular,  $\mu_n(A_i^p) = 0 \implies \mu(A_i^p) = 0$ . Without loss of generality, assume  $\alpha_1 < \alpha_2 < \dots$  and  $n \geq \alpha_1$ . Let  $p_n \geq 1$  be the unique integer such that  $\alpha_{p_n} \leq n < \alpha_{p_n+1}$ . As  $n \rightarrow \infty$ , we have  $p_n \rightarrow \infty$ .

**Step 2.** The idea is essentially to define  $X'_n = X_n \mid X_n \in A_i^{p_n}$  when  $X \in A_i^{p_n}$ . However, a small adjustment is needed to ensure that  $X'_n \stackrel{d}{=} X_n$ . We define  $X'_n$  as follows:

$$X'_n = \begin{cases} Y_{ni}, & U \leq 1 - \varepsilon_{p_n}, X' \in A_i^{p_n}, \mu_n(A_i^{p_n}) > 0, \\ Z_n, & U > 1 - \varepsilon_{p_n}, \end{cases}$$

where  $(Y_{ni})$ ,  $(Z_n)_{n \geq 1}$ ,  $U$ , and  $X'$  are independent random variables defined on the same probability space such that

- $U$  is uniform on  $[0, 1]$ ;
- $X \stackrel{d}{=} X'$ ;
- $\mathcal{L}(Y_{ni}) = \mu_n(\cdot \mid A_i^{p_n})$ ;
- the distribution of  $Z_n$  is defined by

$$\mathbb{P}(Z_n \in A) = \frac{1}{\varepsilon_{p_n}} \sum_{\substack{0 \leq i \leq k_{p_n} \\ \mu_n(A_i^{p_n}) > 0}} \mu_n(A \mid A_i^{p_n}) (\mu_n(A_i^{p_n}) - (1 - \varepsilon_{p_n})\mu(A_i^{p_n})).$$

This quantity indeed defines a positive measure by  $(\star)$ , with total mass 1, since

$$\mu_n(A_i^{p_n}) = 0 \implies \mu(A_i^{p_n}) = 0.$$

We check that  $X'_n \stackrel{d}{=} X_n$ :

$$\begin{aligned}
 \mathbb{P}(X'_n \in A) &= \sum_{\substack{0 \leq i \leq k_{p_n} \\ \mu_n(A_i^{p_n}) > 0}} \mathbb{P}(U \leq 1 - \varepsilon_{p_n}, X' \in A_i^{p_n}, Y_{ni} \in A) \\
 &\quad + \mathbb{P}(U > 1 - \varepsilon_{p_n}, Z_n \in A) \\
 &= (1 - \varepsilon_{p_n}) \sum_{\substack{0 \leq i \leq k_{p_n} \\ \mu_n(A_i^{p_n}) > 0}} \underbrace{\mathbb{P}(X' \in A_i^{p_n})}_{\mu(A_i^{p_n})} \mu_n(A | A_i^{p_n}) + \varepsilon_{p_n} \mathbb{P}(Z_n \in A) \\
 &= \mu_n(A),
 \end{aligned}$$

since

$$\mathbb{P}(Z_n \in A) = \frac{1}{\varepsilon_{p_n}} \left( \mu_n(A) - (1 - \varepsilon_{p_n}) \sum_{\substack{0 \leq i \leq k_{p_n} \\ \mu_n(A_i^{p_n}) > 0}} \mu(A_i^{p_n}) \mu_n(A | A_i^{p_n}) \right).$$

**Step 3.** We check that  $X'_n \xrightarrow{\text{a.s.}} X'$ . Let

$$E_p = \{X' \notin A_0^p, U \leq 1 - \varepsilon_p\}.$$

We have

$$\mathbb{P}(E_p^c) \leq \mathbb{P}(X' \in A_0^p) + \mathbb{P}(U > 1 - \varepsilon_p) \leq \frac{1}{2^{p-1}}.$$

Thus  $\sum_{p \geq 1} \mathbb{P}(E_p^c) < \infty$ . By Borel–Cantelli, almost surely,  $E_p$  occurs for all sufficiently large indices. Thus, almost surely,  $E_{p_n}$  occurs for all sufficiently large  $n$ . By definition of  $X'_n$ , this implies that there exists  $1 \leq i \leq k_{p_n}$  such that  $X', X'_n \in A_i^{p_n}$ , and hence  $d(X', X'_n) \leq \varepsilon_{p_n}$ . Thus  $X'_n \xrightarrow{\text{a.s.}} X'$ .  $\square$

**Remark C.5.** When  $E$  is Polish, one can show that the common probability space can be taken to be  $[0, 1]$  equipped with Lebesgue measure.

## C.2 Total variation distance

Let  $(S, \mathcal{B})$  be a measurable space. For statements involving  $\mathbb{P}(X = Y)$  or  $\mathbb{P}(X \neq Y)$ , we also assume that the diagonal

$$\Delta_S = \{(x, x) : x \in S\}$$

belongs to  $\mathcal{B} \otimes \mathcal{B}$ . This holds, in particular, when  $S$  is a separable metric space equipped with its Borel sigma-algebra.

**Definition C.6.** For  $\mu, \nu \in \mathcal{M}_1(S)$ , set

$$d_{VT}(\mu, \nu) = \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|.$$

For  $S$ -valued random variables  $X, Y$ , set

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

The following results can be proved:

- $d_{VT}$  is a metric on  $\mathcal{M}_1(S)$  under which it is complete.
- If  $\mu$  and  $\nu$  are absolutely continuous with respect to a  $\sigma$ -finite measure  $\rho$  on  $(S, \mathcal{B})$ , then

$$d_{VT}(\mu, \nu) = \frac{1}{2} \int_S \left| \frac{d\mu}{d\rho} - \frac{d\nu}{d\rho} \right| d\rho.$$

This allows us to show that

$$d_{VT}(\mu, \nu) = \frac{1}{2} \sup \left\{ \left| \int f d\mu - \int f d\nu \right| : f : S \rightarrow \mathbb{R} \text{ measurable, } \|f\|_\infty \leq 1 \right\},$$

and that, if  $X$  and  $Y$  have distributions  $\mu$  and  $\nu$ , respectively,

$$d_{VT}(\mu, \nu) = \inf \{ \mathbb{P}(X' \neq Y') : (X', Y') \text{ is a coupling of } X, Y \},$$

with the infimum attained.

- When  $S$  is a metric space equipped with its Borel sigma-algebra,  $d_{LP} \leq d_{VT}$ .
- When  $S$  is countable, equipped with the discrete metric, and  $\mathcal{B} = \mathcal{P}(S)$ , convergence in  $d_{VT}$  is equivalent to weak convergence (**caution:** this is false in general), and

$$d_{VT}(\mu, \nu) = \frac{1}{2} \sum_{x \in S} |\mu(\{x\}) - \nu(\{x\})|.$$

**Remark C.7.** A  $p$ -metric on  $S$  is a function  $\delta$  with values in  $[0, \infty]$ , defined on the set of joint distributions of  $S$ -valued random variables  $(X, Y)$ , such that, for random variables  $X, Y, Z$  defined on the same space:

- (i) if  $X = Y$  almost surely, then  $\delta(X, Y) = 0$ ;
- (ii)  $\delta(X, Y) = \delta(Y, X)$ ;
- (iii)  $\delta(X, Y) \leq \delta(X, Z) + \delta(Z, Y)$ .

The quantity

$$\underline{\delta}(X, Y) = \inf \{ \delta(X', Y') : X' \stackrel{d}{=} X, Y' \stackrel{d}{=} Y \}$$

is called the *minimal metric* associated with  $\delta$ . Thus  $d_{VT}$  is the minimal metric associated with  $\delta(X, Y) = \mathbb{P}(X \neq Y)$ .

**Remark C.8.** For a nonempty set  $\mathcal{F}$  of bounded measurable functions  $S \rightarrow \mathbb{R}$ , and  $S$ -valued random variables  $X, Y$ , set

$$d_{\mathcal{F}}(X, Y) = \sup \{ |\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]| : f \in \mathcal{F} \}.$$

This quantity is a *dual pseudometric*, possibly infinite, that depends only on the distributions of  $X$  and  $Y$ . It defines a metric between the distributions under consideration when it is finite and  $\mathcal{F}$  separates them. Thus  $d_{VT}$  is the dual metric defined by any of the three families

$$\mathcal{F} = \{f : S \rightarrow \mathbb{R} \text{ measurable, } \|f\|_\infty \leq \tfrac{1}{2}\},$$

$$\mathcal{F} = \{f : S \rightarrow \mathbb{R} \text{ measurable, } 0 \leq f \leq 1\},$$

$$\mathcal{F} = \{\mathbb{1}_A : A \in \mathcal{B}\}.$$

**Reference.** Frank den Hollander's lecture notes *Probability Theory: The Coupling Method* (Section 2):

<http://websites.math.leidenuniv.nl/probability/lecturenotes/CouplingLectures.pdf>.

### C.3 The Lévy–Prokhorov metric

Let  $(E, d)$  be a separable metric space. Recall that, for  $\mu, \nu \in \mathcal{M}_1(E)$ ,

$$d_{LP}(\mu, \nu) = \inf\{\varepsilon > 0 : \forall A \in \mathcal{B}(E), \mu(A) \leq \nu(A^\varepsilon) + \varepsilon\},$$

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

**Definition C.9.** For random variables  $X, Y$  defined on the same probability space and taking values in  $E$ , set

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

This is the Ky Fan metric, and one can check that it metrizes convergence in probability.

**Theorem C.10** (Strassen). Let  $(E, d)$  be separable, and let  $\mu, \nu \in \mathcal{M}_1(E)$ . Then

$$d_{LP}(\mu, \nu) = \inf\{\alpha(X, Y) : (X, Y) \text{ is a coupling with } \mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu\}.$$

If  $\mu$  and  $\nu$  are tight (which is always the case if  $E$  is complete), then the infimum is attained.

*Proof of one inequality.* Let  $(X, Y)$  be such a coupling. We show that  $d_{LP}(\mu, \nu) \leq \alpha(X, Y)$ . Let  $\varepsilon > \alpha(X, Y)$ , so that  $\mathbb{P}(d(X, Y) \geq \varepsilon) < \varepsilon$ . Then

$$\begin{aligned} \mu(A) &= \mathbb{P}(X \in A) \leq \mathbb{P}(X \in A, d(X, Y) < \varepsilon) + \mathbb{P}(X \in A, d(X, Y) \geq \varepsilon) \\ &\leq \mathbb{P}(Y \in A^\varepsilon) + \varepsilon = \nu(A^\varepsilon) + \varepsilon, \end{aligned}$$

hence  $d_{LP}(\mu, \nu) \leq \varepsilon$ . □

The other inequality is delicate and can be proved using Hall's marriage theorem; see Billingsley, *Convergence of Probability Measures*, second edition, Theorem 6.9, or Dudley, *Real Analysis and Probability*, Corollary 11.6.4.

**Remark C.11.** Thus  $d_{LP}$  is the minimal metric associated with  $\alpha$ .

We end this section with a useful coupling result in the spirit of Skorokhod's theorem.

**Theorem C.12** (Dudley). Let  $(E, d)$  be separable, and let  $X_n, Y_n$  be  $E$ -valued random variables such that  $d_{LP}(X_n, Y_n) \rightarrow 0$ . Then there exist a probability space and random variables  $(X'_n, Y'_n)_{n \geq 1}$  defined on it such that

$$\forall n \geq 1, \quad X'_n \stackrel{d}{=} X_n, \quad Y'_n \stackrel{d}{=} Y_n, \quad d(X'_n, Y'_n) \rightarrow 0 \quad \text{a.s.}$$

See Dudley, *Real Analysis and Probability*, Theorem 11.7.1.

## C.4 Monotonicity and stochastic order

Here,  $(E, d)$  is a Polish space.

**Definition C.13.** A relation  $\preceq$  is a partial order if, for every  $x, y, z \in E$ :

- (i)  $x \preceq x$ ;
- (ii)  $x \preceq y, y \preceq z \implies x \preceq z$ ;
- (iii)  $x \preceq y, y \preceq x \implies x = y$ .

**Example C.14.** •  $\mathbb{R}$  equipped with the usual order.

- $\{0, 1\}^{\mathbb{Z}^2}$ , with  $(a_i)_{i \in \mathbb{Z}^2} \preceq (b_i)_{i \in \mathbb{Z}^2}$  if, for every  $i$ ,  $a_i = 1 \implies b_i = 1$ . If  $a_i = 1$ , site  $i$  is said to be occupied, and vacant otherwise.

**Definition C.15.** An event  $A$  is said to be *increasing* if  $x \in A \implies \{y \in E : x \preceq y\} \subset A$ .

**Example C.16.** For  $E = \{0, 1\}^{\mathbb{Z}^2}$ , the event

$$A = \{\text{there exists an infinite path of adjacent occupied sites}\}$$

is increasing.

**Definition C.17.** Let  $\mu, \nu \in \mathcal{M}_1(E)$ . We say that  $\nu$  *stochastically dominates*  $\mu$ , and write  $\mu \preceq \nu$ , if

$$\forall A \in \mathcal{B}(E) \text{ increasing}, \quad \mu(A) \leq \nu(A),$$

or, equivalently,

$$\int_E f \, d\mu \leq \int_E f \, d\nu \quad \forall f : E \rightarrow \mathbb{R} \text{ measurable, bounded, nondecreasing.}$$

By nondecreasing, we mean  $x \preceq y \implies f(x) \leq f(y)$ .

**Example C.18.** For  $E = \mathbb{R}$ ,  $\mu \preceq \nu$  if and only if, for every  $x \in \mathbb{R}$ ,  $\mu([-\infty, x]) \geq \nu([-\infty, x])$ .

**Theorem C.19** (Strassen). Suppose that  $\{(x, y) \in E^2 : x \preceq y\}$  is closed. Let  $\mu, \nu \in \mathcal{M}_1(E)$ . Then the following two assertions are equivalent:

- (i)  $\mu \preceq \nu$ ;
- (ii) there exists a coupling  $(X, Y)$  such that  $\mathcal{L}(X) = \mu$ ,  $\mathcal{L}(Y) = \nu$ , and  $\mathbb{P}(X \preceq Y) = 1$ .

The proof relies on functional analysis; see Lindvall, *Lectures on the Coupling Method*, Chapter IV, or Lindvall, *On Strassen's Theorem on Stochastic Domination*, ECP 4 (1999), 51–59.

## 5 Donsker's theorem and applications

Related exercises: Sheet 5, p. 179.

In this chapter

|     |                   |    |
|-----|-------------------|----|
| 5.1 | Donsker's theorem | 55 |
| 5.2 | The arcsine law   | 58 |
| 5.3 | Lévy's theorem    | 60 |

### 5.1 Donsker's theorem

The following inequality will be useful in proving Donsker's invariance principle.

**Proposition 5.1** (Skorokhod's inequality). Let  $(X_i)_{i \geq 1}$  be i.i.d. random variables taking values in  $\mathbb{R}^d$ , with mean zero, such that  $c = \mathbb{E}[|X_1|^2] < \infty$ . Set  $S_0 = 0$  and  $S_n = X_1 + \cdots + X_n$  for  $n \geq 1$ . Then, for all  $\lambda > 0$  and  $n \geq 1$ ,

$$\mathbb{P} \left( \max_{1 \leq i \leq n} |S_i| > \lambda \sqrt{n} \right) \leq 2 \mathbb{P} \left( |S_n| > (\lambda - \sqrt{2c}) \sqrt{n} \right),$$

where  $|\cdot|$  denotes the Euclidean norm.

*Proof.* If  $c = 0$ , then  $X_i = 0$  almost surely for every  $i$ , and the inequality is immediate. Assume  $c > 0$ . If  $0 < \lambda < \sqrt{2c}$ , the right-hand side equals 2. It therefore remains to consider  $\lambda \geq \sqrt{2c}$ . Let  $T = \inf\{k \geq 1 : |S_k| > \lambda \sqrt{n}\}$ ; we write

$$\begin{aligned} \mathbb{P} \left( \max_{1 \leq i \leq n} |S_i| > \lambda \sqrt{n} \right) &= \mathbb{P}(T \leq n) \\ &\leq \mathbb{P}(|S_n| > (\lambda - \sqrt{2c}) \sqrt{n}) + \mathbb{P}(T \leq n, |S_n| \leq (\lambda - \sqrt{2c}) \sqrt{n}). \end{aligned}$$

Now, for  $1 \leq j \leq n$ , on the event  $\{T = j\}$ , we have  $|S_j| > \lambda \sqrt{n}$ ; and if also  $|S_n| \leq (\lambda - \sqrt{2c}) \sqrt{n}$ , then  $|S_n - S_j| > \sqrt{2cn}$  by the triangle inequality. By independence and Markov's inequality,

$$\begin{aligned} \mathbb{P}(T = j, |S_n - S_j| > \sqrt{2cn}) &= \mathbb{P}(T = j) \mathbb{P}(|S_{n-j}| > \sqrt{2cn}) \\ &\leq \mathbb{P}(T = j) \frac{(n-j) \mathbb{E}[|X_1|^2]}{2cn} \leq \frac{\mathbb{P}(T = j)}{2}. \end{aligned}$$

We conclude that

$$\mathbb{P} \left( \max_{1 \leq i \leq n} |S_i| > \lambda \sqrt{n} \right) \leq \mathbb{P}(|S_n| > (\lambda - \sqrt{2c}) \sqrt{n}) + \frac{1}{2} \mathbb{P}(T \leq n),$$

which gives the result.  $\square$

We begin with some notation. Let  $(X_i)_{i \geq 1}$  be i.i.d. random variables in  $\mathbb{R}^d$  such that  $\mathbb{E}[X_1] = 0$  and  $\text{Cov}(X_1) = I_d$ , the identity matrix in  $\mathbb{R}^d$ . Set  $S_0 = 0$ ,  $S_n = X_1 + \cdots + X_n$

for  $n \geq 1$ , and, by linear interpolation,

$$S_t = (1 - \{t\})S_{[t]} + \{t\}S_{[t]+1} \quad (t \geq 0).$$

Finally, for  $n \geq 1$ , set

$$S_t^{(n)} = \frac{S_{nt}}{\sqrt{n}}, \quad t \geq 0,$$

so that  $S^{(n)}$  is a random variable taking values in  $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ .

**Theorem 5.2 (Donsker).** The sequence  $S^{(n)}$  converges in distribution in  $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$  to a process  $(B_t)_{t \geq 0}$  called *standard d-dimensional Brownian motion*. Its law is characterized by the following properties:  $B_0 = 0$  almost surely and, for all  $0 = t_0 \leq t_1 \leq \dots \leq t_k$ , the random variables  $(B_{t_i} - B_{t_{i-1}})_{1 \leq i \leq k}$  are independent, with respective Gaussian distributions  $\mathcal{N}(0, (t_i - t_{i-1})I_d)$ .

*Proof.* **Step 1: convergence of finite-dimensional distributions.** Fix  $0 = t_0 < t_1 < \dots < t_k$ . By the CLT,

$$\left( \frac{S_{[nt_i]} - S_{[nt_{i-1}]}}{\sqrt{n}} : 1 \leq i \leq k \right) \xrightarrow{d} (N_1, N_2, \dots, N_k),$$

where  $N_1, \dots, N_k$  are independent and  $N_i$  has distribution  $\mathcal{N}(0, (t_i - t_{i-1})I_d)$ . By the continuous mapping theorem,

$$\left( \frac{S_{[nt_i]}}{\sqrt{n}} : 1 \leq i \leq k \right) \xrightarrow{d} (N_1, N_1 + N_2, \dots, N_1 + \dots + N_k).$$

To remove the integer parts, note that

$$\left| S_t^{(n)} - \frac{S_{[nt]}}{\sqrt{n}} \right| \leq \frac{|X_{[nt]+1}|}{\sqrt{n}},$$

which converges in probability to 0 for fixed  $t$ . Thus

$$\left( S_{t_i}^{(n)} - \frac{S_{[nt_i]}}{\sqrt{n}} : 1 \leq i \leq k \right) \xrightarrow{\mathbb{P}} 0.$$

By Slutsky's lemma, we deduce that

$$(S_{t_i}^{(n)} : 1 \leq i \leq k) \xrightarrow{d} (N_1, N_1 + N_2, \dots, N_1 + \dots + N_k),$$

which proves convergence of the finite-dimensional distributions.

**Step 2: tightness.** We must check that, for every  $N \geq 1$ , the restrictions  $(S_{[0, N]}^{(n)})_{n \geq 1}$  are tight. To simplify notation, we do this for  $N = 1$  (the general case is identical). Under additional moment assumptions, one can use Kolmogorov's criterion. In the general case, we will show that

$$\forall \varepsilon, \eta > 0, \exists \delta \in ]0, 1[, \forall t \in [0, 1], \limsup_{n \rightarrow \infty} \frac{1}{\delta} \mathbb{P} \left( \sup_{t \leq s \leq (t+\delta) \wedge 1} |S_s^{(n)} - S_t^{(n)}| > \eta \right) \leq \varepsilon,$$

using Skorokhod's maximal inequality.

a) To this end, first note that, for  $0 < \delta < 1$  and  $0 \leq t \leq 1$ ,

$$\sup_{t \leq s \leq t+\delta} |S_{ns} - S_{nt}| \leq 2 \max_{0 \leq i \leq [n\delta]+2} |S_{[nt]+i} - S_{[nt]}|.$$

Indeed, we write

$$|S_{ns} - S_{nt}| \leq |S_{ns} - S_{[nt]}| + |S_{[nt]} - S_{[nt]}|.$$

On the one hand,

$$|S_{[nt]} - S_{[nt]}| \leq |S_{[nt]+1} - S_{[nt]}|,$$

since  $(S_{[nt]+u} : 0 \leq u \leq 1)$  traces a line segment. On the other hand,

$$|S_{ns} - S_{[nt]}| \leq \max(|S_{[ns]} - S_{[nt]}|, |S_{[ns]+1} - S_{[nt]}|),$$

since  $(S_{[ns]+u} : 0 \leq u \leq 1)$  traces a line segment. Since  $[ns] + 1 \leq [nt] + [n\delta] + 2$ , the desired inequality follows.

b) Thus, for all  $\eta > 0$ ,  $\delta \in ]0, 1[$  and  $t \in [0, 1]$ ,

$$\begin{aligned} \mathbb{P} \left( \sup_{t \leq s \leq t+\delta} |S_s^{(n)} - S_t^{(n)}| > \eta \right) &\leq \mathbb{P} \left( \max_{1 \leq i \leq [n\delta]+2} |S_{[nt]+i} - S_{[nt]}| > \frac{\eta\sqrt{n}}{2} \right) \\ &= \mathbb{P} \left( \max_{1 \leq i \leq [n\delta]+2} |S_i| > \frac{\eta\sqrt{n}}{2\sqrt{[n\delta]+2}} \sqrt{[n\delta]+2} \right) \\ &\leq 2\mathbb{P} \left( \frac{|S_{[n\delta]+2}|}{\sqrt{[n\delta]+2}} > \frac{\eta}{2} \frac{1}{\sqrt{\delta+2/n}} - \sqrt{2d} \right), \end{aligned}$$

by Skorokhod's inequality, since  $\mathbb{E}[|X_1|^2] = d$ . By the CLT,

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{t \leq s \leq t+\delta} |S_s^{(n)} - S_t^{(n)}| > \eta \right) \leq 2\mathbb{P} \left( |\mathcal{N}(0, I_d)| > \frac{\eta}{2\sqrt{\delta}} - \sqrt{2d} \right).$$

Now  $\mathbb{P}(|\mathcal{N}(0, I_d)| \geq x)$  decays exponentially fast as  $x \rightarrow \infty$ . We can therefore find  $0 < \delta < 1$  such that

$$\frac{1}{\delta} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{t \leq s \leq t+\delta} |S_s^{(n)} - S_t^{(n)}| > \eta \right) \leq \varepsilon.$$

□

**Corollary 5.3.** Let  $(X_i)_{i \geq 1}$  be i.i.d. real-valued random variables with mean zero. Write  $\sigma^2 = \mathbb{E}[X_1^2] \in ]0, \infty[$ . Then, for every  $a > 0$ ,

$$\mathbb{P} \left( \max_{0 \leq i \leq n} S_i > a\sigma\sqrt{n} \right) \rightarrow 2 \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

*Proof.* Writing  $S_t^{(n)} = S_{nt}/(\sigma\sqrt{n})$  for  $0 \leq t \leq 1$ , we have

$$(S_t^{(n)} : 0 \leq t \leq 1) \xrightarrow{d} (B_t)_{0 \leq t \leq 1}$$

by Donsker's theorem, where  $(B_t)_{0 \leq t \leq 1}$  is a standard Brownian motion. Since

$$\sup_{0 \leq t \leq 1} S_t^{(n)} = \frac{1}{\sigma\sqrt{n}} \max_{0 \leq i \leq n} S_i$$

and the map  $\sup : \mathcal{C}([0, 1], \mathbb{R}) \rightarrow \mathbb{R}$ ,  $f \mapsto \sup f$ , is continuous, the result follows using the fact that

$$\sup_{0 \leq t \leq 1} B_t \stackrel{d}{=} |B_1|$$

(which can be seen using the reflection principle).  $\square$

A property of the “continuous” setting (the explicit law of  $\sup_{0 \leq t \leq 1} B_t$ ) has been used to obtain an asymptotic result in the discrete setting (the distribution function of  $\max_{0 \leq k \leq n} S_k$ ). In the next two examples, passing to the limit in a discrete result yields interesting information about Brownian motion.

## 5.2 The arcsine law

**Theorem 5.4.** Let  $(B_t)_{0 \leq t \leq 1}$  be a standard Brownian motion. Set

$$T = \sup\{t \in [0, 1] : B_t = 0\}.$$

Then  $T$  has density  $\frac{1}{\pi\sqrt{x(1-x)}} \mathbb{1}_{\{0 < x < 1\}}$ , or, equivalently,

$$\mathbb{P}(T \leq t) = \frac{2}{\pi} \arcsin(\sqrt{t}) \quad (0 \leq t \leq 1).$$

*Proof. Step 1.* Set  $S_n = X_1 + \dots + X_n$ . The variables  $(X_i)_{i \geq 1}$  are i.i.d., with distribution  $\mathbb{P}(X_i = \pm 1) = 1/2$ , and we write

$$T_{2n} = \max\{0 \leq k \leq 2n : S_k = 0\}.$$

Let us show that  $T_{2n}/(2n)$  converges in distribution to a random variable with density

$$\frac{1}{\pi\sqrt{x(1-x)}} \mathbb{1}_{\{0 < x < 1\}}.$$

For  $0 \leq j \leq n$ , we have

$$\mathbb{P}(T_{2n} = 2j) = \mathbb{P}(S_{2j} = 0) \mathbb{P}(S_1 \neq 0, \dots, S_{2(n-j)} \neq 0).$$

Now

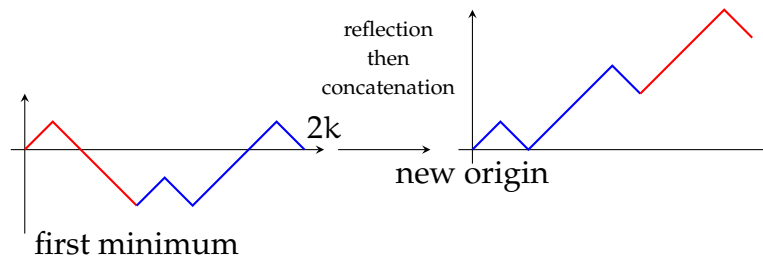
$$\forall k \geq 1, \quad \mathbb{P}(S_1 \neq 0, \dots, S_{2k} \neq 0) = \mathbb{P}(S_{2k} = 0). \quad (*)$$

Let us prove this surprising identity by a neat combinatorial argument.

First,

$$\begin{aligned}\mathbb{P}(S_1 \neq 0, \dots, S_{2k} \neq 0) &= 2\mathbb{P}(S_1 > 0, \dots, S_{2k} > 0) \\ &= 2 \times \frac{1}{2} \mathbb{P}(S_1 \geq 0, \dots, S_{2k-1} \geq 0) \\ &= \mathbb{P}(S_1 \geq 0, \dots, S_{2k} \geq 0).\end{aligned}$$

The second equality follows by moving the origin to  $(1, 1)$ ; the last follows by parity, since  $S_{2k-1} \geq 0 \implies S_{2k} \geq 0$ . We now construct a bijection between paths  $(S_1, \dots, S_{2k})$  such that  $S_{2k} = 0$  and those such that  $S_1 \geq 0, \dots, S_{2k} \geq 0$ .



Take the part before the first time the infimum is attained. Remove it and reflect it across the  $(Oy)$  axis. Concatenate it to the right of the remaining part (in blue) and place the origin at the first point of the blue part (see the figure). After the transformation, we obtain a path such that  $S_1 \geq 0, \dots, S_{2k} \geq 0$ . One checks that this is a bijection (the inverse transformation is easily described). This proves  $(*)$ . Thus,

$$\mathbb{P}(T_{2n} = 2j) = \mathbb{P}(S_{2j} = 0) \mathbb{P}(S_{2(n-j)} = 0).$$

Now

$$\mathbb{P}(S_{2k} = 0) = \binom{2k}{k} \frac{1}{2^{2k}} \underset{k \rightarrow \infty}{\sim} \frac{1}{\sqrt{\pi k}}.$$

Hence, for  $0 < c < 1$ ,

$$\mathbb{P}(T_{2n} = 2\lfloor cn \rfloor) \underset{n \rightarrow \infty}{\sim} \frac{1}{n} \frac{1}{\pi \sqrt{c(1-c)}}.$$

Apply Lemma 6.3 with  $k = 1$ ,  $X_n = T_{2n}/2$  and  $\alpha_n = n$ . The local limit above holds for  $c \in ]0, 1[$ ; for  $c \notin ]0, 1[$ , the probability under consideration is zero for all sufficiently large  $n$ . Thus, for almost every  $c \in \mathbb{R}$ , we obtain the density  $\phi(c) = (\pi \sqrt{c(1-c)})^{-1}$  on  $]0, 1[$ , extended by 0 elsewhere. Since its integral equals 1, the lemma gives the claimed convergence in distribution of  $T_{2n}/(2n)$ .

**Step 2.** We work on the closed subspace

$$\mathcal{C}_0 = \{f \in \mathcal{C}([0, 1], \mathbb{R}) : f(0) = 0\}.$$

For  $f \in \mathcal{C}_0$ , set

$$T(f) = \max\{t \in [0, 1] : f(t) = 0\}.$$

This map is well defined and measurable: for  $\alpha \in [0, 1]$ , the set  $\{f : T(f) \geq \alpha\}$  is closed. It is not continuous everywhere: for  $f_n(x) = x/n$ , we have  $T(f_n) = 0$ , whereas  $T(0) = 1$ .

However, it is continuous at every function  $f$  such that  $0 < T(f) < 1$  and which takes both signs in every neighborhood of  $T(f)$ . Indeed, a pair of values with opposite signs near  $T(f)$  retains its opposite signs after a sufficiently small uniform perturbation, yielding a nearby zero by the intermediate value theorem. On the other hand, for every  $\varepsilon > 0$  such that  $T(f) + \varepsilon \leq 1$ , the minimum of  $|f|$  on  $[T(f) + \varepsilon, 1]$  is strictly positive: no small uniform perturbation can have a zero in this interval.

Almost surely,  $B_1 \neq 0$  and  $B$  takes both signs arbitrarily close to 0, so  $0 < T(B) < 1$ . Moreover, 0 is not the value of any interior local extremum. Indeed, for rational numbers  $0 < p < q < 1$ , the minimum of  $B$  on  $[p, q]$  can be written as

$$B_p + \inf_{0 \leq u \leq q-p} (B_{p+u} - B_p).$$

The two terms are independent and  $B_p$  has a density: this minimum therefore has no atom at 0. The same holds for the maximum. A countable union over intervals with rational endpoints then rules out any interior local extremum with value 0. Thus  $B$  takes both signs in every neighborhood of  $T(B)$ , and almost surely satisfies the preceding continuity condition. By Donsker's theorem and the continuous mapping theorem,

$$T \left( \left( \frac{S_{2nt}}{\sqrt{2n}} : 0 \leq t \leq 1 \right) \right) \xrightarrow{d} T(B).$$

For the simple random walk considered in Step 1, the zeros of the linear interpolation occur at integer times. The left-hand side therefore equals  $T_{2n}/(2n)$ , identifying the law of  $T(B)$ .  $\square$

**Corollary 5.5.** Let  $(X_i)_{i \geq 1}$  be i.i.d. real-valued random variables such that  $\mathbb{E}[X_1] = 0$  and  $\mathbb{E}[X_1^2] \in ]0, \infty[$ . Set  $S_0 = 0$ ,  $S_k = X_1 + \cdots + X_k$ , and extend  $S$  by linear interpolation. Then

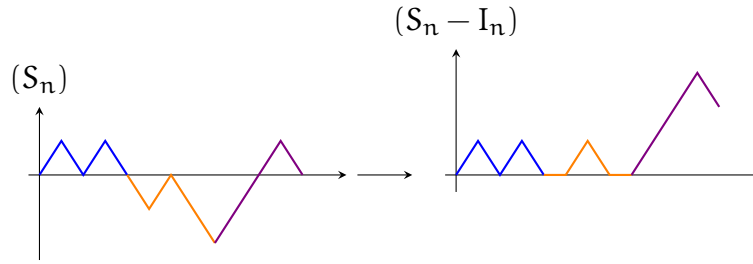
$$\mathbb{P} \left( \frac{1}{2n} \max\{u \in [0, 2n] : S_u = 0\} \leq t \right) \longrightarrow \frac{2}{\pi} \arcsin(\sqrt{t}) \quad (0 < t < 1).$$

### 5.3 Lévy's theorem

Let  $(B_t)_{t \geq 0}$  be a Brownian motion and  $I_t = \inf_{[0, t]} B$  its running infimum.

**Theorem 5.6 (Lévy).** We have  $(B_t - I_t)_{t \geq 0} \stackrel{d}{=} (|B_t|)_{t \geq 0}$ .

*Proof.* Let  $(X_i)_{i \geq 1}$  be i.i.d. random variables such that  $\mathbb{P}(X_i = \pm 1) = 1/2$ . Set  $S_n = X_1 + \cdots + X_n$  and  $I_n = \min_{0 \leq k \leq n} S_k$ .



Comparing the transition probabilities of  $S_n - I_n$  and  $|S_n|$ , we see that they agree away from state 0, where  $S_n - I_n$  can remain with probability  $1/2$ . Set  $R_n = S_n - I_n$  and

$$\theta_n = \#\{1 \leq k \leq n : R_k \neq R_{k-1}\}, \quad \theta_0 = 0.$$

A step is deleted exactly when the walk reaches a new strict minimum, so that

$$n - \theta_n = -I_n.$$

The strong law of large numbers implies  $I_n/n \rightarrow 0$  almost surely. Extending  $\theta$  by linear interpolation, we therefore obtain, almost surely, for every  $T > 0$ ,

$$\sup_{0 \leq t \leq T} \left| \frac{\theta_{nt}}{n} - t \right| \rightarrow 0.$$

Indeed, this supremum is bounded above by  $n^{-1} \max_{0 \leq k \leq \lceil nT \rceil} |S_k|$ , which tends to 0.

For  $m \in \mathbb{N}$ , set

$$\tau_m = \inf\{k \in \mathbb{N} : \theta_k = m\}, \quad Y_m = R_{\tau_m}.$$

These times are almost surely finite since  $\theta_n/n \rightarrow 1$ . The chain  $(Y_m)_{m \geq 0}$  has the same transitions as  $(|S_m|)_{m \geq 0}$ , including at 0; their laws are therefore equal. Extend  $R$  and  $Y$  by linear interpolation. Then, for every  $t \geq 0$ ,

$$R_t = Y_{\theta_t}.$$

By Donsker's theorem,

$$(Y_{nt}/\sqrt{n})_{t \geq 0} \xrightarrow{d} (|B_t|)_{t \geq 0}.$$

Convergence of the time change to the identity, Slutsky's lemma and continuity of composition give

$$(R_{nt}/\sqrt{n})_{t \geq 0} \xrightarrow{d} (|B_t|)_{t \geq 0}.$$

On the other hand,  $R_t = S_t - \inf_{0 \leq u \leq t} S_u$  for the interpolated simple random walk. The map subtracting the running infimum is continuous for uniform convergence on compact sets; Donsker's theorem therefore also gives

$$(R_{nt}/\sqrt{n})_{t \geq 0} \xrightarrow{d} (B_t - I_t)_{t \geq 0}.$$

Uniqueness of the limit in distribution completes the proof. □

## 6 The Brownian bridge

Related exercises: Sheet 5, p. 179.

In this chapter

|     |                                              |    |
|-----|----------------------------------------------|----|
| 6.1 | Definition of the Brownian bridge . . . . .  | 62 |
| 6.2 | The local limit theorem . . . . .            | 63 |
| 6.3 | Convergence to the Brownian bridge . . . . . | 66 |

### 6.1 Definition of the Brownian bridge

Let  $(B_t)_{0 \leq t \leq 1}$  be a standard Brownian motion.

**Theorem 6.1** (Definition of the Brownian bridge). As  $\varepsilon \downarrow 0$ , the law of  $(B_s)_{0 \leq s \leq 1}$  conditional on  $\{|B_1| \leq \varepsilon\}$  converges weakly to a probability measure  $\mathbb{P}_0$  on  $\mathcal{C}([0, 1], \mathbb{R})$ , called the *Brownian bridge law*. Moreover, the law of  $(B_s - sB_1)_{0 \leq s \leq 1}$  is  $\mathbb{P}_0$ .

Informally, the Brownian bridge is “ $(B_t)_{0 \leq t \leq 1}$  conditional on  $B_1 = 0$ ”.

*Proof.* Set  $b_s = B_s - sB_1$  for  $0 \leq s \leq 1$ . The family  $((b_s)_{0 \leq s \leq 1}, B_1)$  is Gaussian and, by the covariance of Brownian motion,

$$\text{Cov}(b_s, B_1) = \text{Cov}(B_s, B_1) - s \text{Var}(B_1) = s - s = 0.$$

Thus  $b$  is independent of  $B_1$ . Let  $F : \mathcal{C}([0, 1], \mathbb{R}) \rightarrow \mathbb{R}$  be bounded and Lipschitz, with constant  $L$ . Since  $\|B - b\|_\infty = |B_1|$ , independence gives, for every  $\varepsilon > 0$ ,

$$\begin{aligned} |\mathbb{E}[F(B) \mid |B_1| \leq \varepsilon] - \mathbb{E}[F(b)]| &= |\mathbb{E}[F(B) - F(b) \mid |B_1| \leq \varepsilon]| \\ &\leq L \mathbb{E}[|B - b|_\infty \mid |B_1| \leq \varepsilon] \leq L\varepsilon. \end{aligned}$$

The characterization of convergence in distribution by bounded Lipschitz functions completes the proof: the limit is the law of  $b$ .  $\square$

In what follows, write  $p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$  for the density of  $B_t$ , for  $t > 0$ .

**Proposition 6.2.** The Brownian bridge law is characterized by the following absolute continuity relation: for  $0 < t < 1$  and  $F : \mathcal{C}([0, t], \mathbb{R}) \rightarrow \mathbb{R}$  bounded continuous or nonnegative measurable,

$$\mathbb{E}[F(b_s : 0 \leq s \leq t)] = \mathbb{E} \left[ F(B_s : 0 \leq s \leq t) \frac{p_{1-t}(-B_t)}{p_1(0)} \right].$$

*Proof.* It suffices to prove the result for bounded continuous  $F : \mathcal{C}([0, t], \mathbb{R}) \rightarrow \mathbb{R}$ . This yields the result for  $F = \mathbb{1}_A$  with  $A$  closed by approximation, then for  $F = \mathbb{1}_A$  with  $A$  Borel by a monotone class argument, then for simple  $F$  by linearity, and finally for measurable  $F \geq 0$  by approximation and monotone convergence. Set  $W_u = B_{t+u} - B_t$ ,

so that  $(W_u)_{0 \leq u \leq 1-t}$  is a Brownian motion independent of  $(B_u)_{0 \leq u \leq t}$ . Then

$$\begin{aligned} & \mathbb{E}[F(B_s : 0 \leq s \leq t) \mid |B_1| \leq \varepsilon] \\ &= \frac{1}{\mathbb{P}(|B_1| \leq \varepsilon)} \mathbb{E}[F(B_s : 0 \leq s \leq t) \mathbb{1}_{\{|B_t + W_{1-t}| \leq \varepsilon\}}] \\ &= \frac{1}{\mathbb{P}(|B_1| \leq \varepsilon)} \mathbb{E}[F(B_s : 0 \leq s \leq t) \mathbb{1}_{\{W_{1-t} \in [-B_t - \varepsilon, -B_t + \varepsilon]\}}] \\ &= \frac{1}{\int_{-\varepsilon}^{\varepsilon} p_1(x) dx} \mathbb{E} \left[ F(B_s : 0 \leq s \leq t) \int_{-B_t - \varepsilon}^{-B_t + \varepsilon} p_{1-t}(x) dx \right]. \end{aligned}$$

Now

$$\frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} p_1(x) dx \longrightarrow p_1(0), \quad \frac{1}{2\varepsilon} \int_{-B_t - \varepsilon}^{-B_t + \varepsilon} p_{1-t}(x) dx \longrightarrow p_{1-t}(-B_t).$$

Moreover, since  $p_{1-t}$  is bounded,

$$\frac{1}{2\varepsilon} \int_{-B_t - \varepsilon}^{-B_t + \varepsilon} p_{1-t}(x) dx \leq \|p_{1-t}\|_{\infty}.$$

By dominated convergence,

$$\mathbb{E}[F(B_s : 0 \leq s \leq t) \mid |B_1| \leq \varepsilon] \longrightarrow \mathbb{E} \left[ F(B_s : 0 \leq s \leq t) \frac{p_{1-t}(-B_t)}{p_1(0)} \right].$$

Since also  $\mathbb{E}[F(B_s : 0 \leq s \leq t) \mid |B_1| \leq \varepsilon] \rightarrow \mathbb{E}[F(b_s : 0 \leq s \leq t)]$ , the result follows.  $\square$

## 6.2 The local limit theorem

**Lemma 6.3.** Let  $k \geq 1$  be an integer; for  $1 \leq i \leq k$ , let  $(\alpha_n^i)$  be a sequence of strictly positive real numbers such that  $\alpha_n^i \rightarrow \infty$ . Consider a random variable  $(X_n^1, \dots, X_n^k)$  taking values in  $\mathbb{Z}^k$  such that, for Lebesgue-almost every  $(\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ ,

$$\alpha_n^1 \alpha_n^2 \cdots \alpha_n^k \mathbb{P}((X_n^1, \dots, X_n^k) = (\lfloor \lambda_1 \alpha_n^1 \rfloor, \dots, \lfloor \lambda_k \alpha_n^k \rfloor)) \longrightarrow \phi(\lambda_1, \dots, \lambda_k),$$

where  $\phi$  is a probability density on  $\mathbb{R}^k$ . Then

$$\left( \frac{X_n^1}{\alpha_n^1}, \dots, \frac{X_n^k}{\alpha_n^k} \right) \xrightarrow{d} X,$$

where  $X$  has density  $\phi$ .

*Proof.* We will use Scheffé's lemma (see [Exercise 1.4](#)): if  $X_n$  has density  $f_n$  on  $\mathbb{R}^k$  and  $f_n \rightarrow f$  almost everywhere, where  $f$  is a probability density, then  $X_n \xrightarrow{d} X$ , where  $X$  has density  $f$ . Let  $(Y_n^1, \dots, Y_n^k)$  be a random variable with density

$$\phi_n(\lambda_1, \dots, \lambda_k) = \alpha_n^1 \cdots \alpha_n^k \mathbb{P}((X_n^1, \dots, X_n^k) = (\lfloor \lambda_1 \alpha_n^1 \rfloor, \dots, \lfloor \lambda_k \alpha_n^k \rfloor))$$

on  $\mathbb{R}^k$ , and let

$$(Z_n^1, \dots, Z_n^k) = (\lfloor \alpha_n^1 Y_n^1 \rfloor, \dots, \lfloor \alpha_n^k Y_n^k \rfloor),$$

which has the same law as  $(X_n^1, \dots, X_n^k)$ . Since  $\phi_n \rightarrow \phi$  almost everywhere, Scheffé's lemma gives  $(Y_n^1, \dots, Y_n^k) \xrightarrow{d} X$ ; and since

$$\left| Y_n^i - \frac{Z_n^i}{\alpha_n^i} \right| \leq \frac{1}{\alpha_n^i} \rightarrow 0 \quad (1 \leq i \leq k),$$

the result follows.  $\square$

**Definition 6.4.** A real-valued random variable  $X$  is called *lattice* if there exist  $b \in \mathbb{R}$  and  $h \in \mathbb{R}_+^*$  such that  $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$ . If  $X$  is not almost surely constant, there is a largest such  $h$ , called the *span* of  $X$ . If it equals 1,  $X$  is called *aperiodic*. In the degenerate case where  $X = b$  almost surely, every  $h > 0$  works and there is no maximal finite span.

**Example 6.5.** If  $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ ,  $X$  has span 2.

**Lemma 6.6.** Let  $X$  be a real-valued random variable and  $\phi$  its characteristic function.

- (1)  $X$  is lattice if and only if there exists  $t \in \mathbb{R}^*$  such that  $|\phi(t)| = 1$ .
- (2)  $X$  has span  $h > 0$  if and only if  $|\phi(2\pi/h)| = 1$  and  $|\phi(t)| < 1$  for  $t \in ]0, 2\pi/h[$ .

*Proof.* (1) If  $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$ , then  $\mathbb{P}((X - b)/h \in \mathbb{Z}) = 1$ , so that

$$\mathbb{E} \left[ e^{2i\pi(X-b)/h} \right] = 1, \quad \left| \mathbb{E} \left[ e^{2i\pi X/h} \right] \right| = |e^{2i\pi b/h}| = 1.$$

Conversely, if  $|\phi(t_0)| = 1$  for  $t_0 \in \mathbb{R}^*$ , let  $a \in \mathbb{R}$  be such that  $\phi(t_0) = e^{ia}$ . Then  $\mathbb{E}[e^{i(t_0 X - a)}] = 1$ . Hence  $\mathbb{E}[\cos(t_0 X - a)] = 1$ . Since  $\cos \leq 1$ , we deduce that, almost surely,  $\cos(t_0 X - a) = 1$ . Thus, almost surely,  $t_0 X - a \in 2\pi\mathbb{Z}$ , and

$$\mathbb{P} \left( X \in \frac{a}{t_0} + \frac{2\pi}{t_0} \mathbb{Z} \right) = 1.$$

(2) The proof of (1) shows that there exist  $b \in \mathbb{R}$  and  $h > 0$  such that  $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$  if and only if  $|\phi(2\pi/h)| = 1$ , which gives the result.  $\square$

**Theorem 6.7 (Local limit theorem).** Let  $(X_i)_{i \geq 1}$  be i.i.d. aperiodic real-valued random variables taking values in  $\mathbb{Z}$ , such that  $a = \mathbb{E}[X_1]$  and  $\sigma^2 = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 \in ]0, \infty[$ . Write  $S_n = X_1 + \dots + X_n$ . Then

$$\sup_{k \in \mathbb{Z}} \left| \sqrt{n} \mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{1}{2} \left( \frac{k - an}{\sigma\sqrt{n}} \right)^2 \right) \right| \rightarrow 0.$$

*Proof.* Set  $X'_i = X_i - a$ ,  $S'_n = S_n - an$ . Write  $\phi(t) = \mathbb{E}[e^{itX'_1}]$ . Since  $\mathbb{E}[e^{itS'_n}] = \sum_{j \in \mathbb{Z} - an} e^{itj} \mathbb{P}(S'_n = j)$ , we obtain, for  $k \in \mathbb{Z} - an$ ,

$$\mathbb{P}(S'_n = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \mathbb{E}[e^{itS'_n}] dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itk} \phi(t)^n dt.$$

Thus, choosing  $u \in \mathbb{R}$  so that  $u\sigma\sqrt{n} \in \mathbb{Z} - an$ ,

$$\sigma\sqrt{n} \mathbb{P}(S'_n = u\sigma\sqrt{n}) = \frac{1}{2\pi} \int_{-\pi\sigma\sqrt{n}}^{\pi\sigma\sqrt{n}} e^{-itu} \phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n dt.$$

But, for every  $u \in \mathbb{R}$ ,

$$\frac{1}{\sqrt{2\pi}} e^{-u^2/2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itu - t^2/2} dt.$$

Set  $q(t) = e^{-t^2/2}$  and

$$g_n(t) = \mathbb{1}_{\{|t| \leq \pi\sigma\sqrt{n}\}} \phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n.$$

The two inversion formulas above give

$$\sup_{k \in \mathbb{Z}} \left| \sigma\sqrt{n} \mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(k - an)^2}{2n\sigma^2}\right) \right| \leq \frac{1}{2\pi} \|g_n - q\|_{L^1(\mathbb{R})}.$$

It therefore suffices to show that  $g_n \rightarrow q$  in  $L^1(\mathbb{R})$ . The zero mean and finite second moment give

$$\phi(t) = 1 - \frac{t^2\sigma^2}{2} + t^2g(t), \quad g(0) = 0, \quad g \text{ continuous.} \quad (*)$$

Indeed, the Taylor remainder satisfies

$$|e^{ix} - 1 - ix + x^2/2| \leq \min(|x|^3/6, x^2).$$

Thus  $|g(t)| \leq \mathbb{E}[\min(|t||X'_1|^3/6, (X'_1)^2)] \rightarrow 0$  by dominated convergence. Moreover,  $(*)$  implies  $|\phi(t)|^2 = 1 - \sigma^2 t^2 + o(t^2)$ . We can therefore choose  $0 < \delta < \pi$  such that

$$|\phi(t)| \leq e^{-\sigma^2 t^2/4} \quad (|t| \leq \delta).$$

Set

$$h_n(t) = \mathbb{1}_{\{|t| \leq \delta\sigma\sqrt{n}\}} \phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n.$$

For every  $t \in \mathbb{R}$ , the expansion  $(*)$  gives  $h_n(t) \rightarrow q(t)$ , while  $|h_n(t)| \leq e^{-t^2/4}$ . By dominated convergence,  $\|h_n - q\|_{L^1(\mathbb{R})} \rightarrow 0$ . Finally, aperiodicity and continuity of  $\phi$  give

$$\rho = \sup_{\delta \leq |t| \leq \pi} |\phi(t)| < 1, \quad \|g_n - h_n\|_{L^1(\mathbb{R})} \leq 2\pi\sigma\sqrt{n} \rho^n \rightarrow 0.$$

The triangle inequality proves  $\|g_n - q\|_{L^1(\mathbb{R})} \rightarrow 0$ . The bound obtained by inversion, divided by  $\sigma$ , completes the proof.  $\square$

**Corollary 6.8.** Under the assumptions of the local limit theorem, assuming  $\mathbb{E}[X_1] = 0$ , we have:

- (1)  $\mathbb{P}(S_n = 0) \sim \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sqrt{n}}$ ;
- (2) there exists  $C > 0$  such that, for all  $n \geq 1$  and  $k \in \mathbb{Z}$ ,  $\mathbb{P}(S_n = k) \leq C/\sqrt{n}$ ;
- (3) if  $(x_n)$  is a sequence of integers such that  $x_n/(\sigma\sqrt{n}) \rightarrow x \in \mathbb{R}$ , then

$$\sqrt{\sigma^2 n} \mathbb{P}(S_n = x_n) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

**Remark 6.9.** One can prove the following refinement, which is useful for large  $|k|$ :

$$\sup_{k \in \mathbb{Z}} \max \left( 1, \left( \frac{k - an}{\sqrt{n}} \right)^2 \right) \left| \sigma\sqrt{n} \mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left( \frac{k - an}{\sigma\sqrt{n}} \right)^2 \right) \right| \rightarrow 0.$$

See Spitzer, *Principles of Random Walk*, Chapter II, Section 7, p. 90.

### 6.3 Convergence to the Brownian bridge

Recall the notation  $p_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$ .

**Theorem 6.10** (Conditional Donsker theorem). Let  $(X_i)_{i \geq 1}$  be i.i.d. aperiodic random variables taking values in  $\mathbb{Z}$ , such that  $\mathbb{E}[X_1] = 0$  and  $\sigma^2 = \mathbb{E}[X_1^2] \in ]0, \infty[$ . Set  $S_n = X_1 + \dots + X_n$  and define  $S_t$  by linear interpolation. Then

$$\left( \frac{S_{nt}}{\sigma\sqrt{n}} : 0 \leq t \leq 1 \right) \text{ under } \mathbb{P}(\cdot | S_n = 0) \xrightarrow{d} (b_t)_{0 \leq t \leq 1}.$$

**Remark 6.11.** Aperiodicity implies that  $\mathbb{P}(S_n = 0) > 0$  for all sufficiently large  $n$  (a consequence of the local limit theorem, for example).

*Proof.* **Step 1: convergence on  $[0, s]$ , for fixed  $0 < s < 1$ .**

Let  $F : \mathcal{C}([0, s], \mathbb{R}) \rightarrow \mathbb{R}$  be bounded continuous. Then

$$\begin{aligned} & \mathbb{E} \left[ F \left( \frac{S_{nt}}{\sigma\sqrt{n}} : 0 \leq t \leq s \right) \middle| S_n = 0 \right] \\ &= \mathbb{E} \left[ F \left( \frac{S_{nt}}{\sigma\sqrt{n}} : 0 \leq t \leq s \right) \frac{Q_{n - \lceil ns \rceil}(-S_{\lceil ns \rceil})}{Q_n(0)} \right], \end{aligned}$$

where  $Q_n(k) = \mathbb{P}(S_n = k)$ . Fix  $s' \in ]s, 1[$ . By Donsker's theorem and the Skorokhod representation theorem, we can construct copies of the paths on  $[0, s']$  such that

$$\sup_{0 \leq t \leq s'} \left| \frac{S_{nt}}{\sigma\sqrt{n}} - B_t \right| \rightarrow 0 \quad \text{almost surely.}$$

We retain the same notation for these copies. Since  $\lceil ns \rceil/n \rightarrow s$  and  $B$  is continuous, in particular we have

$$\frac{S_{\lceil ns \rceil}}{\sigma\sqrt{n}} \rightarrow B_s \quad \text{almost surely.}$$

Using the local limit theorem, we obtain:

(i) there exists  $C > 0$  such that, for all sufficiently large  $n$ ,

$$\frac{Q_{n-\lceil ns \rceil}(-S_{\lceil ns \rceil})}{Q_n(0)} \leq C$$

(using assertions (1) and (2) of Corollary 6.8);

(ii) almost surely,

$$\frac{Q_{n-\lceil ns \rceil}(-S_{\lceil ns \rceil})}{Q_n(0)} \rightarrow \frac{p_1(-B_s/\sqrt{1-s})}{\sqrt{1-s} p_1(0)} = \frac{p_{1-s}(-B_s)}{p_1(0)}$$

(using assertion (3) of Corollary 6.8).

By dominated convergence, we conclude that

$$\begin{aligned} & \mathbb{E} \left[ F \left( \frac{S_{nt}}{\sigma\sqrt{n}} : 0 \leq t \leq s \right) \middle| S_n = 0 \right] \\ & \rightarrow \mathbb{E} \left[ F(B_t : 0 \leq t \leq s) \frac{p_{1-s}(-B_s)}{p_1(0)} \right] = \mathbb{E}[F(b_t : 0 \leq t \leq s)]. \end{aligned}$$

**Step 2: convergence of finite-dimensional distributions on  $[0, 1]$ .** This follows from Step 1 and the fact that both processes equal 0 at  $t = 1$ .

**Step 3: tightness on  $[0, 1]$ .** We use a time-reversal argument: setting  $\hat{S}_i^n = S_n - S_{n-i}$ , we clearly have

$$(S_i)_{0 \leq i \leq n} \stackrel{d}{=} (\hat{S}_i^n)_{0 \leq i \leq n},$$

and this equality in distribution remains true conditional on  $S_n = 0$ . We deduce that, conditional on  $S_n = 0$ ,

$$\begin{aligned} S^{1,n} &:= \left( \frac{S_{nt}}{\sigma\sqrt{n}} : 0 \leq t \leq \frac{3}{4} \right) \xrightarrow{d} \left( b_t : 0 \leq t \leq \frac{3}{4} \right), \\ S^{2,n} &:= \left( \frac{S_{n(1-t)}}{\sigma\sqrt{n}} : 0 \leq t \leq \frac{3}{4} \right) \xrightarrow{d} \left( \tilde{b}_t : 0 \leq t \leq \frac{3}{4} \right), \end{aligned}$$

where  $\tilde{b}$  is a Brownian bridge. These are two separate marginal convergences, which suffice to establish tightness of both sequences. We deduce that  $S^{1,n}$  and  $S^{2,n}$  are tight. Since, for  $\delta < 1/4$ ,

$$\omega \left( \left( \frac{S_{nt}}{\sigma\sqrt{n}} : 0 \leq t \leq 1 \right), \delta \right) = \max(\omega(S^{1,n}, \delta), \omega(S^{2,n}, \delta)),$$

we conclude that  $(S_{nt}/(\sigma\sqrt{n}) : 0 \leq t \leq 1)$  under  $\mathbb{P}(\cdot | S_n = 0)$  is tight.  $\square$

## 7 Discontinuous processes: Skorokhod space

Related exercises: Sheet 6, p. 198.

### In this chapter

|     |                                                        |    |
|-----|--------------------------------------------------------|----|
| 7.1 | Basic properties . . . . .                             | 69 |
| 7.2 | Skorokhod ( $J_1$ ) topology on $\mathbb{D}$ . . . . . | 70 |
| 7.3 | Weak convergence in $\mathbb{D}$ . . . . .             | 73 |
| 7.4 | Tightness criteria . . . . .                           | 75 |
| 7.5 | Optional further material . . . . .                    | 77 |

Write  $\mathbb{D}$  (or  $\mathbb{D}([0, 1], \mathbb{R}^d)$ ) for the set of functions  $x : [0, 1] \rightarrow \mathbb{R}^d$  that

- are right-continuous: for every  $t < 1$ ,  $x(t+) = \lim_{s \downarrow t} x(s) = x(t)$ ;
- have left limits: for every  $0 < t \leq 1$ ,  $x(t-) = \lim_{s \uparrow t} x(s)$  exists.

For  $x \in \mathbb{D}$ , we say that  $x$  is *càdlàg* and write  $\Delta x(t) = x(t) - x(t-)$  for  $0 < t \leq 1$ , and  $\Delta x(0) = 0$ . Càdlàg functions and processes arise naturally in probability:

- distribution functions;
- Poisson processes;
- Lévy processes (càdlàg, with stationary independent increments), which arise in particular in the extension of Donsker's theorem to certain random variables with infinite variance;
- every martingale with respect to a complete right-continuous filtration admits a càdlàg modification.

**Caution.**  $(\mathbb{D}, \|\cdot\|_\infty)$  is a Banach space, but it is not separable. Take  $x^{(u)} = \mathbb{1}_{[u, 1]}$ :  $\|x^{(u)} - x^{(v)}\|_\infty = 1$  for  $u \neq v$ .

We must look for another topology on  $\mathbb{D}$ .

### References.

- Billingsley, *Convergence of Probability Measures*, second edition.
- Jacod–Shiryaev, *Limit Theorems for Stochastic Processes*, Chapter VI.
- Kallenberg, *Foundations of Modern Probability*, second edition.

**Remark 7.1.** We work in  $\mathbb{D}([0, 1], \mathbb{R}^d)$  for simplicity, but what follows extends readily to  $\mathbb{D}([0, 1], S)$ , with  $(S, d)$  Polish.

## 7.1 Basic properties

**Example 7.2.** Let  $(r_i)_{i \geq 1}$  be an enumeration of the rationals in  $]0, 1[$ . We have

$$x(t) = \sum_{i=1}^{\infty} \frac{1}{2^i} \mathbb{1}_{\{r_i \leq t\}} \in \mathbb{D}.$$

**Proposition 7.3.** Let  $x \in \mathbb{D}$ .

- (1) For every  $\varepsilon > 0$ ,  $\#\{t \in [0, 1] : |\Delta x(t)| > \varepsilon\} < \infty$ .
- (2)  $\{t \in [0, 1] : \Delta x(t) \neq 0\}$  is countable.
- (3)  $x$  is bounded.
- (4)  $x$  is a uniform limit of step functions.

*Proof sketch.* (1) If the set is infinite, take a sequence  $(t_n)$  in it. Passing to a subsequence, assume  $t_n \rightarrow t$ . Consider the two cases  $t_n \downarrow t$  and  $t_n \uparrow t$  to obtain a contradiction.

(2) This follows from (1).

(3) By contradiction: if  $|x(t_n)| > n$ , extract a subsequence  $t_n \rightarrow t$  with  $(t_n)$  monotone, and obtain  $\max(|x(t)|, |x(t-)|) = \infty$ , a contradiction.

(4) Introduce  $\omega(x, A) = \sup_{s, t \in A} |x(t) - x(s)|$  for  $A \subset [0, 1]$ . We then use the following lemma. □

**Lemma 7.4.** For every  $\varepsilon > 0$ , there exists  $0 = t_0 < t_1 < \dots < t_r = 1$  such that

$$\max_{0 \leq i \leq r-1} \omega(x, [t_i, t_{i+1}[) \leq \varepsilon.$$

*Proof sketch.* Let  $\varepsilon > 0$ . Write

$$T = \sup \left\{ t \in [0, 1] : \exists 0 = t_0 < t_1 < \dots < t_r = t, \max_{0 \leq i \leq r-1} \omega(x, [t_i, t_{i+1}[) \leq \varepsilon \right\}.$$

Then  $x(0) = x(0+) \implies T > 0$ ; existence of  $x(T-)$  implies that the supremum is attained; and  $x(T) = x(T+)$  implies that  $T < 1$  is impossible. □

We conclude the proof of (4) by taking

$$x_n = \sum_{i=0}^{r-1} x(t_i) \mathbb{1}_{[t_i, t_{i+1}[} \quad \text{on } [0, 1[, \quad x_n(1) = x(1),$$

with a partition associated with  $\varepsilon \downarrow 0$ . We now define the “modified modulus of continuity”.

**Notation.** For  $x : [0, 1] \rightarrow \mathbb{R}^d$  and  $0 < \delta < 1$ , set

$$\omega'(x, \delta) = \inf_{r \geq 1, 0=t_0 < t_1 < \dots < t_r=1, t_{i+1}-t_i > \delta} \max_{0 \leq i \leq r-1} \omega(x, [t_i, t_{i+1}[).$$

**Example 7.5.** Let  $0 < \varepsilon < 1$  and  $x_n = n\mathbb{1}_{[1-\varepsilon,1]}$ . For  $0 < \delta < 1$ , we have

$$\omega'(x_n, \delta) = \begin{cases} 0, & \delta < \min(\varepsilon, 1 - \varepsilon), \\ n, & \delta \geq \min(\varepsilon, 1 - \varepsilon). \end{cases}$$

**Proposition 7.6.** Let  $x, y : [0, 1] \rightarrow \mathbb{R}^d$ .

- (1)  $x \in \mathbb{D}$  if and only if  $\omega'(x, \delta) \rightarrow 0$  as  $\delta \rightarrow 0$  (compare with:  $x$  is continuous if and only if  $\omega(x, \delta) \rightarrow 0$ ).
- (2)  $\omega'(x, \delta) \leq 2\|x - y\|_\infty + \omega'(y, \delta)$ .
- (3) If  $x \in \mathbb{D}$  and  $0 < \delta < 1$ , then

- (a)  $\omega'(x, \delta) \leq \omega(x, 2\delta)$ ,
- (b)  $\omega(x, \delta) \leq 2\omega'(x, \delta) + \sup_{0 \leq t \leq 1} |\Delta x(t)|$ .

The proof is left as an exercise (we have already proved the forward implication in (1)); see Billingsley, p. 123.

## 7.2 Skorokhod ( $J_1$ ) topology on $\mathbb{D}$

Two functions  $x$  and  $y$  are close in the uniform norm if the graph of  $x(t)$  can be transformed into that of  $y(t)$  by a small perturbation of the vertical coordinates, keeping the horizontal coordinates fixed. In  $\mathbb{D}$ , we will also allow a small time deformation.

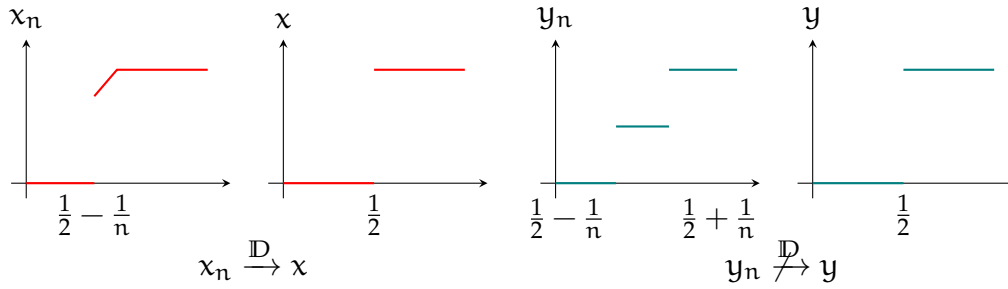
**Definition 7.7.** A *time change* is a function  $\lambda : [0, 1] \rightarrow [0, 1]$  such that  $\lambda(0) = 0$ ,  $\lambda(1) = 1$ , and  $\lambda$  is strictly increasing and continuous. Write  $\Lambda$  for the set of time changes.

**Definition 7.8.** Let  $x_n \in \mathbb{D}$  and  $x \in \mathbb{D}$ . We say that  $x_n \rightarrow x$  in the Skorokhod topology if, for every  $n \geq 1$ , there exists  $\lambda_n \in \Lambda$  such that

$$\|\lambda_n - \text{Id}\|_\infty \rightarrow 0, \quad \|x_n \circ \lambda_n - x\|_\infty \rightarrow 0.$$

Equivalently, we can require  $\|x_n - x \circ \lambda_n\|_\infty \rightarrow 0$ . We write  $x_n \xrightarrow{\mathbb{D}} x$ .

**Example 7.9.** The first graphs illustrate convergence in  $\mathbb{D}$ : a slightly shifted jump, whose size tends to 1, is followed by a small continuous variation. The second graphs illustrate the failure of convergence in  $\mathbb{D}$  when two jumps of size  $1/2$  approach each other to form a jump of size 1.



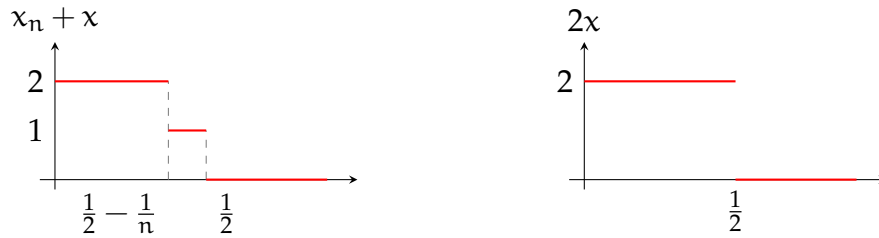
**Remark 7.10.** For  $\lambda_n, \lambda \in \Lambda$ ,  $\|\lambda_n - \lambda\|_\infty \rightarrow 0$  if and only if  $\lambda_n$  converges pointwise to  $\lambda$  (Dini's theorem).

**Remark 7.11.** If we set, for  $x, y \in \mathbb{D}$ ,

$$d_S(x, y) = \inf_{\lambda \in \Lambda} \max(\|\lambda - \text{Id}\|_\infty, \|x \circ \lambda - y\|_\infty),$$

one checks that  $d_S$  is a metric on  $\mathbb{D}$  that metrizes the preceding notion of convergence, and that  $(\mathbb{D}, d_S)$  is a separable metric space.

**Caution.**  $(\mathbb{D}, d_S)$  does not carry a normed vector space topology: if  $x_n = \mathbb{1}_{[0, 1/2 - 1/n[}$  and  $x = \mathbb{1}_{[0, 1/2[}$ , we have  $x_n \xrightarrow{\mathbb{D}} x$ , but  $x_n + x \not\xrightarrow{\mathbb{D}} 2x$ .



However, one checks that, if  $x_n \xrightarrow{\mathbb{D}} x$  and  $y_n \xrightarrow{\mathbb{D}} y$  with  $y$  continuous, then  $x_n + y_n \xrightarrow{\mathbb{D}} x + y$ .

For  $t \in [0, 1]$ , write  $\Pi_t : \mathbb{D} \rightarrow \mathbb{R}^d$ ,  $x \mapsto x(t)$ .

**Proposition 7.12** (Examples of continuous functionals). (1)  $x \mapsto \|x\|_\infty$  is continuous for  $d_S$ .

(2)  $x \mapsto \sup_{0 \leq t \leq 1} |\Delta x(t)|$  is continuous for  $d_S$ .

(3)  $\Pi_0$  and  $\Pi_1$  are continuous.

(4) Let  $t \in ]0, 1[$ .  $\Pi_t$  is continuous at every function  $x$  that is continuous at  $t$ .

(5) Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  be continuous.  $x \mapsto \int_0^1 f(x(s)) ds$  is continuous for  $d_S$ .

*Proof.* Let  $x_n \xrightarrow{\mathbb{D}} x$  and  $\lambda_n \in \Lambda$  be such that  $x_n \circ \lambda_n \rightarrow x$  uniformly and  $\lambda_n \rightarrow \text{Id}$  uniformly.

(1) We have

$$|\|x_n\|_\infty - \|x\|_\infty| = |\|x_n \circ \lambda_n\|_\infty - \|x\|_\infty| \leq \|x_n \circ \lambda_n - x\|_\infty \rightarrow 0.$$

(2) We have  $|\Delta x(t)| = |x(t) - x(t-)| \leq 2\|x\|_\infty$ , and since

$$\sup_{0 \leq t \leq 1} |\Delta x_n(t)| = \sup_{0 \leq t \leq 1} |\Delta(x_n \circ \lambda_n)(t)|,$$

we obtain

$$\begin{aligned} \left| \sup_{0 \leq t \leq 1} |\Delta x_n(t)| - \sup_{t \in [0,1]} |\Delta x(t)| \right| &\leq \sup_{t \in [0,1]} |\Delta(x_n \circ \lambda_n - x)(t)| \\ &\leq 2\|x_n \circ \lambda_n - x\|_\infty \rightarrow 0. \end{aligned}$$

(3) This follows from the fact that, for every  $\lambda \in \Lambda$ , we have  $\lambda(0) = 0$  and  $\lambda(1) = 1$ .

(4) We write

$$|x_n(t) - x(t)| \leq \|x_n - x \circ \lambda_n^{-1}\|_\infty + |x \circ \lambda_n^{-1}(t) - x(t)| \rightarrow 0$$

by continuity of  $x$  at  $t$ .

(5) Let  $x_n \xrightarrow{\mathbb{D}} x$ . Since  $\sup_n \|x_n\|_\infty \vee \|x\|_\infty < \infty$ , we can assume  $f$  is bounded. Set  $D_x = \{t \in [0, 1] : \Delta x(t) \neq 0\}$ , which is countable. Then

$$\int_a^b f(x_n(t)) \, dt = \int_a^b \underbrace{f(x_n(t)) \mathbb{1}_{\{t \notin D_x\}}}_{\rightarrow f(x(t)) \mathbb{1}_{\{t \notin D_x\}}} \, dt.$$

We conclude by dominated convergence. □

**Remark 7.13.** In general, the functions  $f_1 : x \mapsto x(t)$ ,  $f_2 : x \mapsto \omega'(x, \delta)$  and

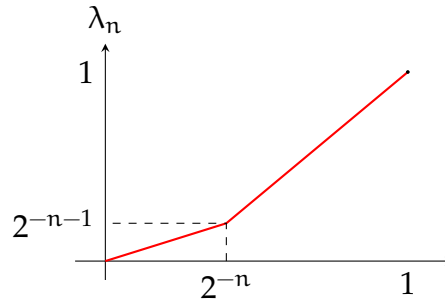
$$f_3 : x \mapsto \sup_{a \leq t \leq b} |\Delta x(t)| \quad (0 \leq a \leq b \leq 1)$$

are not continuous (although  $x \mapsto \omega(x, \delta)$  is continuous for  $x \in \mathcal{C}([0, 1])$ ). For example, fix  $0 < \delta < 1/2$  and consider, for  $n$  sufficiently large that  $\delta + 2^{-n} < 1 - \delta$ ,  $x_n = \mathbb{1}_{[0, \delta + 2^{-n}[}$  and  $x = \mathbb{1}_{[0, \delta[}$ . Then  $x_n \xrightarrow{\mathbb{D}} x$ , but

$$\begin{aligned} x_n(\delta) = 1 &\not\rightarrow x(\delta) = 0, \\ \omega'(x_n, \delta) = 0 &\not\rightarrow \omega'(x, \delta) = 1, \\ x_n(\delta) - x_n(\delta-) = 0 &\not\rightarrow x(\delta) - x(\delta-) = -1. \end{aligned}$$

However, one can show that  $f_2$  and  $f_3$  are upper semicontinuous.

**Example 7.14.** Let  $x_n = \mathbb{1}_{[0, 2^{-n}[}$ . Considering the time change  $\lambda_n$  depicted below, we see that  $d_S(x_n, x_{n+1}) \leq 2^{-n}$ .



In particular,  $(x_n)$  is Cauchy but does not converge (the only possible limit is the zero function). Thus,  $(\mathbb{D}, d_S)$  is not complete.

However:

**Theorem 7.15.** There exists a metric  $d'_S$  on  $\mathbb{D}$  such that the topologies induced by  $d_S$  and  $d'_S$  coincide, and  $(\mathbb{D}, d'_S)$  is complete.

See Billingsley, Theorem 12.2.

### 7.3 Weak convergence in $\mathbb{D}$

**Proposition 7.16.** The product sigma-algebra on  $\mathbb{D}$  (the smallest sigma-algebra making the maps  $\Pi_t : x \mapsto x(t)$  measurable) is the Borel sigma-algebra of  $(\mathbb{D}, d_S)$ .

For  $t < 1$  and integers  $p$  sufficiently large that  $t + 1/p \leq 1$ , the map

$$\Pi_t^p : x \mapsto p \int_t^{t+1/p} x(s) \, ds$$

is continuous, hence measurable. Since  $\Pi_t = \lim_{p \rightarrow \infty} \Pi_t^p$  by right-continuity,  $\Pi_t$  is measurable. Finally,  $\Pi_1$  is continuous. For the reverse inclusion, write  $\mathcal{T}$  for the product sigma-algebra and set

$$A_n x(t) = x(k2^{-n}) \quad \text{if } k2^{-n} \leq t < (k+1)2^{-n}, \quad A_n x(1) = x(1).$$

The map  $A_n : (\mathbb{D}, \mathcal{T}) \rightarrow (\mathbb{D}, \mathcal{B}(\mathbb{D}))$  is measurable: it depends on only finitely many coordinates, and the map assigning the corresponding step function to these values is continuous for the uniform norm, hence for the  $J_1$  topology.

Let us check that  $A_n x \rightarrow x$  in  $\mathbb{D}$  for every  $x \in \mathbb{D}$ . Fix  $\varepsilon > 0$  and choose a partition  $0 = t_0 < \dots < t_r = 1$  such that  $\omega(x, [t_i, t_{i+1}[) \leq \varepsilon$  for every  $i$ . For all sufficiently large  $n$ , the points  $s_i^{(n)} = 2^{-n} \lceil 2^n t_i \rceil$  are distinct. Let  $\lambda_n$  be the piecewise affine time change that maps each  $t_i$  to  $s_i^{(n)}$ . Then  $\|\lambda_n - \text{Id}\|_\infty \leq 2^{-n}$ . If  $t \in [t_i, t_{i+1}[$ , the grid coordinate used by  $A_n x(\lambda_n(t))$  still belongs to  $[t_i, t_{i+1}[$ ; thus

$$\|A_n x \circ \lambda_n - x\|_\infty \leq \varepsilon.$$

At time 1, we also have  $A_n x(\lambda_n(1)) = A_n x(1) = x(1)$ . We deduce that  $d_S(A_n x, x) \rightarrow 0$ . The identity map from  $(\mathbb{D}, \mathcal{T})$  to  $(\mathbb{D}, \mathcal{B}(\mathbb{D}))$  is therefore a pointwise limit of measurable maps taking values in a metric space, which gives  $\mathcal{B}(\mathbb{D}) \subset \mathcal{T}$ .

**Remark 7.17.** The product sigma-algebra is also the smallest sigma-algebra making the maps  $(\Pi_t)_{t \in D}$  measurable, where  $D \subset [0, 1]$  is dense and contains 1. Indeed,

$$\Pi_t = \lim_{\substack{s \in D \\ s \downarrow t}} \Pi_s \quad \text{pointwise.}$$

Thus, finite-dimensional distributions characterize the measures in  $\mathcal{M}_1(\mathbb{D})$ . However:

**Caution.** Convergence in distribution does not imply convergence of finite-dimensional distributions (the maps  $\Pi_t$  are not continuous).

**Definition 7.18.** Let  $X$  be a random variable in  $\mathbb{D}$ . We say that  $t \in [0, 1]$  is a *fixed discontinuity* of  $X$  if  $\mathbb{P}(|\Delta X_t| > 0) > 0$ . Write  $D_X$  for the set of these points.

**Proposition 7.19.** If  $X$  is a random variable in  $\mathbb{D}$ ,  $D_X$  is at most countable.

*Proof.* For  $\varepsilon, \delta > 0$ , set

$$D_X^{\varepsilon, \delta} = \{t \in [0, 1] : \mathbb{P}(|\Delta X_t| > \varepsilon) \geq \delta\}.$$

If this set were infinite, it would contain a sequence of distinct deterministic times  $(t_n)$ . Each càdlàg path has only finitely many jumps of size greater than  $\varepsilon$ ; thus

$$\mathbb{1}_{\{|\Delta X_{t_n}| > \varepsilon\}} \longrightarrow 0 \quad \text{almost surely.}$$

By dominated convergence,  $\mathbb{P}(|\Delta X_{t_n}| > \varepsilon) \rightarrow 0$ , contradicting the lower bound  $\delta$ . All the sets  $D_X^{\varepsilon, \delta}$  are therefore finite, and

$$D_X = \bigcup_{p, q \geq 1} D_X^{1/p, 1/q}$$

is at most countable. □

**Theorem 7.20.** Let  $X^n, X$  be random variables taking values in  $\mathbb{D}$ . Then  $X^n \xrightarrow{d} X$  if and only if:

(i)  $X^n \xrightarrow{\text{f.d.}} X$  outside  $D_X \setminus \{1\}$ , that is, for all  $t_1, \dots, t_k \in ([0, 1] \setminus D_X) \cup \{1\}$ ,

$$(X_{t_1}^n, \dots, X_{t_k}^n) \xrightarrow{d} (X_{t_1}, \dots, X_{t_k});$$

(ii)  $(X^n)$  is tight in  $\mathcal{M}_1(\mathbb{D})$ .

*Proof.* The forward implication for (ii) follows from Prokhorov's theorem. For (i), if  $t \notin D_X$ , then  $\Pi_t$  is continuous  $\mathbb{P}_X$ -almost everywhere. The map  $\Pi_1$ , in turn, is continuous on all of  $\mathbb{D}$ , even when  $1 \in D_X$ . The continuous mapping theorem therefore gives the claimed finite-dimensional convergences.

Conversely, by Prokhorov's theorem, it suffices to show that if  $X^{\varphi(n)} \xrightarrow{d} Y$ , then

$X \stackrel{d}{=} Y$ . To this end, by (i),  $X$  and  $Y$  have the same finite-dimensional distributions on  $([0, 1] \setminus (D_X \cup D_Y)) \cup \{1\}$ , and therefore have the same law.  $\square$

## 7.4 Tightness criteria

The analogue of the Arzelà–Ascoli theorem in  $\mathbb{D}$  is the following.

**Theorem 7.21.** A family  $\mathcal{F} \subset \mathbb{D}$  is relatively compact if and only if

- (1)  $\sup_{x \in \mathcal{F}} \|x\|_\infty < \infty$ ;
- (2)  $\forall \varepsilon > 0, \exists \delta > 0, \sup_{x \in \mathcal{F}} \omega'(x, \delta) \leq \varepsilon$ .

As for continuous functions, we deduce:

**Theorem 7.22.** A sequence  $(X^n)_{n \geq 1}$  of random variables in  $\mathbb{D}$  is tight if and only if

- (1)  $(\|X^n\|_\infty)_{n \geq 1}$  is tight in  $\mathbb{R}$ ;
- (2)  $\forall \varepsilon, \eta > 0, \exists \delta > 0, \limsup_{n \rightarrow \infty} \mathbb{P}(\omega'(X^n, \delta) > \eta) \leq \varepsilon$ .

**Remark 7.23.** Condition (1) cannot be replaced by “ $(X^n(0))_{n \geq 1}$  is tight”, as in  $\mathcal{C}([0, 1], \mathbb{R})$ : take, for example,  $X^n = n\mathbb{1}_{[U, 1]}$ , with  $U$  uniform on  $[0, 1]$ .

However, one can prove that, when (2) holds,  $(1) \iff (1') \iff (1'')$ , where

(1') There exists a dense set  $T \subset [0, 1]$  containing 1 such that, for every  $t \in T$ ,  $(X^n(t))_{n \geq 1}$  is tight.

(1'') The sequences  $(X^n(0))_{n \geq 1}$  and  $(\sup_{0 \leq t \leq 1} |\Delta X^n(t)|)_{n \geq 1}$  are tight.

See Billingsley, second edition, corollary to Theorem 13.2, p. 140.

**Application.** Let  $(U_i)_{i \geq 1}$  be i.i.d. random variables uniform on  $[0, 1]$  and let  $\lambda > 0$ . For  $t \leq 1$ , set

$$X_t^{(n)} = \sum_{k=1}^{\lfloor nt \rfloor} \mathbb{1}_{\{U_k \leq \lambda/n\}}.$$

Then  $(X_t^{(n)})_{0 \leq t \leq 1} \xrightarrow{d} (N_t^\lambda)_{0 \leq t \leq 1}$ , where  $N^\lambda$  is a Poisson process with intensity  $\lambda$ .

*Proof sketch.* It suffices to consider integers  $n \geq \lambda$ . *Finite-dimensional distributions.*  $N^\lambda$  has no fixed discontinuities. Let  $0 = t_0 < t_1 < \dots < t_k \leq 1$ . The increments  $X_{t_{i+1}}^{(n)} - X_{t_i}^{(n)}$ ,  $0 \leq i < k$ , are independent for fixed  $n$  and have distributions

$$\text{Bin}\left(\lfloor nt_{i+1} \rfloor - \lfloor nt_i \rfloor, \frac{\lambda}{n}\right) \xrightarrow{d} \text{Poi}(\lambda(t_{i+1} - t_i)).$$

*Tightness.* First,  $\|X^{(n)}\|_\infty \stackrel{d}{=} \text{Bin}(n, \lambda/n)$  converges in distribution, hence is tight.

For  $0 < \delta < 1/2$ , write  $A_n(\delta)$  for the event that two distinct jumps of  $X^{(n)}$  are at most  $\delta$  apart, and  $B_n(\delta)$  for the event that a jump occurs in  $]0, \delta] \cup [1 - \delta, 1]$ . On the complement of  $A_n(\delta) \cup B_n(\delta)$ , the partition formed by 0, the jump times and 1 has all its intervals of length strictly greater than  $\delta$ . Consequently,  $\omega'(X^{(n)}, \delta) = 0$ .

For  $n \geq \lambda$ , independence of the indicators gives

$$\mathbb{P}(A_n(\delta)) \leq \sum_{\substack{1 \leq i < j \leq n \\ j-i \leq n\delta}} \mathbb{P}(U_i \leq \lambda/n, U_j \leq \lambda/n) \leq n \lfloor n\delta \rfloor \frac{\lambda^2}{n^2} \leq \lambda^2 \delta.$$

There are at most  $2n\delta + 2$  grid points in  $]0, \delta] \cup [1 - \delta, 1]$ , so

$$\mathbb{P}(B_n(\delta)) \leq 2\lambda\delta + \frac{2\lambda}{n}.$$

Thus, for every  $\eta > 0$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\omega'(X^{(n)}, \delta) > \eta) \leq (\lambda^2 + 2\lambda)\delta \xrightarrow[\delta \downarrow 0]{} 0,$$

which completes the proof of tightness.  $\square$

We conclude with several other tightness criteria. The first is an analogue of Kolmogorov's criterion adapted to càdlàg paths (see Billingsley, second edition, Section 13, in particular the proof of Theorem 13.5).

**Theorem 7.24** (Billingsley's criterion). Let  $(X^n)_{n \geq 1}$  be a sequence of random variables in  $\mathbb{D}$ . Assume that:

- (1)  $(\|X^n\|_\infty)_{n \geq 1}$  is tight;
- (2) there exist  $\alpha, b > 0$  and a continuous nondecreasing function  $F : [0, 1] \rightarrow \mathbb{R}_+$  such that, for every  $n \geq 1$  and all  $0 \leq r \leq s \leq t \leq 1$ ,

$$\mathbb{E}[|X^n(t) - X^n(s)|^\alpha |X^n(s) - X^n(r)|^\alpha] \leq (F(t) - F(r))^{1+b};$$

- (3) for every  $\varepsilon > 0$ ,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} (\mathbb{P}(|X^n(\delta) - X^n(0)| \geq \varepsilon) + \mathbb{P}(|X^n(1) - X^n(1 - \delta)| \geq \varepsilon)) = 0.$$

Then  $(X^n)_{n \geq 1}$  is tight.

**Remark 7.25.** Conditions (1) and (2) alone do not suffice: the sequence  $X^n = \mathbb{1}_{[0, 1/n[}$  satisfies them, but is not tight in  $\mathbb{D}$ . However, condition (3) is not necessary for every tight sequence: in particular, it excludes the constant sequence  $X^n = \mathbb{1}_{\{1\}}$ .

Under assumption (2), with  $F$  continuous, condition (3) is satisfied if, for some random

variable  $X$  taking values in  $\mathbb{D}$ , we have  $X^n \xrightarrow{\text{f.d.}} X$  on  $[0, 1] \setminus D_X$  and  $X$  is almost surely continuous at 1. The bound in (2) allows us to compare the boundary increments with those computed at nearby times outside  $D_X$ ; continuity of  $F$  then allows these times to tend to the endpoints.

We conclude with Aldous's criterion, which is particularly useful for random walks or, more generally, martingales.

For  $X^n \in \mathbb{D}$ , write  $(\mathcal{F}_t^n)_{0 \leq t \leq 1}$  for its natural filtration and

$$\mathcal{T}^n = \{\text{stopping times for } \mathcal{F}^n \text{ that are } \leq 1\}.$$

**Theorem 7.26** (Aldous). Assume that

- (1)  $(\|X^n\|_\infty)_{n \geq 1}$  is tight;
- (2) for every sequence  $h_n \downarrow 0$  and every sequence  $\tau_n \in \mathcal{T}^n$ ,

$$\left| X_{(\tau_n + h_n) \wedge 1}^n - X_{\tau_n}^n \right| \xrightarrow{\mathbb{P}} 0.$$

Then  $(X^n)_{n \geq 1}$  is tight.

See, for example, Kallenberg, Theorem 16.11, for a proof.

For example, this theorem can be used to prove the following striking result.

**Theorem 7.27.** Let  $(X_i^n)_{i, n \geq 1}$  be random variables such that, for every  $n \geq 1$ ,  $(X_i^n)_{i \geq 1}$  are i.i.d. Set  $S_k^n = X_1^n + \cdots + X_k^n$  and assume there exists a sequence of integers  $m_n \rightarrow \infty$  such that  $S_{m_n}^n$  converges in distribution to a limit denoted by  $X$ .

Then there exists  $(X_t)_{0 \leq t \leq 1} \in \mathbb{D}$  with stationary independent increments (that is, a Lévy process) such that  $X_1 \stackrel{d}{=} X$  and

$$(S_{[m_n t]}^n : 0 \leq t \leq 1) \xrightarrow{d} (X_t)_{0 \leq t \leq 1}.$$

See Kallenberg, Theorem 16.14.

If  $(X_i)_{i \geq 1}$  are i.i.d., with mean zero and finite variance, taking  $X_i^n = X_i/\sqrt{n}$  and  $m_n = n$  gives essentially Donsker's theorem.

## 7.5 Optional further material

### 7.5.1 The space $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$

Write  $\mathbb{D}_m = \mathbb{D}([0, m], \mathbb{R}^d)$ ,  $m > 0$ , and  $d_m$  for the Skorokhod metric naturally defined on it. Let  $\mathbb{D}_\infty = \mathbb{D}([0, \infty[, \mathbb{R}^d)$ .

**Definition 7.28.** We say that  $x_n \xrightarrow{\mathbb{D}_\infty} x$  if, for every  $n \geq 1$ , there exists  $\lambda_n \in \Lambda_\infty$ , where

$$\Lambda_\infty = \{\text{increasing homeomorphisms from } \mathbb{R}_+ \text{ onto } \mathbb{R}_+\},$$

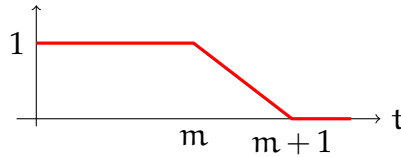
such that

- 1)  $\|\lambda_n - \text{Id}\|_\infty \rightarrow 0$ ;
- 2) for every  $m \geq 1$ ,  $\|(\lambda_n \circ \lambda_n - x)\mathbb{1}_{[0,m]}\|_\infty \rightarrow 0$ .

**Remark 7.29.** The restriction map  $\mathbb{D}_\infty \rightarrow \mathbb{D}_m$ ,  $x \mapsto x\mathbb{1}_{[0,m]}$ , is not continuous in  $\mathbb{D}_\infty$  if  $x$  has a jump at  $m$ .

To define a metric  $d_\infty$  such that  $d_\infty(x_n, x) \rightarrow 0$  if and only if  $x_n \xrightarrow{\mathbb{D}_\infty} x$ , introduce, for every  $m \geq 1$ , the function

$$g_m(t) = \begin{cases} 1, & 0 \leq t \leq m, \\ m+1-t, & m \leq t \leq m+1, \\ 0, & t \geq m+1. \end{cases}$$



For  $x, y \in \mathbb{D}_\infty$ , set

$$d_\infty(x, y) = \sum_{m \geq 1} 2^{-m} (d_{m+1}((g_m x)|_{[0,m+1]}, (g_m y)|_{[0,m+1]}) \wedge 1).$$

This metric metrizes the convergence defined above. The resulting topology is Polish: it admits a compatible complete separable metric.

**Tightness criteria.** For  $m \geq 1$  and  $0 < \delta < m$ , define

$$\omega'_m(x, \delta) = \inf_{\substack{0=t_0 < t_1 < \dots < t_r=m \\ t_{i+1}-t_i > \delta, 0 \leq i \leq r-1}} \max_{0 \leq i \leq r-1} \omega(x, [t_i, t_{i+1}[).$$

A sequence  $(X^n)$  is tight in  $\mathbb{D}_\infty$  if and only if, for each integer  $m \geq 1$ , the sequence  $((g_m X^n)|_{[0,m+1]})$  is tight in  $\mathbb{D}_{m+1}$ . We therefore apply the preceding criteria with  $\|g_m X^n\|_\infty$  and  $\omega'_{m+1}(g_m X^n, \delta)$ . See Billingsley, Section 16.

### 7.5.2 $\mathcal{C}$ -tightness

Note that  $\mathcal{C} = \mathcal{C}([0, 1], \mathbb{R}^d)$  is a closed subset of  $\mathbb{D}$ , since it is

$$\left\{ x \in \mathbb{D} : \sup_{0 \leq t \leq 1} |\Delta x(t)| = 0 \right\}.$$

**Definition 7.30.** A sequence  $(X^n)$  of random variables in  $\mathbb{D}$  is called  $\mathcal{C}$ -tight if it is tight in  $\mathbb{D}$  ( $\mathbb{D}$ -tight) and every subsequential limit is supported on  $\mathcal{C}$ : if  $X^{\varphi(n)} \xrightarrow{d} X$ , then  $\mathbb{P}(X \in \mathcal{C}) = 1$ .

**Proposition 7.31.** Let  $X^n, X$  be random variables in  $\mathbb{D}$  such that  $X^n \xrightarrow{d} X$ . Assume that

$$\sup_{0 \leq t \leq 1} |\Delta X^n(t)| \xrightarrow{\mathbb{P}} 0.$$

Then  $\mathbb{P}(X \in \mathcal{C}) = 1$ .

*Proof.* The map

$$J : \mathbb{D} \longrightarrow \mathbb{R}_+, \quad J(x) = \sup_{0 \leq t \leq 1} |\Delta x(t)|$$

is continuous, as proved in the proposition on continuous functionals. Thus  $J(X^n) \xrightarrow{d} J(X)$ . Now  $J(X^n) \xrightarrow{\mathbb{P}} 0$  by assumption; by uniqueness of the limit in distribution,  $J(X) = 0$  almost surely. Since a càdlàg function without jumps is continuous, we obtain  $\mathbb{P}(X \in \mathcal{C}) = 1$ .  $\square$

See also Billingsley, Theorem 13.4, p. 142.

**Theorem 7.32.** Let  $(X^n)$  be a sequence of random variables in  $\mathbb{D}$ . It is  $\mathcal{C}$ -tight if and only if

- (1)  $(X_0^n)_{n \geq 1}$  is tight in  $\mathbb{R}^d$ ;
- (2) for all  $\varepsilon, \eta > 0$ , there exists  $\delta > 0$  such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\omega(X^n, \delta) > \eta) \leq \varepsilon.$$

See Billingsley, corollary on p. 142.

We readily deduce a càdlàg version of Donsker's theorem: if  $(X_i)_{i \geq 1}$  are i.i.d., with mean zero and variance  $\sigma^2 \in ]0, \infty[$ , writing  $S_n = X_1 + \cdots + X_n$ , we have

$$\left( \frac{S_{[nt]}}{\sigma\sqrt{n}} : 0 \leq t \leq 1 \right) \xrightarrow{d} (B_t : 0 \leq t \leq 1),$$

where convergence is in distribution in  $\mathbb{D}$ .

### 7.5.3 Point processes

For point processes, convergence of finite-dimensional distributions implies convergence in distribution.

**Theorem 7.33.** Write  $\mathbb{D}_c$  for the set of counting functions on  $[0, 1]$ , that is, nondecreasing càdlàg functions taking values in  $\mathbb{Z}$ , whose jumps equal 1 at points of discontinuity. Let  $X^n, X$  be random variables taking values in  $\mathbb{D}_c$ . Assume that  $\mathbb{P}(X(1-) = X(1)) = 1$ .

Then  $X^n \xrightarrow{d} X$  if and only if there exists a countable dense set  $T \subset [0, 1]$  containing 0 and 1, such that, for all  $t_1, \dots, t_k \in T$ ,

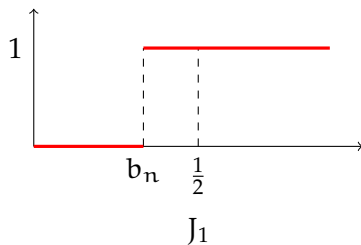
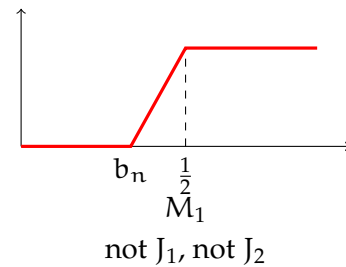
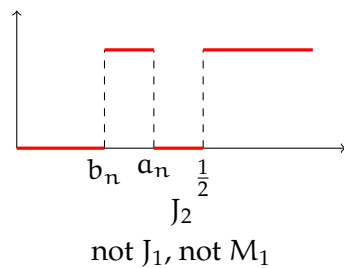
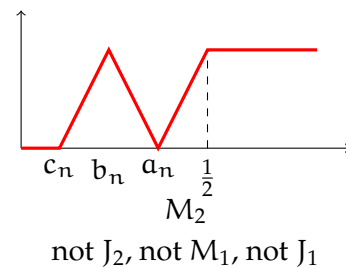
$$(X^n_{t_1}, \dots, X^n_{t_k}) \xrightarrow{d} (X_{t_1}, \dots, X_{t_k}).$$

See Billingsley, Theorem 12.6.

### 7.5.4 Other topologies

We have considered the Skorokhod  $J_1$  topology on  $\mathbb{D}([0, 1], \mathbb{R})$ , but there are three others ( $J_2, M_1, M_2$ ) in which the following convergences hold, with

$$x = \mathbb{1}_{[1/2, 1]}, \quad a_n = \frac{1}{2} - \frac{1}{2n}, \quad b_n = \frac{1}{2} - \frac{1}{n}, \quad c_n = \frac{1}{2} - \frac{3}{2n} \quad (n \geq 4).$$



See W. Whitt, *Stochastic Process Limits*, Chapter 11.5.

## 8 Poisson random measures

Related exercises: Sheet 7, p. 206.

### In this chapter

|     |                                                          |    |
|-----|----------------------------------------------------------|----|
| 8.1 | Review of Poisson distributions . . . . .                | 81 |
| 8.2 | Poisson random measures . . . . .                        | 82 |
| 8.3 | Multiple points . . . . .                                | 87 |
| 8.4 | Campbell's formulas (additive and exponential) . . . . . | 89 |
| 8.5 | Transformations of Poisson random measures . . . . .     | 91 |
| 8.6 | Poisson random measures with a time parameter . . . . .  | 92 |
| 8.7 | Palm's formula . . . . .                                 | 94 |

One objective is to define a notion of “rain” in a fairly general space, where drops arrive independently and rarely according to a distribution described by a measure  $\mu$ . When  $\mu$  is infinite, the expression “a point chosen at random according to  $\mu$ ” has no real meaning. However, we will be able to give meaning to “infinitely many points chosen according to  $\mu$ ” through Poisson random measures (we will see that Poisson distributions arise naturally).

**References.** Kingman, *Poisson Processes*; Last and Penrose, *Lectures on the Poisson Process*.

### 8.1 Review of Poisson distributions

For  $\lambda \in [0, \infty]$ , a random variable  $X$  taking values in  $\mathbb{N} \cup \{+\infty\}$  is said to have a *Poisson distribution with parameter  $\lambda$* , written  $X \sim \mathcal{P}(\lambda)$ , if

- for  $\lambda = 0$ ,  $X = 0$  almost surely;
- for  $\lambda = \infty$ ,  $X = \infty$  almost surely;
- for  $0 < \lambda < \infty$ ,  $\mathbb{P}(X = k) = e^{-\lambda} \lambda^k / k!$ , for  $k \geq 0$ .

We have  $\mathbb{E}[e^{-uX}] = \exp(-\lambda(1 - e^{-u}))$  for  $u > 0$  (with the convention  $0 \times \infty = 0$ ). The following two properties will be very useful.

**Proposition 8.1.** • *Superposition.* If  $(\lambda_i)_{i \in I}$  is an at most countable family in  $[0, \infty]$  and  $(X_i)_{i \in I}$  are independent random variables with  $X_i \sim \mathcal{P}(\lambda_i)$ , then

$$\sum_{i \in I} X_i \sim \mathcal{P}\left(\sum_{i \in I} \lambda_i\right).$$

- *Thinning.* Let  $X \sim \mathcal{P}(\lambda)$  with  $\lambda \in [0, \infty]$ . Let  $(U_k)_{k \geq 1}$  be i.i.d. random variables taking values in  $\mathbb{N}$ , independent of  $X$ , with  $\mathbb{P}(U_k = i) = p_i$  for  $i \geq 1$ . Then the variables

$$X_i = \sum_{k=1}^X \mathbb{1}_{\{U_k=i\}}, \quad i \geq 1,$$

are independent and  $X_i \sim \mathcal{P}(\lambda p_i)$  (with  $0 \times \infty = 0$ ).

The proof, which can be carried out using characteristic functions, is left as an exercise.

## 8.2 Poisson random measures

We work in the fairly general setting of a measurable space  $(E, \mathcal{E})$ .

**Definition 8.2.** Write  $\mathcal{M}_s(E)$  for the set of measures  $\mu$  on  $E$  that can be written as  $\mu = \sum_{i \in I} \mu_i$ , where  $I$  is countable and each  $\mu_i$  is a finite measure. The elements of  $\mathcal{M}_s(E)$  are called *s-finite measures*.

**Example 8.3.** A  $\sigma$ -finite measure is s-finite, but the converse is false (take  $\mu = \sum_{k=1}^{\infty} \delta_1$ ). The advantage of working with  $\mathcal{M}_s(E)$  is closure under countable sums (which fails for  $\sigma$ -finite measures).

**Remark 8.4.** One can check that Fubini's theorems apply to s-finite measures using the monotone convergence theorem (see Theorem A.13 in Last and Penrose).

**Definition 8.5.** Write  $\mathcal{M}_{\text{atom}}(E)$  for the set of countable sums of atoms of the form  $\sum_{i \in I} \delta_{x_i}$ , where  $I$  is countable and  $(x_i)_{i \in I} \in E^I$ . Equip  $\mathcal{M}_{\text{atom}}(E)$  with the smallest sigma-algebra making the following maps measurable:

$$\mathcal{M}_{\text{atom}}(E) \longrightarrow \mathbb{N} \cup \{+\infty\}, \quad \mu \longmapsto \mu(A), \quad A \in \mathcal{E}.$$

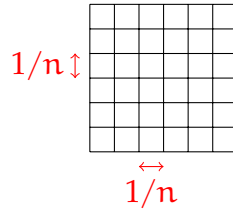
**Example 8.6.** 1)  $N : \Omega \rightarrow \mathcal{M}_{\text{atom}}(E)$  is a random variable if and only if, for every  $A \in \mathcal{E}$  and every  $k \geq 0$ ,  $\{N(A) = k\}$  is measurable, since

$$\{N(A) = \infty\} = \left( \bigcup_{k \geq 0} \{N(A) = k\} \right)^c.$$

2) The map  $E \rightarrow \mathcal{M}_{\text{atom}}(E)$ ,  $x \mapsto \delta_x$ , is measurable.

3) For  $A \in \mathcal{E}$ , the map  $\mathcal{M}_{\text{atom}}(E) \rightarrow \mathcal{M}_{\text{atom}}(E)$ ,  $\mu \mapsto \mu|_A$ , is measurable, where  $\mu|_A(B) = \mu(A \cap B)$  for  $B \in \mathcal{E}$ .

Before defining a Poisson random measure, here is some motivation: to define “uniform” rain on  $\mathbb{R}^2$ , it is natural to discretize  $\mathbb{R}^2$  using a grid of mesh size  $1/n$  and choose a point in each cell, independently, with probability  $1/n^2$ .



Then the number of points falling in a region  $A$  has approximately distribution  $\text{Bin}(n^2 \text{Area}(A), 1/n^2)$ , which converges in distribution to  $\mathcal{P}(\text{Area}(A))$ , and the numbers of points falling in disjoint regions are independent.

**Definition 8.7.** Let  $(E, \mathcal{E})$  be a measurable space and  $\mu$  an s-finite measure on  $E$ . A *Poisson random measure* (or *Poisson point cloud*) with intensity  $\mu$  is a random variable  $N$  taking values in  $\mathcal{M}_{\text{atom}}(E)$  such that

- ① for every  $A \in \mathcal{E}$ ,  $N(A) \sim \mathcal{P}(\mu(A))$ ;
- ② if  $(A_i)_{1 \leq i \leq k}$  is a family of disjoint measurable subsets of  $E$ , the variables  $(N(A_i))_{1 \leq i \leq k}$  are independent.

**Remark 8.8.** • For each fixed  $A$ , almost surely,  $N(A) = \infty \iff \mu(A) = \infty$ .

- If singletons are measurable in  $E$  and  $\mu(\{x\}) > 0$ , then  $\mathbb{P}(N(\{x\}) \geq 2) > 0$ .
- $\mathbb{E}[e^{-\lambda N(A)}] = e^{-\mu(A)(1-e^{-\lambda})}$ .

**Theorem 8.9.** For every measure  $\mu \in \mathcal{M}_s(E)$ , there exists a Poisson random measure with intensity  $\mu$ , unique in law in  $\mathcal{M}_{\text{atom}}(E)$ .

To prove this theorem, we use the following useful lemma.

**Lemma 8.10** (Closure under sums). Let  $(N_i)_{i \geq 1}$  be a countable family of independent Poisson random measures with intensities  $\mu_i$ . Then  $N = \sum_{i \geq 1} N_i$  is a Poisson random measure with intensity  $\mu = \sum_{i \geq 1} \mu_i$ .

*Proof.* Let  $A \in \mathcal{E}$ . By the superposition principle for Poisson distributions,

$$N(A) \sim \mathcal{P}\left(\sum_{i \geq 1} \mu_i(A)\right) = \mathcal{P}(\mu(A)).$$

If  $(A_j)_{1 \leq j \leq k}$  are disjoint measurable subsets, the variables  $(N_i(A_j))_{i \geq 1, 1 \leq j \leq k}$  are independent. Grouping them into blocks, the variables  $(\sum_{i=1}^{\infty} N_i(A_j))_{1 \leq j \leq k}$  are independent.  $\square$

*Proof of existence of Poisson random measures.* By the preceding lemma, it suffices to consider the case where  $\mu$  is finite. If  $\mu(E) = 0$ , the empty point cloud  $M = 0$  works. Thus assume  $0 < \mu(E) < \infty$ . Let  $L \sim \mathcal{P}(\mu(E))$  and let  $(X_i)_{i \geq 1}$  be i.i.d. random variables with law  $\mu(\cdot)/\mu(E)$ , independent of  $L$ . Set

$$M = \sum_{i=1}^L \delta_{X_i}.$$

Let us check that  $M$  works. It is indeed a random variable (see the example on measurability of  $x \mapsto \delta_x$ ).

Let  $(A_j)_{1 \leq j \leq k}$  be disjoint measurable subsets of  $E$  and set  $B = (\bigcup_{j=1}^k A_j)^c$ . For  $i \geq 1$ , define

$$U_i = \begin{cases} j & \text{if } X_i \in A_j, \\ 0 & \text{if } X_i \in B. \end{cases}$$

The variables  $(U_i)_{i \geq 1}$  are i.i.d. By the thinning principle for Poisson distributions, the variables

$$M(A_j) = \sum_{i=1}^L \mathbb{1}_{\{X_i \in A_j\}} = \sum_{i=1}^L \mathbb{1}_{\{U_i=j\}}$$

are independent and  $M(A_j) \sim \mathcal{P}(\mu(E)\mu(A_j)/\mu(E)) = \mathcal{P}(\mu(A_j))$ .  $\square$

*Proof of uniqueness in law of Poisson random measures.* Sets of the form

$$\{\nu \in \mathcal{M}_{\text{atom}}(E) : \nu(A_1) = k_1, \dots, \nu(A_\ell) = k_\ell\},$$

with measurable  $A_i$  and  $k_i \geq 0$ , form a  $\pi$ -system generating the sigma-algebra on  $\mathcal{M}_{\text{atom}}(E)$ . Any such set can be written as a finite disjoint union of sets of the form

$$\{\nu \in \mathcal{M}_{\text{atom}}(E) : \nu(B_1) = j_1, \dots, \nu(B_n) = j_n\}, \quad (7)$$

with disjoint measurable  $(B_i)$  and  $j_i \geq 0$ . Thus a probability measure on  $\mathcal{M}_{\text{atom}}(E)$  is characterized by its values on sets of the form (7).

If two Poisson random measures satisfy the two defining conditions with the same intensity, their laws agree on the sets (7), and therefore coincide.  $\square$

**Remark 8.11.** The preceding proof shows that, if  $\mu$  is  $s$ -finite, we can construct a Poisson random measure of the form

$$M = \sum_{i=1}^K \delta_{X_i}, \quad (8)$$

where  $K$  takes values in  $\mathbb{N} \cup \{+\infty\}$  and the  $X_i$  are random variables taking values in  $E$ . When  $\mu(E) = \infty$ , we have  $K = \infty$  almost surely. In general, the positions  $X_i$  are neither independent nor identically distributed.

If  $M$  is a Poisson random measure, in general it cannot be written in the form (8),

since in general the atoms cannot be enumerated measurably.

Indeed, let  $E = [0, 1]$  and suppose there exist measurable maps  $\pi_i : \mathcal{M}_{\text{atom}}(E) \rightarrow E$  and  $L : \mathcal{M}_{\text{atom}}(E) \rightarrow \mathbb{N} \cup \{+\infty\}$  such that

$$\forall \nu \in \mathcal{M}_{\text{atom}}(E), \quad \nu = \sum_{i=1}^{L(\nu)} \delta_{\pi_i(\nu)}.$$

Let  $\lambda$  be Lebesgue measure on  $[0, 1]$ ,  $\mu = \sum_{n \geq 1} \lambda$  and  $M$  a Poisson random measure with intensity  $\mu$ . For each fixed Borel set  $A$ ,  $M(A)$  is almost surely equal to 0 if  $\lambda(A) = 0$  and to  $\infty$  otherwise. The sigma-algebra  $\sigma(M)$  therefore consists of events with probability 0 or 1: its generators have this property, and events of probability 0 or 1 form a sigma-algebra.

Since  $M(E) = \infty$  almost surely, we have  $L(M) = \infty$  almost surely. The variable  $Z = \pi_1(M) \in [0, 1]$  is independent of itself, so

$$\text{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = 0.$$

Thus  $Z = c$  almost surely for some deterministic  $c \in [0, 1]$ . The enumeration then forces  $M(\{c\}) \geq 1$  almost surely, whereas  $\mu(\{c\}) = 0$  forces  $M(\{c\}) = 0$  almost surely. This is a contradiction: additional “finiteness” assumptions are therefore necessary.

If  $E$  is a measurable subset of a Polish metric space, writing  $\mathcal{M}_{\text{atom}}^{<\infty}(E)$  for the finite measures in  $\mathcal{M}_{\text{atom}}(E)$ , the atoms of measures in  $\mathcal{M}_{\text{atom}}^{<\infty}(E)$  can be enumerated measurably (Proposition 6.2 in Last and Penrose). We deduce that, if  $M$  is a Poisson random measure with  $\sigma$ -finite intensity  $\mu$  on a Polish space  $E$ , then we can write

$$M = \sum_{i=1}^{L(M)} \delta_{\pi_i(M)},$$

with  $L$  and the  $\pi_i$  measurable (Corollary 6.5 in Last and Penrose).

**Important consequence.** To prove a property that depends only on the law of  $M$ , we may assume without loss of generality that  $M$  has the form (8), constructed as in the existence proof for Poisson random measures.

**Example 8.12** (A Poisson process is a Poisson random measure). Let  $\lambda > 0$  and let  $(X_i)_{i \geq 1}$  be i.i.d. random variables with distribution  $\text{Exp}(\lambda)$ . Set

$$M = \sum_{k=1}^{\infty} \delta_{X_1 + \dots + X_k}.$$

Then  $M$  is a Poisson random measure with intensity  $\lambda \mathbb{1}_{\{x \geq 0\}} dx$ .

*Proof.* Set  $S_n = X_1 + \dots + X_n$ . The strong law gives  $S_n/n \rightarrow 1/\lambda$  almost surely, hence  $S_n \rightarrow \infty$ . Fix  $T > 0$  and set  $N_T = M([0, T])$ . For  $n \geq 1$ , the change of variables

$s_i = x_1 + \dots + x_i$ , followed by independence of  $X_{n+1}$ , gives

$$\begin{aligned} \mathbb{P}(N_T = n, S_1 \in ds_1, \dots, S_n \in ds_n) \\ &= \lambda^n e^{-\lambda s_n} e^{-\lambda(T-s_n)} \mathbb{1}_{\{0 < s_1 < \dots < s_n < T\}} ds_1 \cdots ds_n \\ &= \left( e^{-\lambda T} \frac{(\lambda T)^n}{n!} \right) \left( \frac{n!}{T^n} \mathbb{1}_{\{0 < s_1 < \dots < s_n < T\}} \right) ds_1 \cdots ds_n. \end{aligned}$$

The second factor is the density of the order statistics of  $n$  independent uniform random variables on  $[0, T]$ . The case  $n = 0$  gives  $\mathbb{P}(N_T = 0) = \mathbb{P}(X_1 > T) = e^{-\lambda T}$ . Thus  $N_T \sim \mathcal{P}(\lambda T)$  and, conditional on  $N_T = n$ , the ordered positions have the law of these order statistics. The finite-intensity construction therefore shows that  $M|_{[0, T]}$  is a Poisson random measure with intensity  $\lambda \mathbb{1}_{[0, T]}(t) dt$ .

To conclude, let  $A_1, \dots, A_r$  be disjoint Borel subsets of  $\mathbb{R}_+$  and  $k_1, \dots, k_r \in \mathbb{N}$ . Since  $M(A_j \cap [0, T]) \uparrow M(A_j)$ , continuity from above of probabilities gives

$$\mathbb{P}(M(A_j) \leq k_j \text{ for every } j) = \lim_{T \rightarrow \infty} \prod_{j=1}^r e^{-\theta_{j,T}} \sum_{\ell=0}^{k_j} \frac{\theta_{j,T}^\ell}{\ell!}, \quad \theta_{j,T} = \lambda \text{Leb}(A_j \cap [0, T]).$$

Each factor tends to the distribution function at  $k_j$  of a  $\mathcal{P}(\lambda \text{Leb}(A_j))$  law; it tends to zero if this parameter is infinite. The same equalities apply to every subfamily of the  $A_j$ . They yield the Poisson distributions of the counts and their independence, including when some counts are almost surely infinite. This completes the proof.  $\square$

**Proposition 8.13** (Closure under restriction). Let  $M$  be a Poisson random measure with intensity  $\mu$ . Let  $(A_i)_{i \in I}$  be disjoint measurable subsets of  $E$ . Then  $(M|_{A_i})_{i \in I}$  are independent Poisson random measures, and  $M|_{A_i}$  has intensity  $\mu|_{A_i}$ .

This is an immediate consequence of the definition.

We deduce the following result, which is important in practice for reducing to a finite intensity measure.

**Corollary 8.14.** Let  $M$  be a Poisson random measure with intensity  $\mu$ , assumed  $\sigma$ -finite. Then we can write  $M = \sum_{i=1}^{\infty} M_i$ , where  $(M_i)_{i \geq 1}$  are independent Poisson random measures with finite intensities.

*Proof.* Write  $E = \bigcup_{i=1}^{\infty} A_i$  with  $(A_i)_{i \geq 1}$  disjoint elements of  $\mathcal{E}$  of finite  $\mu$ -measure. Then  $M = \sum_{i=1}^{\infty} M|_{A_i}$ , where  $(M|_{A_i})_{i \geq 1}$  are independent Poisson random measures with finite intensities  $\mu|_{A_i}$ .  $\square$

**Remark 8.15.** If  $\mu$  is  $s$ -finite, it is unclear whether  $M$  can be decomposed “measurably” as  $M = \sum_{i=1}^{\infty} M_i$ , where  $(M_i)_{i \geq 1}$  are independent Poisson random measures with finite intensities. But when  $\mu$  is  $\sigma$ -finite, we have just seen that this is possible.

### 8.3 Multiple points

Assume that  $\{x\} \in \mathcal{E}$  for every  $x \in E$ . The atoms of a measure  $\sum_{i \in I} \delta_{x_i} \in \mathcal{M}_{\text{atom}}(E)$  are not necessarily distinct.

If  $M$  is a Poisson random measure with intensity  $\mu$ ,  $M(\{x\}) \sim \mathcal{P}(\mu(\{x\}))$ . Thus

$$\forall x \in E, \mu(\{x\}) = 0 \iff \forall x \in E, \mathbb{P}(M(\{x\}) \in \{0, 1\}) = 1.$$

In fact, we will see that

$$\forall x \in E, \mu(\{x\}) = 0 \iff \mathbb{P}(\forall x \in E, M(\{x\}) \in \{0, 1\}) = 1,$$

under suitable assumptions ensuring measurability of the event.

**Definition 8.16.** Write

$$\mathcal{M}_{\text{atom}}^{\text{simple}}(E) = \left\{ \mu = \sum_{i \in I} \delta_{x_i} \in \mathcal{M}_{\text{atom}}(E) : x_i \neq x_j \text{ for } i \neq j \right\}.$$

Recall that  $\mathcal{M}_{\text{atom}}(E)$  is equipped with the smallest sigma-algebra for which the maps  $\mu \mapsto \mu(A)$ , taking values in  $\mathbb{N} \cup \{+\infty\}$ , are measurable for every  $A \in \mathcal{E}$ .

**Remark 8.17.** In general,  $\mathcal{M}_{\text{atom}}^{\text{simple}}(E)$  is not a measurable subset of  $\mathcal{M}_{\text{atom}}(E)$ . Indeed, let  $(X_i)_{i \geq 1}$  be i.i.d. random variables uniform on  $[0, 1] = E$ . Set  $M_1 = \sum_{n \geq 1} \delta_{X_n}$  and  $M_2 = \sum_{n \geq 1} 2\delta_{X_n}$ . Then  $M_1$  and  $M_2$  are two Poisson random measures with s-finite intensity  $\mu$  given by

$$\mu(A) = \begin{cases} 0 & \text{if } \lambda(A) = 0, \\ \infty & \text{if } \lambda(A) > 0, \end{cases}$$

where  $\lambda$  is Lebesgue measure on  $[0, 1]$ . The measure  $\mu$  is s-finite since  $\mu = \sum_{n=1}^{\infty} \lambda$ . The variables  $M_1$  and  $M_2$  have the same law,  $M_1 \in \mathcal{M}_{\text{atom}}^{\text{simple}}(E)$  almost surely and  $M_2 \notin \mathcal{M}_{\text{atom}}^{\text{simple}}(E)$ ; hence  $\mathcal{M}_{\text{atom}}^{\text{simple}}(E)$  is not measurable.

Finiteness and measurability assumptions are necessary.

**Lemma 8.18.** Assume that  $D = \{(x, x) : x \in E\} \in \mathcal{E} \otimes \mathcal{E}$ . Then

$$\mathcal{M}_{\text{atom}}^{\text{simple}, f}(E) = \{\mu \in \mathcal{M}_{\text{atom}}^{\text{simple}}(E) : \mu(E) < \infty\}$$

is measurable in  $\mathcal{M}_{\text{atom}}(E)$ .

*Proof.* Set  $\mathcal{M}_{\text{atom}}^f(E) = \{\mu \in \mathcal{M}_{\text{atom}}(E) : \mu(E) < \infty\}$ , a measurable subset of  $\mathcal{M}_{\text{atom}}(E)$ , and

$$\Phi : \mathcal{M}_{\text{atom}}^f(E) \longrightarrow \mathcal{M}_{\text{atom}}^f(E \times E), \quad \sum_{i=1}^m \delta_{x_i} \longmapsto \sum_{\substack{1 \leq i, j \leq m \\ i \neq j}} \delta_{(x_i, x_j)}.$$

Since  $\mathcal{M}_{\text{atom}}^{\text{simple}, f}(E) = \{\mu \in \mathcal{M}_{\text{atom}}^f(E) : \Phi(\mu)(D) = 0\}$ , it suffices to check that  $\mu \mapsto$

$\Phi(\mu)(D)$  is measurable. To this end, consider

$$\mathcal{C} = \{C \in \mathcal{E} \otimes \mathcal{E} : \mu \mapsto \Phi(\mu)(C) \text{ is measurable on } \mathcal{M}_{\text{atom}}^f(E)\}.$$

For finite  $\mu$  and  $A, B \in \mathcal{E}$ ,  $\Phi(\mu)(A \times B) = \mu(A)\mu(B) - \mu(A \cap B)$ . One checks that  $\mathcal{C}$  is a monotone class containing the measurable rectangles. This completes the proof. One can similarly show that  $\Phi$  is measurable.  $\square$

**Proposition 8.19.** Assume that  $\mu$  is  $\sigma$ -finite and  $D = \{(x, x) : x \in E\} \in \mathcal{E} \otimes \mathcal{E}$ . Let  $M$  be a Poisson random measure with intensity  $\mu$ . Then

$$\forall x \in E, \mu(\{x\}) = 0 \iff M \in \mathcal{M}_{\text{atom}}^{\text{simple}}(E) \text{ almost surely.}$$

This means there exists an event  $A$  such that  $\mathbb{P}(A) = 1$  and  $\omega \in A \Rightarrow M(\omega) \in \mathcal{M}_{\text{atom}}^{\text{simple}}(E)$ . If  $\mathbb{P}$  is complete (every subset of an event of probability zero is measurable), this is therefore equivalent to  $\mathbb{P}(M \in \mathcal{M}_{\text{atom}}^{\text{simple}}(E)) = 1$ . In this case,  $M$  is said to be *simple*, or to have *no multiple points*.

Recall that in general  $\mathcal{M}_{\text{atom}}^{\text{simple}}(E)$  is not measurable in  $\mathcal{M}_{\text{atom}}(E)$ .

*Proof.* For the reverse implication, suppose that  $\mu(\{x\}) = c > 0$  for some  $x \in E$ . Then

$$\mathbb{P}(M(\{x\}) \geq 2) = 1 - e^{-c} - ce^{-c} > 0,$$

which contradicts the almost sure absence of multiple points.

For the forward implication, first consider the case where  $\mu$  is finite. If  $\mu(E) = 0$ , then  $M(E) = 0$  almost surely, so  $M$  is almost surely simple. From now on, assume  $\mu(E) > 0$ . By the lemma,  $\mathbb{P}(M \in \mathcal{M}_{\text{atom}}^{\text{simple},f}(E))$  depends only on the law of  $M$ . Thus, we may assume that  $M$  is constructed from random variables as in the existence proof for Poisson random measures:

$$M = \sum_{i=1}^L \delta_{X_i},$$

where  $L \sim \mathcal{P}(\mu(E))$  and  $(X_i)_{i \geq 1}$  are i.i.d. with law  $\mu(\cdot)/\mu(E)$ , independent of  $L$ . Then, for  $1 \leq i \neq j$ ,

$$\mathbb{P}(X_i = X_j) = \frac{1}{\mu(E)^2} \iint_{E^2} \mathbb{1}_{\{x=y\}} \mu(dx) \mu(dy) = 0.$$

Hence  $\mathbb{P}(M \in \mathcal{M}_{\text{atom}}^{\text{simple},f}(E)) = 1$ .

In the general  $\sigma$ -finite case, write  $E = \bigcup_{i=1}^{\infty} A_i$  with disjoint  $A_i$  and  $\mu(A_i) < \infty$  for every  $i \geq 1$ . We have seen that  $M = \sum_{i=1}^{\infty} M|_{A_i}$ , where  $(M|_{A_i})_{i \geq 1}$  are independent Poisson random measures with respective intensities  $\mu|_{A_i}$ , which are finite. Now let

$$A = \bigcap_{i=1}^{\infty} \{M|_{A_i} \in \mathcal{M}_{\text{atom}}^{\text{simple},f}(E)\}.$$

By the finite case,  $\mathbb{P}(A) = 1$ . Moreover,  $\omega \in A \Rightarrow M(\omega) \in \mathcal{M}_{\text{atom}}^{\text{simple}}(E)$ .  $\square$

## 8.4 Campbell's formulas (additive and exponential)

Throughout this section,  $\mu \in \mathcal{M}_s(E)$  is fixed and  $M$  is a Poisson random measure with intensity  $\mu$ , on an arbitrary measurable space  $(E, \mathcal{E})$ .

**Theorem 8.20** (Additive formula). 1) *Nonnegative case.* Let  $f : E \rightarrow \mathbb{R}_+$  be measurable. Then the random variable  $M(f) = \int_E f(x)M(dx)$  has expectation

$$\mathbb{E}[M(f)] = \int_E f(x)\mu(dx).$$

2) *Integrable case.* Let  $f : E \rightarrow \mathbb{R}$  be  $\mu$ -integrable. Then  $M(f) = \int_E f(x)M(dx)$  is integrable and

$$\mathbb{E}[M(f)] = \int_E f(x)\mu(dx).$$

*Proof.* 1) Write  $f \geq 0$  as an increasing limit of simple functions  $f_n = \sum_{i=1}^{N_n} c_i \mathbb{1}_{A_i}$ . In particular,

$$\mathbb{E}[M(f_n)] = \sum_{i=1}^{N_n} c_i \mu(A_i) = \mu(f_n).$$

Then  $M(f_n) \uparrow M(f)$ . Thus  $M(f)$  is measurable and, by monotone convergence,  $\mathbb{E}[M(f)] = \mu(f)$ .

2) Write  $f = f^+ - f^-$  and apply 1) to  $f^+$  and  $f^-$ . □

**Theorem 8.21** (Exponential formula). 1) *Nonnegative case.* Let  $f : E \rightarrow \mathbb{R}_+$  be measurable. We have

$$\mathbb{E}[e^{-M(f)}] = \exp \left( - \int_E (1 - e^{-f(x)})\mu(dx) \right).$$

2) *Integrable case.* Let  $f : E \rightarrow \mathbb{R}$  be  $\mu$ -integrable. We have

$$\mathbb{E}[e^{iM(f)}] = \exp \left( \int_E (e^{if(x)} - 1)\mu(dx) \right).$$

*Proof.* 1) Write  $f \geq 0$  as an increasing limit of simple functions  $f_n = \sum_{i=1}^{N_n} c_i \mathbb{1}_{A_i}$ , with disjoint  $A_i$ . Then  $M(f_n) = \sum_{i=1}^{N_n} c_i M(A_i)$  is a linear combination of independent Poisson random variables, and we conclude using the form of the Laplace transform of a Poisson random variable and monotone convergence.

2) Write  $f = \lim_{n \rightarrow \infty} f_n$ , with  $f = f^+ - f^-$ ,  $f_n = f_n^+ - f_n^-$ ,  $0 \leq f_n^+ \uparrow f^+$ ,  $0 \leq f_n^- \uparrow f^-$  and  $f_n^+, f_n^-$  simple. As in 1), an explicit calculation gives

$$\mathbb{E}[e^{iM(f_n)}] = \exp \left( \int_E (e^{if_n(x)} - 1)\mu(dx) \right).$$

On the one hand, the right-hand side converges to  $\exp(\int_E (e^{if(x)} - 1)\mu(dx))$  by domi-

nated convergence, using

$$|e^{if_n(x)} - 1| \leq |f_n(x)| = f_n^+(x) + f_n^-(x) \leq f^+(x) + f^-(x) = |f(x)|.$$

On the other hand,  $M(f_n) = M(f_n^+) - M(f_n^-)$ , with  $M(f_n^+) \uparrow M(f^+)$  and  $M(f_n^-) \uparrow M(f^-)$  by monotone convergence. Thus  $M(f_n) \rightarrow M(f)$ . Since  $|e^{iM(f_n)}| = 1$ , we also have  $\mathbb{E}[e^{iM(f_n)}] \rightarrow \mathbb{E}[e^{iM(f)}]$  by dominated convergence, completing the proof.  $\square$

**Remark 8.22** (Characterization by the Laplace transform). Let  $\mu$  be an  $s$ -finite measure. A random variable  $N$  taking values in  $\mathcal{M}_{\text{atom}}(E)$  is a Poisson random measure with intensity  $\mu$  if and only if, for every measurable function  $f : E \rightarrow \mathbb{R}_+$ ,

$$\mathbb{E}[e^{-N(f)}] = \exp \left( - \int_E (1 - e^{-f(x)}) \mu(dx) \right).$$

It suffices to use the functions  $f = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ , with  $c_i \geq 0$  and disjoint measurable  $A_i$ , followed by uniqueness of the joint Laplace transform.

If  $\mu$  is  $\sigma$ -finite, we can also use the following characterization: for every real-valued  $\mu$ -integrable function  $f$ , we have  $N(|f|) < \infty$  almost surely and

$$\mathbb{E}[e^{iN(f)}] = \exp \left( \int_E (e^{if(x)} - 1) \mu(dx) \right).$$

First apply this identity to linear combinations of indicators of sets of finite  $\mu$ -measure, then exhaust  $E$  by such sets. This second characterization does not extend to measures that are merely  $s$ -finite.

For measurable  $f : E \rightarrow \mathbb{R}_+$ , by the additive formula,

$$\mathbb{E} \left[ \int f(x) M(dx) \right] < \infty \iff \int f d\mu < \infty.$$

The following result gives a necessary and sufficient condition for  $\int f(x) M(dx) < \infty$  almost surely.

**Proposition 8.23.** Let  $f : E \rightarrow \mathbb{R}_+$  be measurable. The following alternative holds:

- (a) if  $\int_E \min(1, f) d\mu < \infty$ , then almost surely  $\int_E f(x) M(dx) < \infty$ ;
- (b) if  $\int_E \min(1, f) d\mu = \infty$ , then almost surely  $\int_E f(x) M(dx) = \infty$ .

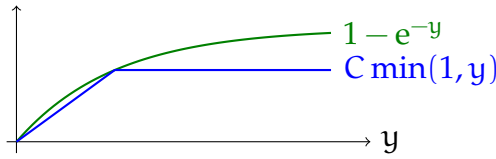
*Proof.* Note that, for  $x \in [0, \infty]$ ,  $x < \infty \iff \lim_{\lambda \downarrow 0} e^{-\lambda x} = 1$ . Thus, by the exponential formula,

$$\begin{aligned} \mathbb{P} \left( \int f(x) M(dx) < \infty \right) &= \lim_{\lambda \downarrow 0} \mathbb{E}[e^{-\lambda M(f)}] \\ &= \lim_{\lambda \downarrow 0} \exp \left( - \int (1 - e^{-\lambda f(x)}) \mu(dx) \right). \end{aligned}$$

If  $\int_E \min(1, f) d\mu < \infty$ , for  $\lambda \leq 1$ , we have  $1 - e^{-\lambda f(x)} \leq \min(1, f(x))$ , and the result

follows by dominated convergence.

If  $\int_E \min(1, f) d\mu = \infty$ , note that  $C \min(1, y) \leq 1 - e^{-y}$  for the constant  $C = 1 - e^{-1}$ .



Thus, for  $0 < \lambda < 1$ ,

$$\int (1 - e^{-\lambda f(x)}) \mu(dx) \geq C \int \min(1, \lambda f(x)) \mu(dx) \geq \lambda C \int \min(1, f(x)) \mu(dx) = \infty.$$

Hence  $\mathbb{P}(\int f(x) M(dx) < \infty) = 0$ . □

**Example 8.24.** If  $\mu(dx) = x^{-3/2} \mathbb{1}_{\{x>0\}} dx$  and  $f(x) = x$ , then  $M(f) < \infty$  almost surely, but  $\mathbb{E}[M(f)] = \infty$ .

## 8.5 Transformations of Poisson random measures

**Proposition 8.25** (Mapping theorem). Let  $M$  be a Poisson random measure with  $s$ -finite intensity  $\mu$  on  $(E, \mathcal{E})$  and let  $f : (E, \mathcal{E}) \rightarrow (F, \mathcal{G})$  be measurable. Then  $f_* M$  is a Poisson random measure with intensity  $f_* \mu$  on  $F$ .

*Proof.* First, let us check that  $f_*$  is the map

$$\mathcal{M}_{\text{atom}}(E) \longrightarrow \mathcal{M}_{\text{atom}}(F), \quad \sum_{i \in I} \delta_{x_i} \longmapsto \sum_{i \in I} \delta_{f(x_i)},$$

and that it is measurable. For  $B \in \mathcal{G}$ , we indeed have

$$\left( \sum_{i \in I} \delta_{f(x_i)} \right) (B) = \#\{i \in I : x_i \in f^{-1}(B)\} = \left( \sum_{i \in I} \delta_{x_i} \right) (f^{-1}(B)).$$

For  $B \in \mathcal{G}$  and  $k \geq 0$ ,

$$\begin{aligned} f_*^{-1}(\{\nu \in \mathcal{M}_{\text{atom}}(F) : \nu(B) = k\}) &= \{\nu \in \mathcal{M}_{\text{atom}}(E) : (f_* \nu)(B) = k\} \\ &= \{\nu \in \mathcal{M}_{\text{atom}}(E) : \nu(f^{-1}(B)) = k\}, \end{aligned}$$

which is measurable. Thus  $f_* M \in \mathcal{M}_{\text{atom}}(F)$  and, if  $B \in \mathcal{G}$ ,

$$(f_* M)(B) = M(f^{-1}(B)) \sim \mathcal{P}(\mu(f^{-1}(B))) = \mathcal{P}((f_* \mu)(B)).$$

Moreover, if  $B_1, \dots, B_k \in \mathcal{G}$  are disjoint, their preimages  $f^{-1}(B_1), \dots, f^{-1}(B_k)$  are disjoint, so the variables  $(f_* M)(B_i) = M(f^{-1}(B_i))$  are independent. □

**Remark 8.26.** Another advantage of working here with  $s$ -finite measures rather than  $\sigma$ -finite ones is that the image of a  $\sigma$ -finite measure need not be  $\sigma$ -finite.

**Proposition 8.27** (Marking theorem). Assume that  $M = \sum_{k=1}^L \delta_{X_k}$ , where  $L \in \mathbb{N} \cup \{+\infty\}$  and  $X_k \in E$  are random variables, is a Poisson random measure with intensity  $\mu \in \mathcal{M}_s(E)$ . Let  $(Y_i)_{i \geq 1}$  be i.i.d. random variables in  $(F, \mathcal{G})$ , with law  $\mu'$ , independent of  $(L, X_1, X_2, \dots)$ . Then

$$\widetilde{M} = \sum_{k=1}^L \delta_{(X_k, Y_k)}$$

is a Poisson random measure on  $E \times F$  with intensity  $\mu \otimes \mu'$ .

*Proof.* Let  $f : E \times F \rightarrow \mathbb{R}_+$  be measurable. We compute

$$\begin{aligned} \mathbb{E}[\exp(-\widetilde{M}(f))] &= \mathbb{E} \left[ \mathbb{E} \left[ \exp \left( - \sum_{k=1}^L f(X_k, Y_k) \right) \middle| L, (X_k)_{k \geq 1} \right] \right] \\ &= \mathbb{E} \left[ \prod_{k=1}^L \int_F \exp(-f(X_k, y)) \mu'(dy) \right] = \mathbb{E}[\exp(-M(F))], \end{aligned}$$

where  $F(x) = -\ln \int_F e^{-f(x,y)} \mu'(dy) \geq 0$ . Thus, by the exponential formula,

$$\begin{aligned} \mathbb{E}[\exp(-\widetilde{M}(f))] &= \exp \left( - \int_E (1 - e^{-F(x)}) \mu(dx) \right) \\ &= \exp \left( - \int_{E \times F} (1 - e^{-f(x,y)}) (\mu \otimes \mu')(dx dy) \right). \end{aligned}$$

This gives the result. □

**Example 8.28.** Let  $M = \sum_{k=1}^L \delta_{X_k}$  be a Poisson random measure with intensity  $\mu$  on  $E$ , given with this random enumeration, and let  $(Y_i)_{i \geq 1}$  be an i.i.d. sequence with distribution  $\text{Bernoulli}(p)$ , independent of  $(L, X_1, X_2, \dots)$ . Then

$$M_p = \sum_{k=1}^L Y_k \delta_{X_k}$$

is a Poisson random measure with intensity  $p\mu$ , since  $M_p$  is identified with the restriction to  $E \times \{1\}$  of  $\widetilde{M} = \sum_{k=1}^L \delta_{(X_k, Y_k)}$ , which has intensity  $\mu \otimes \text{Bernoulli}(p)$ .

## 8.6 Poisson random measures with a time parameter

We work with a measurable space  $(E, \mathcal{E})$ , with  $\{x\} \in \mathcal{E}$  for every  $x \in E$ . Assume that  $\mathbb{P}$  is complete and  $\mu$  is  $\sigma$ -finite on  $E$ .

**Definition 8.29.** A *Poisson point process* on  $E$  with intensity  $\mu$  is a Poisson random measure  $M$  on  $\mathbb{R}_+ \times E$  with intensity  $dt \otimes \mu$ .

**Proposition 8.30.** Almost surely, for every  $t \geq 0$ ,  $M(\{t\} \times E) \leq 1$ .

*Proof.* Choose measurable sets  $E_n \uparrow E$  such that  $\mu(E_n) < \infty$ , and set  $\pi(t, x) = t$ . For  $m, n \geq 1$ , the restriction and mapping properties show that

$$N_{m,n} = \pi_* (M|_{[0,m] \times E_n})$$

is a Poisson random measure on  $\mathbb{R}_+$  with intensity  $\mu(E_n) \mathbb{1}_{[0,m]}(t) dt$ . This intensity is finite and has no atoms, so  $N_{m,n}$  is almost surely simple.

On the countable intersection of these events of probability 1, we have, for all  $m, n \geq 1$  and  $t \in [0, m]$ ,

$$M(\{t\} \times E_n) = N_{m,n}(\{t\}) \leq 1.$$

Letting  $n \rightarrow \infty$  and then choosing  $m \geq t$ , we obtain  $M(\{t\} \times E) \leq 1$  for every  $t \geq 0$ .  $\square$

**Definition 8.31.** For  $t \geq 0$ , set

$$e(t) = \begin{cases} x & \text{if } M(\{(t, x)\}) = 1, \\ \partial & \text{otherwise, where } \partial \text{ is a cemetery state.} \end{cases}$$

**Proposition 8.32.** 1) For every  $t \geq 0$ ,  $e(t)$  is a random variable.

2) If  $\mu(E) = \infty$ ,  $\{t : e(t) \neq \partial\}$  is almost surely dense.

3) If  $0 < \mu(E) < \infty$ , almost surely,  $\{t : e(t) \neq \partial\}$  is infinite and discrete and can be ordered using random variables  $0 < T_1 < T_2 < \dots$  satisfying:

(a)  $T_1, T_2 - T_1, T_3 - T_2, \dots$  are i.i.d. with distribution  $\text{Exp}(\mu(E))$ ;

(b) the variables  $(e(T_i))_{i \geq 1}$  are i.i.d. with law  $\mu(\cdot)/\mu(E)$ ;

(c)  $(e(T_i))_{i \geq 1}$  is independent of  $(T_i)_{i \geq 1}$ .

*Proof.* 1) For  $A \in \mathcal{E}$ ,

$$\{e(t) \in A\} = \{M(\{t\} \times A) = 1\}, \quad \{e(t) = \partial\} = \{M(\{t\} \times E) = 0\}.$$

2) If  $\mu(E) = \infty$ , for all  $a < b$  in  $\mathbb{Q}_+$ ,  $M([a, b] \times E) \sim \mathcal{P}(\infty)$ , hence equals  $\infty$  almost surely.

3) The law of  $(T_i, e(T_i))_{i \geq 1}$  depends only on the law of  $M$ . For example,

$$T_1 = \inf\{t \geq 0 : M([0, t] \times E) \geq 1\}, \quad \{e(T_i) \in A\} = \{M(\{T_i\} \times A) = 1\}.$$

The property to be proved therefore depends only on the law of  $M$ . Now construct  $M$  as follows. Let  $P = \sum_{i=1}^{\infty} \delta_{T_i}$  be a Poisson random measure with intensity  $\mu(E) dt$  on

$\mathbb{R}_+$  and let  $(Y_i)_{i \geq 1}$  be i.i.d. variables with law  $\mu(\cdot)/\mu(E)$ , independent of  $P$ . Then, by marking,

$$M = \sum_{i=1}^{\infty} \delta_{(T_i, Y_i)}$$

is a Poisson random measure with intensity  $dt \otimes \mu$ . Moreover, we have seen that  $(P([0, t]))_{t \geq 0}$  is a Poisson process with intensity  $\mu(E)$ , and we can indeed order  $T_1 < T_2 < \dots$  satisfying (a). Properties (b) and (c) follow from  $e(T_i) = Y_i$ .  $\square$

**Example 8.33.** Take  $\mu(dx) = x^{-3/2} \mathbb{1}_{\{x > 0\}} dx$  on  $E = \mathbb{R}_+^*$ . We have  $\mu(E) = \infty$ . For  $t \geq 0$ , set

$$X_t = \sum_{s \leq t} e(s) \mathbb{1}_{\{e(s) \neq \emptyset\}} = M(f_t), \quad f_t(s, x) = x \mathbb{1}_{\{s \leq t\}}.$$

Since  $\int_{\mathbb{R}_+ \times \mathbb{R}_+^*} (1 \wedge f_t(s, x)) ds \mu(dx) < \infty$ , for every  $t \geq 0$ , almost surely  $X_t < \infty$ ; by monotonicity, almost surely,  $X_t < \infty$  for every  $t \geq 0$ . Nevertheless, for every  $t > 0$ ,  $\int_{\mathbb{R}_+ \times \mathbb{R}_+^*} f_t(s, x) ds \mu(dx) = \infty$ , so  $\mathbb{E}[X_t] = \infty$ .

Using the properties of Poisson random measures, one can check that  $(X_t)_{t \geq 0}$  is a process with stationary independent increments, and it is càdlàg by construction: it is a nondecreasing Lévy process (a *subordinator*).

## 8.7 Palm's formula

**Theorem 8.34** (Palm's formula / Mecke equation). Let  $\mu$  be  $\sigma$ -finite and  $M$  a Poisson random measure with intensity  $\mu$ . Let  $f : E \times \mathcal{M}_{\text{atom}}(E) \rightarrow \mathbb{R}_+$  be measurable. Then

$$\mathbb{E} \left[ \int_E f(x, M) M(dx) \right] = \int_E \mathbb{E}[f(x, M + \delta_x)] \mu(dx). \quad (9)$$

Palm's formula describes the point cloud seen from a distinguished point: the left-hand side sums over all points of the cloud, with their multiplicities, while the right-hand side adds a point at  $x$  to a Poisson point cloud with the original law and integrates against  $\mu(dx)$ . In general, this identity does not correspond to ordinary conditioning on  $M(\{x\}) \geq 1$ ; the distinction is essential when  $\mu$  has atoms. This result can be viewed as an extension of the additive formula that takes the atoms into account.

*Proof.* The map  $(x, \nu) \mapsto \nu + \delta_x$  is jointly measurable, since, for every  $A \in \mathcal{E}$ ,

$$(\nu + \delta_x)(A) = \nu(A) + \mathbb{1}_A(x).$$

Choose  $E_n \uparrow E$  such that  $\mu(E_n) < \infty$ , and set

$$H = \{\nu \in \mathcal{M}_{\text{atom}}(E) : \nu(E_n) < \infty \text{ for every } n\}.$$

This set is measurable and  $\mathbb{P}(M \in H) = 1$ . Replacing  $M$  by the zero measure on the null event  $\{M \notin H\}$  if necessary, we can work with a version taking values in  $H$ . On

H, the map  $\nu \mapsto \nu|_{E_n}$  is a finite kernel. A monotone class argument therefore gives measurability of

$$\nu \mapsto \int_{E_n} f(x, \nu) \nu(dx).$$

Letting  $n$  increase yields measurability of the integral over  $E$ . These observations, together with joint measurability of adding an atom and Tonelli's theorem, justify both sides of (9) and show that they depend only on the law of  $M$ .

*Step 1:  $\mu$  finite.* If  $\mu(E) = 0$ , then  $M = 0$  almost surely and both sides of the formula vanish. Thus assume  $\mu(E) > 0$ . Write  $M = \sum_{i=1}^L \delta_{X_i}$ , where  $L \sim \mathcal{P}(\mu(E))$  and  $(X_i)_{i \geq 1}$  are i.i.d. with law  $\mu(\cdot)/\mu(E)$ , independent of  $L$ . Then

$$\begin{aligned} \mathbb{E} \left[ \int_E f(x, M) M(dx) \right] &= \sum_{k=1}^{\infty} \frac{\mu(E)^k}{k!} e^{-\mu(E)} \mathbb{E} \left[ \sum_{i=1}^k f \left( X_i, \sum_{j=1}^k \delta_{X_j} \right) \right] \\ &= \sum_{k=1}^{\infty} \frac{\mu(E)^k}{k!} e^{-\mu(E)} k \mathbb{E} \left[ f \left( X_k, \sum_{j=1}^k \delta_{X_j} \right) \right] \\ &= \sum_{k=1}^{\infty} \frac{\mu(E)^k}{k!} e^{-\mu(E)} k \int_E \mathbb{E} \left[ f \left( x, \sum_{j=1}^{k-1} \delta_{X_j} + \delta_x \right) \right] \frac{\mu(dx)}{\mu(E)} \\ &= \int_E \left( \sum_{k=1}^{\infty} \frac{\mu(E)^{k-1}}{(k-1)!} e^{-\mu(E)} \mathbb{E} \left[ f \left( x, \sum_{j=1}^{k-1} \delta_{X_j} + \delta_x \right) \right] \right) \mu(dx) \\ &= \int_E \mathbb{E}[f(x, M + \delta_x)] \mu(dx). \end{aligned}$$

*Step 2:  $\mu$   $\sigma$ -finite.* Write  $E = \bigsqcup_{i \geq 1} A_i$ , with  $\mu(A_i) < \infty$ , and set  $M_i = M|_{A_i}$  and  $R_i = M|_{A_i^c}$ . These two restrictions are independent. For each fixed atomic measure  $r$ , apply the finite case to  $M_i$  and the function  $(x, \nu) \mapsto f(x, \nu + r)$ , then integrate in  $r$  with respect to the law of  $R_i$ . Independence and Tonelli's theorem give

$$\mathbb{E} \left[ \int_{A_i} f(x, M) M(dx) \right] = \int_{A_i} \mathbb{E}[f(x, M + \delta_x)] \mu(dx).$$

Summing over  $i \geq 1$  and applying Tonelli again, we obtain (9). □

**Remark 8.35.** If the atoms of  $M$  can be enumerated measurably (as is the case if  $\mu$  is  $\sigma$ -finite and  $E$  is Polish), Palm's formula can be written as

$$\mathbb{E} \left[ \int_E f(x, M - \delta_x) M(dx) \right] = \int_E \mathbb{E}[f(x, M)] \mu(dx).$$

## Part II

# Exercises

### Contents

|   |                                                                      |     |
|---|----------------------------------------------------------------------|-----|
| 1 | Weak convergence in $\mathbb{R}^k$ . . . . .                         | 97  |
| 2 | Modes of convergence (almost sure, in probability, $L^1$ ) . . . . . | 110 |
| 3 | Probability measures on a metric space . . . . .                     | 124 |
| 4 | The space of continuous functions on a compact set . . . . .         | 156 |
| 5 | Donsker's theorem, Brownian bridge . . . . .                         | 179 |
| 6 | The space of càdlàg functions . . . . .                              | 198 |
| 7 | Poisson random measures . . . . .                                    | 206 |

# 1 Weak convergence in $\mathbb{R}^k$

Related lecture notes: [Chapter 1](#), p. 6.

We write  $\mathcal{M}_1(\mathbb{R}^k)$  for the set of probability measures on  $\mathbb{R}^k$ , where  $k \geq 1$  is a fixed integer, and  $\Rightarrow$  denotes weak convergence in this space.

## 1.1 Exercises

**Exercise 1.1.** [solution p. 100] Let  $(\mu_n)_{n \geq 1}$  be a sequence of measures in  $\mathcal{M}_1(\mathbb{R})$  and  $\mu \in \mathcal{M}_1(\mathbb{R})$ . Assume that  $\mu_n \Rightarrow \mu$ . Are there any implications between the following assertions?

- (a)  $\mu_n$  has a density for all sufficiently large  $n$
- (b)  $\mu$  has a density
- (c)  $\mu_n$  is atomic for all sufficiently large  $n$
- (d)  $\mu$  is atomic.

Recall that an atomic measure is a measure of the form  $\sum \alpha_i \delta_{b_i}$  for sequences  $\alpha_i \in \mathbb{R}_+$  and  $b_i \in \mathbb{R}$ , and that in  $\mathbb{R}^n$  saying that a measure has a density means that it has one with respect to Lebesgue measure (the usual convention is simply to say that it has a density).

\* \* \*

**Exercise 1.2.** [solution p. 101] Show that a family  $(\mu_i)_{i \in I}$  of measures in  $\mathcal{M}_1(\mathbb{R})$  is tight if and only if there exists a measurable function  $f : \mathbb{R} \rightarrow [0, \infty)$  such that  $f(x) \rightarrow \infty$  as  $|x| \rightarrow \infty$  and  $\sup_{i \in I} \int_{\mathbb{R}} f d\mu_i < \infty$ .

\* \* \*

**Exercise 1.3.** [solution p. 101] Let  $(X_n)_{n \geq 1}$  be real-valued random variables. Show that the sequence  $(X_n)_{n \geq 1}$  is tight if and only if, for every  $\varepsilon > 0$  and every sequence  $(c_n)_{n \geq 1}$  of strictly positive real numbers such that  $c_n \rightarrow 0$ , we have  $\mathbb{P}(c_n |X_n| \geq \varepsilon) \rightarrow 0$ .

\* \* \*

**Exercise 1.4.** [solution p. 102] (**Riesz's theorem and Scheffé's lemma**) Let  $(E, \mathcal{A}, \mu)$  be a measure space, where  $\mu$  is a positive measure (not necessarily finite). Let  $(f_n)_{n \geq 1}$  be a sequence of measurable functions from  $E$  to  $\mathbb{R}$  such that:

$$f_n \rightarrow f \quad \mu - \text{almost everywhere.} \quad (1)$$

- (1) Let  $p \in [1, \infty[$ . Assume that the convergence (1) holds, that for every  $n \geq 1$ ,  $f_n \in L^p(\mu)$ ,  $f \in L^p(\mu)$  and  $\|f_n\|_p \rightarrow \|f\|_p$  as  $n \rightarrow \infty$ . Prove that  $f_n \rightarrow f$  in  $L^p(\mu)$  (Riesz's theorem).

You may introduce the function  $g_n = 2^p(|f_n|^p + |f|^p) - |f_n - f|^p$ .

- (2) Prove Scheffé's lemma:

$$\text{if (1), } f_n \text{ and } f \text{ are } \mu\text{-integrable, } f_n \geq 0 \text{ and } \int_E f_n d\mu \xrightarrow{n \rightarrow \infty} \int_E f d\mu$$

$$\text{then } \int_E |f_n - f| d\mu \xrightarrow{n \rightarrow \infty} 0.$$

- (3) Let  $(X_n)_{n \geq 1}$  be a sequence of nonnegative integrable random variables converging almost surely to  $X$ . Assume that  $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X] < \infty$ . Show that  $X_n$  converges to  $X$  in  $L^1$ .
- (4) In this question, assume that  $E = \mathbb{R}$  and  $\mathcal{A} = \mathcal{B}(\mathbb{R})$ . Let  $(X_n)_{n \geq 1}$  be a sequence of real-valued random variables such that  $X_n$  has distribution  $f_n d\mu$ . Show that if  $f_n$  converges  $\mu$ -almost everywhere to a probability density  $f$ , then  $X_n \Rightarrow X$ , where  $X$  has distribution  $f d\mu$ .

\* \* \*

**Exercise 1.5.** [solution p. 102] (**Laplace transform**) Let  $\mu \in \mathcal{M}_1(\mathbb{R}_+)$ . Its Laplace transform is defined by the integral  $L_\mu(t) = \int_{\mathbb{R}_+} e^{-tx} \mu(dx)$  for  $t \geq 0$ .

- (1) Verify that  $L_\mu$  is of class  $C^\infty$  on  $]0, \infty[$ .
- (2) Show that for every  $x > 0$  we have

$$\mu([0, x[) + \frac{1}{2}\mu(\{x\}) = \lim_{t \rightarrow \infty} \sum_{k=0}^{\lfloor tx \rfloor} \frac{(-t)^k}{k!} L_\mu^{(k)}(t),$$

where  $L_\mu^{(k)}$  denotes the  $k$ th derivative of  $L_\mu$ . Deduce that if  $\mu, \nu \in \mathcal{M}_1(\mathbb{R}_+)$  have the same Laplace transform, then  $\mu = \nu$ .

- (3) Let  $(\mu_n)_{n \geq 1}$  be a sequence of measures in  $\mathcal{M}_1(\mathbb{R}_+)$ . Assume that  $L_{\mu_n}$  converges pointwise on  $\mathbb{R}_+$  to a function  $L$  that is right-continuous at 0. Show that there exists  $\mu \in \mathcal{M}_1(\mathbb{R}_+)$  such that  $L = L_\mu$  and  $\mu_n \Rightarrow \mu$ .

*You may draw inspiration from the proof of Lévy's theorem.*

\* \* \*

**Exercise 1.6.** [solution p. 104] Let  $(\mu_n)_{n \geq 1}$  be a sequence of measures in  $\mathcal{M}_1(\mathbb{R})$  and  $\mu \in \mathcal{M}_1(\mathbb{R})$ .

- (1) Assume that  $(\mu_n)_{n \geq 1}$  is tight. Show that

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall n \geq 1, \quad |x - y| \leq \delta \implies |\phi_{\mu_n}(x) - \phi_{\mu_n}(y)| \leq \varepsilon.$$

(in other words, the sequence  $(\phi_{\mu_n})_{n \geq 1}$  is uniformly equicontinuous).

- (2) Assume that  $\mu_n \Rightarrow \mu$ . Show that  $\phi_{\mu_n}$  converges to  $\phi_\mu$  uniformly on every compact set. Give an example in which the convergence is not uniform.
- (3) Is the converse of the statement in the first question true? In other words, if the sequence  $(\phi_{\mu_n})_{n \geq 1}$  is uniformly equicontinuous, is the sequence  $(\mu_n)_{n \geq 1}$  tight?

**Exercise 1.7.** [solution p. 105] Let  $(\mu_n)_{n \geq 1}$  be a sequence of measures in  $\mathcal{M}_1(\mathbb{R})$  and let  $\mu \in \mathcal{M}_1(\mathbb{R})$  have no atoms. Show that  $\mu_n \Rightarrow \mu$  if and only if  $\sup_{s \in \mathbb{R}} |F_{\mu_n}(s) - F_\mu(s)| \rightarrow 0$ .

\* \* \*

**Exercise 1.8.** [solution p. 106] (**Glivenko–Cantelli theorem**) Let  $(X_n)_{n \geq 1}$  be a sequence of independent real-valued random variables with common distribution  $\mu$ . Consider  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ , the empirical measure of these random variables. Write  $F_n$  for the distribution function of  $\mu_n$  and  $F$  for that of  $\mu$ . The aim of this exercise is to prove the Glivenko–Cantelli theorem:

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

(1) Prove  $(\star)$  when  $\mu$  is the uniform distribution on  $[0, 1]$ .

For  $0 < x < 1$ , set  $G(x) = \inf\{s \in \mathbb{R} : F(s) \geq x\}$  (called the generalized inverse of  $F$ ). One can verify that  $G$  is nondecreasing and left-continuous on  $]0, 1[$ . Let  $(Y_i)_{i \geq 1}$  be a sequence of independent random variables uniformly distributed on  $]0, 1[$ .

(2) (a) Show that for every  $t \in \mathbb{R}$  and  $x \in ]0, 1[$  we have

$$F(t) \geq x \iff t \geq G(x). \quad (2)$$

(b) Show that  $(G(Y_i))_{i \geq 1}$  are independent random variables with common distribution  $\mu$ .

(c) Show that the sequences of functions  $(F_n)_{n \geq 1}$  and  $(A_n \circ F)_{n \geq 1}$  have the same distribution, where  $A_n$  is the empirical distribution function of  $Y_1, \dots, Y_n$ .

(3) Deduce  $(\star)$ .

\* \* \*

**Exercise 1.9.** [solution p. 107] (**Characterization of distribution functions on  $\mathbb{R}^k$** ) For  $x = (x_1, \dots, x_k) \in \mathbb{R}^k$  and  $y = (y_1, \dots, y_k) \in \mathbb{R}^k$ , write  $x \leq y$  if  $x_i \leq y_i$  for every  $1 \leq i \leq k$ . Let  $F : \mathbb{R}^k \rightarrow [0, 1]$  be a function. It is said to be *right-continuous* if  $F(x^n) \rightarrow F(x)$  whenever  $x^n \downarrow x$ . It is said to be *proper* if  $F(x) \rightarrow 1$  as  $\min_i x_i \rightarrow \infty$  and  $F(x) \rightarrow 0$  as  $\min_i x_i \rightarrow -\infty$ . It is said to be a *distribution function* if there exists a measure  $\mu \in \mathcal{M}_1(\mathbb{R}^k)$  such that  $F(x) = \mu(\{y \in \mathbb{R}^k : y \leq x\})$ .

(1) Give an example of a function  $F : \mathbb{R}^2 \rightarrow [0, 1]$  that is right-continuous, proper and nondecreasing in each variable, but is not a distribution function.

We say that  $F$  has *nonnegative increments* if, for every rectangle  $]x, y] = ]x_1, y_1] \times \dots \times ]x_k, y_k]$ , we have  $F([x, y]) := \sum_u s(u)F(u) \geq 0$ , where the sum runs over all corners  $u$  of  $]x, y]$  and  $s(u) = (-1)^p$  with  $p = \sum_{i=1}^k \mathbb{1}_{u_i = x_i}$ .

- (2) Show that  $F : \mathbb{R}^k \rightarrow [0, 1]$  is a distribution function if and only if it is right-continuous, proper and has nonnegative increments.

For a probabilistic proof of the converse, you may set  $\mathbf{1} = (1, \dots, 1)$  and justify the existence, for every  $n \geq 1$ , of a probability measure  $\mu_n$  supported on  $(2^{-n}\mathbb{Z})^k$  such that

$$\mu_n(\{2^{-n}\mathbf{x}\}) = F([2^{-n}(\mathbf{x} - \mathbf{1}), 2^{-n}\mathbf{x}]), \quad \mathbf{x} \in \mathbb{Z}^k.$$

Then construct random variables  $X^n$  with distribution  $\mu_n$  such that  $X^m - 2^{-m}\mathbf{1} < X^n \leq X^m$  coordinatewise for  $m < n$ , and consider  $X = \lim_n X^n$ .

\* \* \*

**Exercise 1.10.** [solution p. 108] Let  $(E, d)$  be a metric space equipped with its Borel sigma-algebra, and let  $f : E \rightarrow \mathbb{R}$  be measurable. Show that the set of discontinuity points of  $f$  is measurable.

*Hint.* For  $\varepsilon, \delta > 0$ , you may verify that  $U_{\varepsilon, \delta} := \{x \in E : \exists y, z \in B(x, \varepsilon), |f(y) - f(z)| > \delta\}$  is open.

\* \* \*

**Exercise 1.11.** [solution p. 108] Let  $\mu_n, \mu$  be probability measures on  $\mathbb{R}$ . Write  $\lambda$  for Lebesgue measure on  $\mathbb{R}$  and  $\phi_\nu(t) = \int e^{itx} \nu(dx)$  for a probability measure  $\nu$  on  $\mathbb{R}$ .

- (1) Prove that  $\mu_n \Rightarrow \mu$  if and only if the characteristic function of  $\mu_n$  converges pointwise to the characteristic function of  $\mu$  at Lebesgue-almost every point.
- (2) Is it true that  $\mu_n \Rightarrow \mu$  if and only if the characteristic function of  $\mu_n$  converges pointwise on a countable dense set to the characteristic function of  $\mu$ ?

\* \* \*

**Exercise 1.12.** [solution p. 109] Let  $(X_n)_{n \geq 1}$  and  $X$  be real-valued random variables. Let  $D$  be a countable dense subset of  $\mathbb{R}$ .

- (1) Alix says: “if  $X_n$  converges in distribution to  $X$  as  $n \rightarrow \infty$ , then for every  $u \in D$  we have  $\mathbb{P}(X_n \leq u) \rightarrow \mathbb{P}(X \leq u)$  as  $n \rightarrow \infty$ .” Is he right? Justify your answer.
- (2) Billie says: “if for every  $u \in D$  we have  $\mathbb{P}(X_n \leq u) \rightarrow \mathbb{P}(X \leq u)$  as  $n \rightarrow \infty$ , then  $X_n$  converges in distribution to  $X$  as  $n \rightarrow \infty$ .” Is she right? Justify your answer.

## 1.2 Solutions

**Solution to Exercise 1.1.** [statement p. 97] No:

- if  $\mu_n$  has a density,  $\mu$  may either have a density (take  $\mu = \mu_n$ ), be atomic (take  $\mu_n(dx) = \frac{n}{2} \mathbb{1}_{[-1/n, 1/n]}(x) dx$ , which converges weakly to  $\delta_0$ ), or even be singular with respect to Lebesgue measure (consider the devil’s staircase).
- if  $\mu_n$  is atomic,  $\mu$  may either be atomic (take  $\mu = \mu_n$ ) or have a density (take  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{i/n}$ , which converges weakly to Lebesgue measure on  $[0, 1]$ ).

**Solution to Exercise 1.2.** [statement p. 97]  $\Leftarrow$  Set  $C = 1 + \sup_{i \in I} \int_{\mathbb{R}} f d\mu_i$ . Let  $\varepsilon > 0$  and choose  $A > 0$  such that  $f(x) \geq C/\varepsilon$  for  $|x| \geq A$ . For  $i \in I$ , we write

$$C \geq \int_{\mathbb{R}} f d\mu_i \geq \int_{\mathbb{R} \setminus [-A, A]} f d\mu_i \geq \frac{C}{\varepsilon} \int_{\mathbb{R} \setminus [-A, A]} d\mu_i,$$

so that  $\mu_i(\mathbb{R} \setminus [-A, A]) \leq \varepsilon$ .

$\Rightarrow$  Let  $(A_n)_{n \geq 1}$  be a sequence of strictly positive real numbers such that for every  $n \geq 1$  and  $i \in I$  we have

$$\mu_i(\mathbb{R} \setminus [-A_n, A_n]) \leq \frac{1}{n^3}.$$

Without loss of generality, we may assume that the sequence  $(A_n)$  is strictly increasing and  $A_n \rightarrow \infty$ . Define  $f$  by  $f(x) = 0$  if  $|x| \leq A_1$  and  $f(x) = n$  if  $A_n < |x| \leq A_{n+1}$ , so that  $f$  is measurable,  $f(x) \rightarrow \infty$ , and for  $i \in I$ ,

$$\begin{aligned} \int_{\mathbb{R}} f d\mu_i &= \sum_{n=1}^{\infty} \int_{[-A_{n+1}, A_{n+1}] \setminus [-A_n, A_n]} f d\mu_i \\ &\leq \sum_{n=1}^{\infty} n \mu_i([-A_{n+1}, A_{n+1}] \setminus [-A_n, A_n]) \\ &\leq \sum_{n=1}^{\infty} n \mu_i(\mathbb{R} \setminus [-A_n, A_n]) \\ &\leq \sum_{n=1}^{\infty} \frac{1}{n^2}, \end{aligned}$$

which proves the result.

\* \* \*

**Solution to Exercise 1.3.** [statement p. 97] Assume that the sequence  $(X_n)_{n \geq 1}$  is tight. Let  $\varepsilon > 0$  and let  $(c_n)_{n \geq 1}$  be a sequence of strictly positive real numbers such that  $c_n \rightarrow 0$ . Let  $\eta > 0$ . By tightness, there exists  $M > 0$  such that  $\mathbb{P}(|X_n| \geq M) \leq \eta$ . For sufficiently large  $n$ ,  $\varepsilon/c_n \geq M$ , and then

$$\mathbb{P}(c_n |X_n| \geq \varepsilon) \leq \mathbb{P}(|X_n| \geq M) \leq \eta,$$

which proves the result.

Conversely, argue by contradiction, assuming that the sequence  $(X_n)_{n \geq 1}$  is not tight. We can then find  $\varepsilon > 0$  and a subsequence indexed by  $\phi(n)$  such that  $\mathbb{P}(|X_{\phi(n)}| \geq n) \geq \varepsilon$  for every  $n$ . Define  $c_i = 1$  for  $1 \leq i < \phi(1)$  and  $c_i = 1/k$  for  $\phi(k) \leq i < \phi(k+1)$ . This sequence is strictly positive, tends to 0 and satisfies  $c_{\phi(k)} = 1/k$ . Then, for every  $n \geq 1$ ,

$$\mathbb{P}(c_{\phi(n)} |X_{\phi(n)}| \geq 1) \geq \varepsilon,$$

contradiction.

**Solution to Exercise 1.4.** [statement p. 97]

(1) By the triangle inequality, we observe that

$$\begin{aligned} |f_n - f|^p &\leq (|f_n| + |f|)^p \leq 2^p \max(|f_n|^p, |f|^p) \\ &\leq 2^p (|f_n|^p + |f|^p), \end{aligned}$$

so that  $g_n \geq 0$ . Apply Fatou's lemma to  $g_n$ :

$$\begin{aligned} 2^{p+1} \int_E |f|^p d\mu &= \int_E \left( \liminf_{n \rightarrow \infty} g_n \right) d\mu \\ &\leq \liminf_{n \rightarrow \infty} \int_E 2^p (|f_n|^p + |f|^p) d\mu \\ &\quad - \limsup_{n \rightarrow \infty} \int_E |f_n - f|^p d\mu. \end{aligned}$$

We deduce that

$$\limsup_{n \rightarrow \infty} \int_E |f_n - f|^p d\mu \leq 0,$$

which proves the result.

(2) This is simply Riesz's theorem for  $p = 1$ .

(3) This is simply Scheffé's lemma.

(4) Let  $s \in \mathbb{R}$ . We show that  $\mathbb{P}(X_n \leq s) \rightarrow \mathbb{P}(X \leq s)$ . To this end, writing  $A = ]-\infty, s]$ , observe that

$$\left| \int_A f_n d\mu - \int_A f d\mu \right| \leq \int_{\mathbb{R}} |f_n - f| d\mu \xrightarrow{n \rightarrow \infty} 0,$$

which completes the proof.

**Remark.** The same proof shows that in fact  $\mathbb{P}(X_n \in A) \rightarrow \mathbb{P}(X \in A)$  for every Borel set  $A$ . This implies convergence in distribution, but the converse is false (take, for example,  $\mu_n = \delta_{1/n} \Rightarrow \delta_0$  with  $A = \{0\}$ ).

\* \* \*

**Solution to Exercise 1.5.** [statement p. 98]

(1) On every compact interval  $[a, b] \subset ]0, \infty[$  and for every integer  $k \geq 0$ ,

$$\left| \frac{\partial^k}{\partial t^k} e^{-tu} \right| = u^k e^{-tu} \leq u^k e^{-au}.$$

The upper bound is bounded on  $\mathbb{R}_+$  and hence  $\mu$ -integrable. Differentiation under the integral sign, to every order, gives  $L_\mu \in C^\infty([0, \infty[)$  and

$$L_\mu^{(k)}(t) = (-1)^k \int_{\mathbb{R}_+} u^k e^{-tu} \mu(du).$$

(2) Let  $x > 0$ . By the previous question and Fubini's theorem,

$$\sum_{k=0}^{\lfloor tx \rfloor} \frac{(-t)^k}{k!} L_{\mu}^{(k)}(t) = \int_{\mathbb{R}_+} \sum_{k=0}^{\lfloor tx \rfloor} \frac{(tu)^k}{k!} e^{-tu} \mu(du).$$

Now

$$\sum_{k=0}^{\lfloor tx \rfloor} \frac{(tu)^k}{k!} e^{-tu} = \mathbb{P}(\text{Poisson}(tu) \leq \lfloor tx \rfloor).$$

Let us show that

$$\mathbb{P}(\text{Poisson}(tu) \leq \lfloor tx \rfloor) \xrightarrow{t \rightarrow \infty} \begin{cases} 0 & \text{if } x < u \\ \frac{1}{2} & \text{if } x = u \\ 1 & \text{if } x > u \end{cases}$$

by distinguishing three cases:

- $x = u$ : Let  $(Y_i)_{i \geq 1}$  be i.i.d. Poisson random variables with parameter  $x$ , set  $m = \lfloor t \rfloor$ , and take a Poisson random variable  $Z_t$  with parameter  $(t - m)x$ , independent of the  $Y_i$ . Then

$$\begin{aligned} \mathbb{P}(\text{Poisson}(tx) \leq \lfloor tx \rfloor) \\ = \mathbb{P}\left(\frac{Y_1 + \dots + Y_m - mx + Z_t - (t - m)x}{\sqrt{tx}} \leq \frac{\lfloor tx \rfloor - tx}{\sqrt{tx}}\right). \end{aligned}$$

The CLT gives  $(Y_1 + \dots + Y_m - mx)/\sqrt{tx} \Rightarrow \mathcal{N}(0, 1)$ . Moreover,  $\text{Var}(Z_t) = (t - m)x \leq x$ , so  $(Z_t - (t - m)x)/\sqrt{tx} \rightarrow 0$  in probability, while  $(\lfloor tx \rfloor - tx)/\sqrt{tx} \rightarrow 0$ . Slutsky's lemma and continuity of the normal distribution function at 0 give the limit 1/2.

- $x < u$ : then

$$\begin{aligned} \mathbb{P}(\text{Poisson}(tu) \leq \lfloor tx \rfloor) &= \mathbb{P}(tu - \lfloor tx \rfloor \leq tu - \text{Poisson}(tu)) \\ &\leq \mathbb{P}(tu - \lfloor tx \rfloor \leq |tu - \text{Poisson}(tu)|) \\ &\leq \frac{\text{Var}(\text{Poisson}(tu))}{(tu - \lfloor tx \rfloor)^2} \\ &\sim \frac{u}{t(u - x)^2} \end{aligned}$$

as  $t \rightarrow \infty$ , which therefore tends to 0.

- $x > u$ : similarly,

$$\begin{aligned} \mathbb{P}(\text{Poisson}(tu) \leq \lfloor tx \rfloor) &= 1 - \mathbb{P}(\text{Poisson}(tu) > \lfloor tx \rfloor) \\ &= 1 - \mathbb{P}(\text{Poisson}(tu) - tu > \lfloor tx \rfloor - tu) \\ &\geq 1 - \mathbb{P}(|\text{Poisson}(tu) - tu| > \lfloor tx \rfloor - tu) \end{aligned}$$

and if  $u = 0$ , the probability under consideration is 0. If  $u > 0$ , for sufficiently large  $t$ ,

$$\mathbb{P}(|\text{Poisson}(tu) - tu| > \lfloor tx \rfloor - tu) \leq \frac{\text{Var}(\text{Poisson}(tu))}{(\lfloor tx \rfloor - tu)^2} \sim \frac{u}{t(x - u)^2} \rightarrow 0.$$

Thus,

$$\sum_{k=0}^{\lfloor tx \rfloor} \frac{(tu)^k}{k!} e^{-tu} \xrightarrow{t \rightarrow \infty} \mathbb{1}_{[0, x[}(u) + \frac{1}{2} \mathbb{1}_{\{x\}}(u).$$

Using the dominated convergence theorem, we conclude that

$$\begin{aligned} & \int_{\mathbb{R}_+} \sum_{k=0}^{\lfloor tx \rfloor} \frac{(tu)^k}{k!} e^{-tu} \mu(du) \\ & \xrightarrow{t \rightarrow \infty} \int_{\mathbb{R}_+} \left( \mathbb{1}_{[0, x[}(u) + \frac{1}{2} \mathbb{1}_{\{x\}}(u) \right) \mu(du) \\ & = \mu([0, x]) + \frac{1}{2} \mu(\{x\}). \end{aligned}$$

Now suppose that  $\mu, \nu \in \mathcal{M}_1(\mathbb{R}_+)$  have the same Laplace transform. Write  $D$  for the set of atoms of  $\mu$  and  $\nu$ , which is at most countable. By the previous result,  $\mu([x, y]) = \nu([x, y])$  for all  $x, y \in \mathbb{R} \setminus D$ . Since  $\mathbb{R} \setminus D$  is dense, we deduce that  $\mu$  and  $\nu$  agree on every open interval, and hence  $\mu = \nu$  by the monotone class lemma.

- (3) First observe that if  $\mu_n \Rightarrow \mu$ , then  $L_{\mu_n}$  converges pointwise to  $L_\mu$  (since the function  $x \mapsto e^{-tx}$  is bounded and continuous on  $\mathbb{R}_+$ ).

It suffices to show that  $(\mu_n)_{n \geq 1}$  is tight. Every weak subsequential limit is supported on  $\mathbb{R}_+$  and has Laplace transform  $L$ , since  $x \mapsto e^{-tx}$  is bounded and continuous on  $\mathbb{R}_+$ . Question (2) shows that all these limits are equal. Tightness and uniqueness of subsequential limits therefore give the existence of  $\mu \in \mathcal{M}_1(\mathbb{R}_+)$  such that  $\mu_n \Rightarrow \mu$  and  $L = L_\mu$ .

To prove tightness, we use, for every  $K > 0$ ,

$$(1 - e^{-1})\mu_n([K, \infty[) \leq \int_{\mathbb{R}_+} (1 - e^{-x/K}) \mu_n(dx) = 1 - L_{\mu_n}(1/K).$$

By pointwise convergence of the Laplace transforms,

$$\limsup_{n \rightarrow \infty} \mu_n([K, \infty[) \leq \frac{1 - L(1/K)}{1 - e^{-1}} \rightarrow 0 \quad (K \rightarrow \infty),$$

since  $L(0) = 1$  and  $L$  is right-continuous at zero. We absorb the finitely many initial terms by increasing  $K$ : the sequence  $(\mu_n)_{n \geq 1}$  is therefore tight.

\* \* \*

### Solution to Exercise 1.6. [statement p. 98]

- (1) Let  $\varepsilon > 0$  and choose  $M > 0$  such that for every  $n \geq 1$ ,  $\mu_n(\mathbb{R} \setminus [-M, M]) \leq \varepsilon$ . By uniform continuity, there exists  $\delta > 0$  such that  $|x - y| \leq \delta$  and  $t \in [-M, M]$  imply  $|e^{itx} - e^{ity}| \leq \varepsilon$ . We then write:

$$\begin{aligned} |\phi_{\mu_n}(x) - \phi_{\mu_n}(y)| & \leq \int_{[-M, M]} |e^{itx} - e^{ity}| \mu_n(dt) + \int_{\mathbb{R} \setminus [-M, M]} |e^{itx} - e^{ity}| \mu_n(dt) \\ & \leq \varepsilon + 2\mu_n(\mathbb{R} \setminus [-M, M]) \\ & \leq 3\varepsilon \end{aligned}$$

- (2) We know that  $\phi_{\mu_n}$  converges pointwise to  $\phi_\mu$ . This is then a general result: if a sequence of real-valued functions on a compact set is uniformly equicontinuous and converges pointwise, then it converges uniformly. To prove this in our particular case, let  $\varepsilon > 0$  and let  $K$  be a compact set. Let  $\delta > 0$  be such that the implication in the first question holds. Passing to the limit, first observe that  $|x - y| \leq \delta$  implies  $|\phi_\mu(x) - \phi_\mu(y)| \leq \varepsilon$ . By compactness, we can cover  $K$  by finitely many balls  $(B(x_i, \delta))_{1 \leq i \leq k}$ . Let  $N$  be such that  $n \geq N$  implies  $\max_{1 \leq i \leq k} |\phi_{\mu_n}(x_i) - \phi_\mu(x_i)| \leq \varepsilon$ . Then, for  $n \geq N$ , let  $x \in K$ . Writing  $1 \leq i \leq k$  for an integer such that  $x \in B(x_i, \delta)$ :
- $$|\phi_{\mu_n}(x) - \phi_\mu(x)| \leq |\phi_{\mu_n}(x) - \phi_{\mu_n}(x_i)| + |\phi_{\mu_n}(x_i) - \phi_\mu(x_i)| + |\phi_\mu(x_i) - \phi_\mu(x)| \leq 3\varepsilon,$$
- which completes the proof.

Taking, for example,  $X_n = 1/n$  and  $X = 0$ , we have  $\phi_{X_n}(t) = e^{it/n} \rightarrow \phi_X(t)$  for every  $t \in \mathbb{R}$ , but  $\|\phi_{X_n} - \phi_X\|_\infty = 2$  for every  $n \geq 1$ .

- (3) Yes. The tail bound used in the proof of Lévy's theorem gives, for a universal constant  $c > 0$ ,

$$\begin{aligned} \sup_n \mu_n(\mathbb{R} \setminus [-K, K]) &\leq cK \sup_n \int_{-1/K}^{1/K} (1 - \operatorname{Re} \phi_{\mu_n}(u)) du \\ &\leq 2c \sup_n \sup_{|u| \leq 1/K} |1 - \phi_{\mu_n}(u)| \rightarrow 0 \quad (K \rightarrow \infty). \end{aligned}$$

Indeed,  $\phi_{\mu_n}(0) = 1$  for every  $n$  and the family is equicontinuous at zero. This proves tightness directly.

\* \* \*

**Solution to Exercise 1.7.** [statement p. 99] The converse is clear. The forward implication follows from Dini's second theorem. Here is an approach in the context of this exercise. Fix  $k \geq 2$ . Since the function  $F_\mu$  is continuous and nondecreasing, with limit zero at  $-\infty$  and limit 1 at  $\infty$ , there exist points  $s_1 < \dots < s_{k-1}$  such that  $F_\mu(s_i) = \frac{i}{k}$  for every  $1 \leq i \leq k-1$ . By pointwise convergence, for sufficiently large  $n$ , for every  $1 \leq i \leq k-1$ ,

$$\frac{i-1}{k} \leq F_{\mu_n}(s_i) \leq \frac{i+1}{k}.$$

By convention, set  $s_0 = -\infty$  and  $s_k = \infty$ . It follows that for all sufficiently large  $n$  and every  $s \in \mathbb{R}$ , choosing  $s_i$  such that  $s_i \leq s < s_{i+1}$ :

$$F_{\mu_n}(s) \leq F_{\mu_n}(s_{i+1}) \leq \frac{i+2}{k} = F_\mu(s_i) + \frac{2}{k} \leq F_\mu(s) + \frac{2}{k}$$

and similarly

$$F_{\mu_n}(s) \geq F_{\mu_n}(s_i) \geq \frac{i-1}{k} = F_\mu(s_{i+1}) - \frac{2}{k} \geq F_\mu(s) - \frac{2}{k}.$$

Thus, for sufficiently large  $n$ ,

$$\sup_{s \in \mathbb{R}} |F_{\mu_n}(s) - F_\mu(s)| \leq \frac{2}{k},$$

which completes the proof.

**Solution to Exercise 1.8.** [statement p. 99]

- (1) We follow the same approach as in Exercise 7. Fix  $k \geq 2$ . There exist points  $s_1 < \dots < s_{k-1}$  such that  $F(s_i) = \frac{i}{k}$  for every  $1 \leq i \leq k-1$ . By the law of large numbers, almost surely, for sufficiently large  $n$ , for every  $1 \leq i \leq k-1$ ,

$$\frac{i-1}{k} \leq F_{\mu_n}(s_i) \leq \frac{i+1}{k}.$$

By convention, set  $s_0 = -\infty$  and  $s_k = \infty$ . It follows that for all sufficiently large  $n$  and every  $s \in \mathbb{R}$ , choosing  $s_i$  such that  $s_i \leq s < s_{i+1}$ :

$$F_{\mu_n}(s) \leq F_{\mu_n}(s_{i+1}) \leq \frac{i+2}{k} = F_{\mu}(s_i) + \frac{2}{k} \leq F_{\mu}(s) + \frac{2}{k}$$

and similarly

$$F_{\mu_n}(s) \geq F_{\mu_n}(s_i) \geq \frac{i-1}{k} = F_{\mu}(s_{i+1}) - \frac{2}{k} \geq F_{\mu}(s) - \frac{2}{k}.$$

Thus, almost surely, for sufficiently large  $n$ ,

$$\sup_{s \in \mathbb{R}} |F_{\mu_n}(s) - F(s)| \leq \frac{2}{k}.$$

Since this holds for every  $k \geq 2$ , the proof is complete.

- (2) (a) If  $F(t) \geq x$ , then clearly  $\inf\{s \in \mathbb{R} : F(s) \geq x\} \leq t$ , and hence  $G(x) \leq t$ . Conversely, let us show that  $F(t) < x$  implies  $t < G(x)$ . Since  $F$  is right-continuous, there exists  $\varepsilon > 0$  such that  $F(t + \varepsilon) < x$ . We then have  $G(x) \geq t + \varepsilon > t$ .
- (b) They are independent and identically distributed and, for every  $t \in \mathbb{R}$ , the previous question gives

$$\mathbb{P}(G(Y_1) \leq t) = \mathbb{P}(Y_1 \leq F(t)) = F(t),$$

which proves the result.

- (c) The sequences  $(X_i)_{i \geq 1}$  and  $(G(Y_i))_{i \geq 1}$  have the same distribution. Thus, to study a property of the whole sequence, we may work with  $X_i = G(Y_i)$  for every  $i$ . Then, simultaneously for every  $n$  and every  $t$ ,

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{G(Y_i) \leq t\}} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq F(t)\}} = A_n(F(t)).$$

- (3) By (1), almost surely,

$$\lim_{n \rightarrow \infty} \sup_{u \in [0,1]} |A_n(u) - u| = 0.$$

Since  $F(s) \in [0, 1]$ , we then have almost surely

$$\sup_{s \in \mathbb{R}} |A_n(F(s)) - F(s)| \leq \sup_{u \in [0,1]} |A_n(u) - u| \rightarrow 0.$$

The equality in distribution of the entire sequences established in (c) transfers this almost sure property to  $(F_n)_{n \geq 1}$ . The event is measurable because distribution functions are right-continuous and each supremum over  $\mathbb{R}$  can be taken over  $\mathbb{Q}$ .

**Solution to Exercise 1.9.** [statement p. 99]

- (1) Take  $F(x_1, x_2) = \mathbb{1}_{\{x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \geq 1\}}$ . This function is right-continuous, proper and nondecreasing in each variable. If it were the distribution function of a measure  $\mu \in \mathcal{M}_1(\mathbb{R}^2)$ , we would have

$$\mu([0, 1] \times ]0, 1]) = F(1, 1) - F(1, 0) - F(0, 1) + F(0, 0) = -1,$$

which is impossible.

- (2) The forward implication is clear (if  $F$  is the distribution function of  $\mu \in \mathcal{M}_1(\mathbb{R}^k)$ , nonnegativity of the increments follows from the fact that  $\sum_u s(u)F(u) = \mu([x, y])$  – using the notation of the statement).

Set  $\mathbf{1} = (1, \dots, 1)$ . Since  $F$  is proper and has nonnegative increments, there do exist probability measures  $\mu_n$  supported on  $(2^{-n}\mathbb{Z})^k$  such that  $\mu_n(\{2^{-n}x\}) = F([2^{-n}(x - \mathbf{1}), 2^{-n}x])$  for  $x \in \mathbb{Z}^k$ .

Next, by construction,  $F$  is finitely additive on rectangles, so that for  $x \in \mathbb{Z}^k$  and  $1 \leq m < n$ :

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

Construct  $X^1$  with distribution  $\mu_1$ , and then, conditionally on  $X^n = z$ , choose

$$X^{n+1} = z - 2^{-n-1}i, \quad i \in \{0, 1\}^k,$$

with probability

$$\frac{\mu_{n+1}(\{z - 2^{-n-1}i\})}{\mu_n(\{z\})}.$$

The denominator is strictly positive for  $\mu_n$ -almost every  $z$ . Finite additivity of the increments gives

$$\sum_{i \in \{0, 1\}^k} \mu_{n+1}(\{z - 2^{-n-1}i\}) = \mu_n(\{z\}),$$

so these probabilities do define a transition. By induction, we obtain  $X^n$  with distribution  $\mu_n$  and, coordinatewise,  $X^m - 2^{-m}\mathbf{1} < X^n \leq X^m$  for  $m < n$ . The sequence is nonincreasing and bounded below by  $X^1 - 2^{-1}\mathbf{1}$ , so it converges to a random variable  $X \in \mathbb{R}^k$ .

Let us verify that the distribution function of  $X$  is  $F$ . For a dyadic point  $x \in \mathbb{R}^k$  and sufficiently large  $n$ ,  $\mathbb{P}(X^n \leq x) = F(x)$ . Since  $X^n \downarrow X$  coordinatewise, we obtain

$$\mathbb{P}(X < x) \leq F(x) \leq \mathbb{P}(X \leq x).$$

Applying this inequality also at the dyadic point  $x + 2^{-m}\mathbf{1}$ , we have, for every  $m \geq 1$ ,

$$F(x) \leq \mathbb{P}(X \leq x) \leq \mathbb{P}(X < x + 2^{-m}\mathbf{1}) \leq F(x + 2^{-m}\mathbf{1}).$$

Right-continuity gives  $\mathbb{P}(X \leq x) = F(x)$  at dyadic points. Approximating an arbitrary point by a nonincreasing sequence of dyadic points, we conclude by right-continuity of  $F$  and continuity from above of probability measures.

**Solution to Exercise 1.10.** [statement p. 100] First verify that  $U_{\varepsilon,\delta}$  is indeed open. If  $x \in U_{\varepsilon,\delta}$  and  $y, z \in B(x, \varepsilon)$  satisfy  $|f(y) - f(z)| > \delta$ , then for every  $x'$  such that  $d(x, x') < \varepsilon - \max(d(x, y), d(x, z))$  we have  $x' \in U_{\varepsilon,\delta}$ . In particular,  $U_{\varepsilon,\delta}$  is measurable. We conclude by observing that the set of discontinuity points of  $f$  is

$$\bigcup_{\substack{\delta > 0 \\ \delta \in \mathbb{Q}}} \bigcap_{\substack{\varepsilon > 0 \\ \varepsilon \in \mathbb{Q}}} U_{\varepsilon,\delta}.$$

\* \* \*

**Solution to Exercise 1.11.** [statement p. 100]

- (1) The forward implication is clear. For the converse, we adapt the proof of Lévy's theorem 1.18. As in that proof, there exists a constant  $c > 0$  such that for every  $A > 0$

$$\begin{aligned} \mu_n(\mathbb{R} \setminus [-A, A]) &\leq cA \int_{-1/A}^{1/A} (1 - \operatorname{Re} \phi_{\mu_n}(u)) du \\ &\xrightarrow{n \rightarrow \infty} cA \int_{-1/A}^{1/A} (1 - \operatorname{Re} \phi_{\mu}(u)) du \\ &\xrightarrow{A \rightarrow \infty} 0, \end{aligned}$$

where the first convergence follows from the almost everywhere convergence of  $\phi_{\mu_n}$  to  $\phi_{\mu}$  and the dominated convergence theorem, and the second from continuity of  $\phi_{\mu}$  at 0.

Thus  $(\mu_n)_{n \geq 1}$  is tight. Suppose that  $\mu_{\psi(n)} \Rightarrow \nu$ . By Prokhorov's theorem, it suffices to verify that  $\mu = \nu$ . We then have  $\phi_{\mu_{\psi(n)}}(t) \rightarrow \phi_{\nu}(t)$  for every  $t \in \mathbb{R}$ . But  $\phi_{\mu_{\psi(n)}}(t) \rightarrow \phi_{\mu}(t)$  for almost every  $t \in \mathbb{R}$ . Hence  $\phi_{\mu} = \phi_{\nu}$  almost everywhere. Since these two functions are continuous, they are equal everywhere, which completes the proof.

- (2) This is false; for example, take  $\mu_n = \delta_{2\pi n!}$ .

**Solution to Exercise 1.12.** [statement p. 100]

- (1) Alix is wrong. For example, if  $D = \mathbb{Q}$ ,  $X_n = 1/n$  and  $X = 0$ , then  $X_n$  converges in distribution to  $X$ , but  $\mathbb{P}(X_n \leq 0) \not\rightarrow \mathbb{P}(X \leq 0)$  as  $n \rightarrow \infty$ .
- (2) Billie is right.

*First solution.* Writing  $E$  for the set of functions of the form  $\sum_{i=1}^n \alpha_i \mathbb{1}_{]a_i, b_i]}$  with  $\alpha_i \in \mathbb{R}$  and  $a_i, b_i \in D$  with  $a_i < b_i$ , we have  $\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)]$  for every  $f \in E$ . By uniform continuity, every continuous function with compact support is a uniform limit of functions in  $E$ , which shows that  $X_n$  converges in distribution to  $X$  by Proposition 1.5.

*Second solution.* Write  $F$  and  $F_n$  for the distribution functions of  $X$  and  $X_n$ , respectively. Let  $x$  be a continuity point of  $F$ . By monotonicity, for  $u < x < v$  with  $u, v \in D$  we have:

$$F(x) - F_n(v) \leq F(x) - F_n(x) \leq F(x) - F_n(u).$$

Thus, letting  $n \rightarrow \infty$ :

$$F(x) - F(v) \leq \liminf_{n \rightarrow \infty} (F(x) - F_n(x)) \leq \limsup_{n \rightarrow \infty} (F(x) - F_n(x)) \leq F(x) - F(u).$$

We conclude by continuity of  $F$  at  $x$ , letting  $u, v$  tend to  $x$ .

## 2 Modes of convergence (almost sure, in probability, $L^1$ )

Related lecture notes: [Chapter 2](#), p. 14.

### 2.1 Exercises

**Exercise 2.1.** [solution p. 114] **(Method of moments)** Let  $X$  be a real-valued random variable with moments of all orders. Assume that the distribution of  $X$  is *determined by its moments*, meaning that if  $Y$  is a random variable with moments of all orders satisfying  $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$  for every integer  $k \geq 1$ , then  $X$  and  $Y$  have the same distribution.

- (1) Let  $(Z_n)_{n \geq 1}$  be a uniformly integrable sequence of real-valued random variables converging in distribution to  $Z$ . Show that  $\mathbb{E}[Z_n] \rightarrow \mathbb{E}[Z]$ .
- (2) Let  $(X_n)_{n \geq 1}$  be a sequence of real-valued random variables with moments of all orders. Assume that for every integer  $k \geq 1$ ,  $\mathbb{E}[X_n^k] \rightarrow \mathbb{E}[X^k]$ . Show that  $X_n \Rightarrow X$ .

**Remark.** See Exercise 2.5 for sufficient conditions for a random variable to be determined by its moments.

\* \* \*

**Exercise 2.2.** [solution p. 114] Let  $(E, d)$  be a separable metric space and let  $X, X_1, X_2, \dots$  be random variables defined on the same probability space and taking values in  $E$ . Show that if  $X_n \rightarrow X$  almost surely or in probability, then  $X_n \Rightarrow X$ .

\* \* \*

**Exercise 2.3.** [solution p. 115] Let  $(X_i)_{i \geq 1}$  be independent identically distributed real-valued random variables.

- (1) Show that if  $\mathbb{E}[|X_1|] < \infty$ , the sequence  $(\max(X_1, \dots, X_n)/n)_{n \geq 2}$  is uniformly integrable.
- (2) Is the converse true?

\* \* \*

**Exercise 2.4.** [solution p. 115] **(Skorokhod's representation theorem on  $\mathbb{R}$ )** Let  $X$  be a real-valued random variable. Write  $F$  for its distribution function and set, for  $0 < u < 1$ ,  $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ . Recall (see Exercise Sheet 1) that  $F^{-1}(u) \leq x \iff u \leq F(x)$ . Let  $(X_n)_{n \geq 1}$  be a sequence of real-valued random variables converging in distribution to  $X$ .

- (1) Verify that  $F^{-1}$  is a nondecreasing left-continuous function from  $]0, 1[$  to  $\mathbb{R}$ .
- (2) Show that for every  $u \in ]0, 1[$  we have

$$F^{-1}(u) \leq \liminf_{n \rightarrow \infty} F_n^{-1}(u) \leq \limsup_{n \rightarrow \infty} F_n^{-1}(u) \leq F^{-1}(u+),$$

where  $F^{-1}(u+)$  is the right-hand limit of  $F^{-1}$  at  $u$ .

- (3) Deduce that if  $X, X_1, X_2, \dots$  are real-valued random variables such that  $X_n \Rightarrow X$ , there exist a probability space and random variables  $X', X'_1, X'_2, \dots$  defined on it such that  $X \stackrel{d}{=} X', X_i \stackrel{d}{=} X'_i$  for every  $i \geq 1$ , and  $X'_n$  converges almost surely to  $X'$  as  $n \rightarrow \infty$ .

\* \* \*

**Exercise 2.5.** [solution p. 116] Let  $X$  be a real-valued random variable with moments of all orders. Set  $m_k = \mathbb{E}[X^k]$ ,  $M_k = \mathbb{E}[|X|^k]$  for  $k \geq 1$  and  $\phi(t) = \mathbb{E}[e^{itX}]$  for  $t \in \mathbb{R}$ .

- (1) Show that the power series  $\sum_{k=0}^{\infty} \frac{m_k z^k}{k!}$  has a nonzero radius of convergence if and only if  $\sum_{k=0}^{\infty} \frac{M_k z^k}{k!}$  has a nonzero radius of convergence.

In what follows, assume that these radii of convergence are nonzero, and write  $R$  for the radius of convergence of  $\sum_{k=0}^{\infty} \frac{M_k z^k}{k!}$ .

- (2) Show that for every  $h \in \mathbb{R}$ ,

$$\left| e^{ih} - \sum_{k=0}^K \frac{(ih)^k}{k!} \right| \leq \frac{|h|^{K+1}}{(K+1)!}.$$

- (3) Show that for every  $t \in \mathbb{R}$  and  $|h| < R$  we have the power series expansion  $\phi(t+h) = \sum_{k=0}^{\infty} \frac{\phi^{(k)}(t)}{k!} h^k$ .
- (4) Deduce that the distribution of  $X$  is determined by its moments.

\* \* \*

**Exercise 2.6.** [solution p. 117] Let  $(X_i)_{i \in \mathbb{I}}$  be a uniformly integrable family of random variables. Revisiting the construction of the nondecreasing function  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that  $\lim_{x \rightarrow \infty} \frac{\phi(x)}{x} = \infty$  and  $\sup_{i \in \mathbb{I}} \mathbb{E}[\phi(|X_i|)] < \infty$  (proof of the de la Vallée Poussin criterion), verify that  $\phi$  can be chosen so that for every  $x \geq 1$ ,  $\phi'(2x) \leq 2\phi'(x)$ , and then that  $\phi$  can be chosen so that  $\phi(x) \leq x^2$  for  $x \geq 1$ .

\* \* \*

**Exercise 2.7.** [solution p. 118] Let  $(X_n)_{n \geq 1}$  be a uniformly integrable sequence of real-valued random variables.

- (1) For  $n \geq 1$ , set  $S_n = (X_1 + \dots + X_n)/n$ . Show that the sequence of random variables  $(S_n)_{n \geq 1}$  is uniformly integrable.
- (2) Show that the closure of  $\{X_n : n \geq 1\}$  in  $L^1$  is a uniformly integrable family.

\* \* \*

**Exercise 2.8.** [solution p. 118] Let  $(E, d)$  be a metric space and let  $(x_n)_{n \geq 1}$  be a sequence in  $E$ . Assume that  $(\delta_{x_n})_{n \geq 1}$  converges weakly to the distribution of a random variable  $X$ . What can be said about  $X$ ?

**Exercise 2.9.** [solution p. 118] Let  $(X_n)_{n \geq 1}$  and  $X$  be random variables taking values in a separable metric space. For each mode of convergence you know, determine whether the fact that every subsequence of  $(X_n)_{n \geq 1}$  has a further subsequence converging to  $X$  implies that  $X_n$  converges to  $X$ .

\* \* \*

**Exercise 2.10.** [solution p. 119] Let  $(X_n)$  and  $(Y_n)$  be real-valued random variables such that  $0 \leq X_n \leq Y_n$ . Assume that  $X_n$  converges in probability and that  $Y_n$  converges in  $L^1$ . Show that  $X_n$  converges in  $L^1$ .

\* \* \*

**Exercise 2.11.** [solution p. 119] Let  $p, q > 1$  satisfy  $1/p + 1/q = 1$ . Let  $X \in L^q(\Omega, \mathcal{A}, \mathbb{P})$ , and let  $(Y_i)_{i \in I}$  be a family of random variables bounded in  $L^p$  (that is,  $\sup_{i \in I} \mathbb{E} [|Y_i|^p]^{1/p} < \infty$ ). Prove that the family  $(XY_i)_{i \in I}$  is uniformly integrable.

\* \* \*

**Exercise 2.12.** [solution p. 120] Let  $\lambda > 0$  and let  $X$  be a random variable whose distribution is given by  $\mathbb{P}(X \geq a) = a^{-\lambda}$  for every  $a \geq 1$ . Let  $(X_n)_{n \geq 1}$  be a sequence of i.i.d. random variables with the same distribution as  $X$ , and set

$$T_n = \left( \prod_{i=1}^n X_i \right)^{1/n}.$$

Determine the integers  $n$  for which  $T_n \in L^1$ . Do we have

$$\mathbb{E} [|T_n - e^{1/\lambda}|] \longrightarrow 0 \quad \text{as } n \longrightarrow \infty?$$

Justify your answer.

\* \* \*

**Exercise 2.13.** [solution p. 120] Let  $(X_n)_{n \geq 1}$  and  $(Y_n)_{n \geq 1}$  be two sequences of real-valued random variables defined on the same probability space such that  $Y_n \geq X_n \geq 0$  and  $\mathbb{E}[Y_n] = 1$  for every  $n \geq 1$ . Assume that the sequence of random variables  $(X_n)_{n \geq 1}$  converges in distribution to a nonnegative random variable  $X$  such that  $\mathbb{E}[X] = 1$ .

- (1) Show that the sequence of random variables  $((X_n, Y_n))_{n \geq 1}$  converges in distribution to  $(X, X)$ .
- (2) Show that the sequence of random variables  $(Y_n)_{n \geq 1}$  is uniformly integrable.

\* \* \*

**Exercise 2.14.** [solution p. 121] Let  $(X_n)_{n \geq 1}$  be a sequence of random variables taking values in  $[0, 1]$  and let  $p > 0$ .

- (1) Assume that  $\mathbb{E}[X_n] \rightarrow 1$  as  $n \rightarrow \infty$ . Show that for every  $p > 0$ ,  $\mathbb{E}[|X_n - 1|^p] \rightarrow 0$  as  $n \rightarrow \infty$ .

(2) Assume that  $\mathbb{E}[X_n] \rightarrow \alpha$  and  $\mathbb{E}[X_n^2] \rightarrow \alpha^2$  with  $\alpha \in [0, 1[$  as  $n \rightarrow \infty$ . Show that

$$\mathbb{E}[|X_n - \alpha|^p] \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

*Hint.* In both cases, you may first show that  $X_n$  converges in probability.

\* \* \*

**Exercise 2.15.** [solution p. 121] Let  $(X_i)_{i \geq 1}$  and  $X$  be real-valued random variables. Set  $S_n = (X_1 + \cdots + X_n)/n$  for  $n \geq 1$ . Consider the following assertion:

$$(A) : \quad \text{if } X_n \xrightarrow[n \rightarrow \infty]{} X, \quad \text{then } S_n \xrightarrow[n \rightarrow \infty]{} X.$$

Is assertion (A) true for the following modes of convergence?

**(1)** Almost sure convergence

**(2)**  $L^1$  convergence

**(3)** Convergence in probability

**(4)** Convergence in distribution.

Justify your answers.

\* \* \*

**Exercise 2.16.** [solution p. 122] Let  $(X_n)_{n \geq 1}$  and  $(Y_n)_{n \geq 1}$  be real-valued random variables. Assume that for every  $n \geq 1$ , the distribution of  $X_n$  is absolutely continuous with respect to the distribution of  $Y_n$ , with density  $f_n$ . This means that  $f_n : \mathbb{R} \rightarrow \mathbb{R}_+$  is a measurable function such that for every nonnegative measurable function  $F : \mathbb{R} \rightarrow \mathbb{R}_+$  or every bounded measurable function  $F : \mathbb{R} \rightarrow \mathbb{R}$  we have:

$$\mathbb{E}[F(X_n)] = \mathbb{E}[f_n(Y_n)F(Y_n)].$$

**(1)** Assume that  $f_n(Y_n)$  converges in probability to 1 as  $n \rightarrow \infty$  and that  $Y_n$  converges in distribution to a random variable  $Y$ . Show that  $X_n$  converges in distribution to  $Y$ .

**(2)** Assume that  $f_n$  is continuous for every  $n \geq 1$ , that the sequence  $(f_n)_{n \geq 1}$  converges uniformly on  $\mathbb{R}$  to a function  $f$ , that  $Y_n$  converges almost surely to  $Y$ , and finally that  $\mathbb{E}[f(Y)] = 1$ . Prove that  $X_n$  converges in distribution.

\* \* \*

**Exercise 2.17.** [solution p. 123] Let  $(X_k^n)_{k, n \geq 1}$  and  $(X_k)_{k \geq 1}$  be random variables defined on the same probability space, taking values in a separable metric space  $E$ , such that, for every  $k \geq 1$ , almost surely,  $X_k^n = X_k$  for all sufficiently large  $n$ . Prove that there exists a deterministic sequence of integers  $A_n \rightarrow \infty$  such that

$$\mathbb{P}(\text{for every } k \in \{1, 2, \dots, A_n\} \text{ we have } X_k^n = X_k) \xrightarrow[n \rightarrow \infty]{} 1.$$

## 2.2 Solutions

**Solution to Exercise 2.1.** [statement p. 110]

- (1) *First approach.* By Skorokhod's representation theorem (Exercise 2.4), since uniform integrability depends only on the individual distributions, we may assume that the convergence  $Z_n \rightarrow Z$  holds almost surely (and hence in probability). By Proposition 2.8, this convergence also holds in  $L^1$ , which proves the lemma.

*Second approach.* For  $K \geq 0$ , define  $h_K(x) = \max(-K, \min(x, K))$ . This function is bounded and continuous. Uniform integrability gives  $C = \sup_n \mathbb{E} [|Z_n|] < \infty$ . By convergence in distribution,

$$\mathbb{E} [h_K(|Z|)] = \lim_{n \rightarrow \infty} \mathbb{E} [h_K(|Z_n|)] \leq C.$$

Monotone convergence as  $K \rightarrow \infty$  yields  $\mathbb{E} [|Z|] \leq C$ . Moreover,  $|x - h_K(x)| \leq |x| \mathbb{1}_{\{|x| > K\}}$ , so

$$|\mathbb{E} [Z_n] - \mathbb{E} [Z]| \leq \mathbb{E} [|Z_n| \mathbb{1}_{\{|Z_n| > K\}}] + |\mathbb{E} [h_K(Z_n)] - \mathbb{E} [h_K(Z)]| + \mathbb{E} [|Z| \mathbb{1}_{\{|Z| > K\}}].$$

For fixed  $K$ , the middle term tends to 0. Thus,

$$\limsup_{n \rightarrow \infty} |\mathbb{E} [Z_n] - \mathbb{E} [Z]| \leq \sup_n \mathbb{E} [|Z_n| \mathbb{1}_{\{|Z_n| > K\}}] + \mathbb{E} [|Z| \mathbb{1}_{\{|Z| > K\}}].$$

Letting  $K \rightarrow \infty$ , we conclude by uniform integrability and integrability of  $Z$ .

- (2) We prove tightness and then uniqueness of the limit.

*First step: tightness.* Since  $(\mathbb{E} [X_n^2])_{n \geq 2}$  converges, the sequence  $(X_n)_{n \geq 1}$  is tight.

*Second step: uniqueness of the limit.* Let  $\phi$  index a subsequence and let  $Y$  be a random variable such that  $X_{\phi(n)} \Rightarrow Y$ . Let  $k \geq 1$  be an integer. Since  $(\mathbb{E} [(X_{\phi(n)})^{2k}])_{n \geq 1}$  converges,  $(X_{\phi(n)}^k)_{n \geq 1}$  is uniformly integrable. Let us verify that

$$\mathbb{E} [|Y|^k] < \infty \quad \text{and} \quad \mathbb{E} [X_{\phi(n)}^k] \rightarrow \mathbb{E} [Y^k]. \quad (3)$$

The lemma implies (3), and we obtain  $\mathbb{E} [X^k] = \mathbb{E} [Y^k]$ . Since  $X$  is determined by its moments, we deduce that  $X$  and  $Y$  have the same distribution, which completes the proof.

\* \* \*

**Solution to Exercise 2.2.** [statement p. 110] If  $X_n \rightarrow X$  almost surely and  $f : E \rightarrow \mathbb{R}$  is a bounded continuous function, then  $f(X_n) \rightarrow f(X)$  almost surely, and since  $|f(X_n)| \leq \|f\|_\infty < \infty$ , the dominated convergence theorem implies  $\mathbb{E} [f(X_n)] \rightarrow \mathbb{E} [f(X)]$ .

If  $X_n \rightarrow X$  in probability, let  $f : E \rightarrow \mathbb{R}$  be a bounded continuous function. Argue by contradiction and suppose that  $\mathbb{E} [f(X_n)] \not\rightarrow \mathbb{E} [f(X)]$ . There then exist  $\epsilon > 0$  and a subsequence indexed by  $\phi$  such that  $|\mathbb{E} [f(X_{\phi(n)})] - \mathbb{E} [f(X)]| \geq \epsilon$  for every  $n \geq 1$ . Since  $X_{\phi(n)}$  converges in probability to  $X$ , there exists a further subsequence indexed by  $\psi$  such that  $X_{\phi(\psi(n))} \rightarrow X$  almost surely. Then  $f(X_{\phi(\psi(n))}) \rightarrow f(X)$  almost surely. Since  $|f(X_{\phi(\psi(n))})| \leq \|f\|_\infty$ , dominated convergence gives  $\mathbb{E} [f(X_{\phi(\psi(n))})] \rightarrow \mathbb{E} [f(X)]$ , a contradiction.

**Solution to Exercise 2.3.** [statement p. 110]

- (1) If
- $\mathbb{E} [|X_1|] < \infty$
- , we write

$$\left| \frac{\max(X_1, \dots, X_n)}{n} \right| \leq \frac{\sum_{k=1}^n |X_k|}{n}.$$

By the law of large numbers, the sequence  $\frac{\sum_{k=1}^n |X_k|}{n}$  converges in  $L^1$  to  $\mathbb{E} [|X_1|]$ . It is therefore uniformly integrable, which implies that the sequence  $(\frac{\max(X_1, \dots, X_n)}{n})_{n \geq 1}$  is uniformly integrable.

- (2) Yes, if the random variables are nonnegative: in that case,
- $0 \leq X_1/2 \leq \max(X_1, X_2)/2$
- , and hence
- $\mathbb{E} [X_1] < \infty$
- .

Not in general: for example, taking  $X_1$  to be a random variable with density  $\frac{1}{|x|^2} \mathbb{1}_{x \leq -1}$ , we see that  $\max(X_1, \dots, X_n)$  has density  $\frac{n}{|x|^{n+1}} \mathbb{1}_{x \leq -1}$ , so that

$$\mathbb{E} \left[ \left| \frac{\max(X_1, \dots, X_n)}{n} \right|^{3/2} \right] = \frac{1}{n^{3/2}} \int_1^\infty \frac{nx^{3/2}}{x^{n+1}} dx = \frac{1}{n^{3/2}} \frac{2n}{2n-3}.$$

Since the sequence  $(\frac{\max(X_1, \dots, X_n)}{n})_{n \geq 2}$  is bounded in  $L^{3/2}$ , it is uniformly integrable, although  $X_1$  is not integrable.

\* \* \*

**Solution to Exercise 2.4.** [statement p. 110]

- (1) To prove that
- $F^{-1}$
- is nondecreasing, if
- $u < v$
- and
- $F(x) \geq v$
- , then
- $F(x) \geq u$
- , so that
- $F^{-1}(u) \leq x$
- ; taking the infimum over
- $x$
- gives
- $F^{-1}(u) \leq F^{-1}(v)$
- .

For left-continuity, suppose that  $u_n \uparrow u$  but  $F^{-1}(u_n) \not\rightarrow F^{-1}(u)$ . There then exists  $\varepsilon > 0$  such that  $F^{-1}(u_n) \leq F^{-1}(u) - \varepsilon$  for every  $n \geq 1$ . Hence  $u_n \leq F(F^{-1}(u) - \varepsilon)$ . Passing to the limit,  $u \leq F(F^{-1}(u) - \varepsilon)$ , and hence  $F^{-1}(u) \leq F^{-1}(u) - \varepsilon$ , a contradiction.

- (2) Let
- $0 < u < 1$
- .

To show that  $\limsup_{n \rightarrow \infty} F_n^{-1}(u) \leq F^{-1}(u+)$ , it suffices to verify that

$$F(x) \geq u + \varepsilon \implies \limsup_{n \rightarrow \infty} F_n^{-1}(u) \leq x.$$

We prove the contrapositive: suppose there exist a subsequence indexed by  $\phi$  and  $\eta > 0$  such that  $x + \eta \leq F_{\phi(n)}^{-1}(u)$  for sufficiently large  $n$ . Choose a point  $x_0 \in [x, x + \eta[$  at which  $F$  is continuous. Then  $F_{\phi(n)}(x_0) < u$  for sufficiently large  $n$ , and convergence in distribution of  $X_n$  to  $X$  therefore gives  $F(x) \leq F(x_0) \leq u < u + \varepsilon$ .

To show that

$$F^{-1}(u) \leq \liminf_{n \rightarrow \infty} F_n^{-1}(u),$$

consider a continuity point  $x$  of  $F$  such that  $x > \liminf_{n \rightarrow \infty} F_n^{-1}(u)$ . There then exists a subsequence indexed by  $\phi$  such that  $x \geq F_{\phi(n)}^{-1}(u)$  for every  $n \geq 1$ . Thus,  $F_{\phi(n)}(x) \geq u$ , and passing to the limit gives  $F(x) \geq u$ , hence  $x \geq F^{-1}(u)$ . This shows that  $F^{-1}(u) \leq \liminf_{n \rightarrow \infty} F_n^{-1}(u)$ .

- (3) Let  $U$  be a random variable uniformly distributed on  $]0, 1[$ . Set  $X' = F^{-1}(U)$  and  $X'_n = F_n^{-1}(U)$ . By Exercise Sheet 1,  $X \stackrel{d}{=} X'$  and  $X_i \stackrel{d}{=} X'_i$  for every  $i \geq 1$ . By the previous question,  $X'_n(\omega) \rightarrow X'(\omega)$  if  $F^{-1}$  is continuous at  $U(\omega)$ . Since  $F^{-1}$  has at most countably many discontinuities, it follows that  $F^{-1}$  is almost surely continuous at  $U$ , and the desired result follows.

\* \* \*

### Solution to Exercise 2.5. [statement p. 111]

- (1) The implication  $\boxed{\Leftarrow}$  is clear because  $|m_k| \leq M_k$ . For the implication  $\boxed{\Rightarrow}$ , the idea is to show that the radii of convergence of

$$\sum_{\substack{k=0 \\ k \text{ even}}}^{\infty} \frac{M_k z^k}{k!} \quad \text{and} \quad \sum_{\substack{k=0 \\ k \text{ odd}}}^{\infty} \frac{M_k z^k}{k!}$$

are nonzero. For the first, this follows from  $m_k = M_k$  for even  $k$ , and for the second from the inequality  $|x|^{2j-1} \leq 1 + x^{2j}$  for every  $x \in \mathbb{R}$  and  $j \geq 1$ .

- (2) Taylor's formula with integral remainder gives, for  $h \in \mathbb{R}$  and  $K \geq 1$ :

$$e^{ih} = \sum_{k=0}^K \frac{(ih)^k}{k!} + \frac{(ih)^{K+1}}{K!} \int_0^1 (1-u)^K e^{iuh} du,$$

so that

$$\left| e^{ih} - \sum_{k=0}^K \frac{(ih)^k}{k!} \right| \leq \frac{|h|^{K+1}}{K!} \int_0^1 (1-u)^K du = \frac{|h|^{K+1}}{(K+1)!}.$$

- (3) By the previous question, we may write

$$e^{i(t+h)x} = \sum_{k=0}^K \frac{(ihx)^k}{k!} e^{itx} + R_K(x, h)$$

with  $|R_K(x, h)| \leq \frac{|hx|^{K+1}}{(K+1)!}$ . In particular,

$$\phi(t+h) = \sum_{k=0}^K \frac{\mathbb{E}[(iX)^k e^{itX}]}{k!} h^k + \mathbb{E}[R_K(X, h)]$$

Now  $\phi^{(k)}(t) = \mathbb{E}[(iX)^k e^{itX}]$ , and by definition of  $R$ , if  $|h| < R$ , then  $\frac{|h|^{K+1} M_{K+1}}{(K+1)!} \rightarrow 0$ , and the result follows.

- (4) Since  $\phi$  determines the distribution of  $X$ , it suffices to show that  $\phi$  can be expressed in terms of the moments of  $X$ . By the previous question, taking  $t = 0$ ,  $\phi$  can be expressed in terms of the moments of  $X$  on  $] - R, R[$ . Then, expanding again as a power series around points close to  $R$  and  $-R$ , we deduce that  $\phi$  can be expressed in terms of the moments of  $X$  on  $] - 2R, 2R[$ , and then on  $] - nR, nR[$  for every  $n \geq 1$  by induction, which proves the result.

**Remark.** It is possible to find random variables  $X$  and  $Y$  with moments of all orders such that  $\mathbb{E}[X^k] = \mathbb{E}[Y^k]$  for every  $k \geq 1$  but whose distributions differ. An explicit example is given by the densities

$$\frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{(\ln x)^2}{2}\right) \mathbb{1}_{x>0} \quad \text{and} \quad (1 + \sin(2\pi \ln x)) \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{(\ln x)^2}{2}\right) \mathbb{1}_{x>0}.$$

\* \* \*

**Solution to Exercise 2.6.** [statement p. 111] By uniform integrability, choose thresholds  $(x_m)_{m \geq 1}$  such that  $x_m \geq 2m$ ,  $x_{m+1} > 4x_m$  and

$$\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| \geq x_m\}}] \leq 2^{-m}.$$

Set  $h(t) = 0$  for  $t \leq 1$ ,  $h(t) = t - 1$  for  $1 \leq t \leq 2$ , and  $h(t) = 1$  for  $t \geq 2$ , and then

$$g(x) = 1 + \sum_{m \geq 1} h(x/x_m), \quad \phi(x) = \int_0^x g(u) du.$$

The sums are locally finite. Thus  $g$  is continuous and nondecreasing and tends to infinity, and  $\phi$  is of class  $C^1$ , nondecreasing and convex, with  $\phi(x)/x \rightarrow \infty$ . Moreover,

$$\phi(x) \leq x + \sum_{m \geq 1} x \mathbb{1}_{\{x \geq x_m\}},$$

so  $\sup_i \mathbb{E}[\phi(|X_i|)] \leq \sup_i \mathbb{E}[|X_i|] + \sum_m 2^{-m} < \infty$ . A term  $h(2x/x_m) - h(x/x_m)$  can be nonzero only if  $x/2 < x_m < 2x$ . The spacing of the thresholds ensures that there is at most one such term. Consequently,

$$\phi'(2x) = g(2x) \leq g(x) + 1 \leq 2g(x) = 2\phi'(x).$$

Finally,  $g(x) = 1$  on  $[0, 2]$ , and for  $x \geq 2$ ,  $g(x) \leq 1 + \#\{m : 2m < x\} \leq 1 + x/2 \leq x$ . Thus, for  $x \geq 1$ ,

$$\phi(x) = 1 + \int_1^x g(u) du \leq 1 + \int_1^x u du = \frac{x^2 + 1}{2} \leq x^2.$$

**Solution to Exercise 2.7.** [statement p. 111] For both questions, we use the “ $\varepsilon$ - $\delta$ ” characterization of uniform integrability. First, since  $(X_n)_{n \geq 1}$  is uniformly integrable, it is bounded in  $L^1$ : there exists  $C > 0$  such that  $\mathbb{E} [|X_n|] \leq C$  for every  $n \geq 1$ . Also, for fixed  $\varepsilon > 0$ , there exists  $\delta > 0$  such that for every event  $A$  with  $\mathbb{P}(A) \leq \delta$  we have  $\mathbb{E} [|X_n| \mathbb{1}_A] \leq \varepsilon$  for every  $n \geq 1$ .

- (1) By the triangle inequality and linearity of expectation, for every  $n \geq 1$  we have  $\mathbb{E} [|S_n|] \leq C$ , and for every event  $A$  with  $\mathbb{P}(A) \leq \delta$  we have

$$\mathbb{E} [|S_n| \mathbb{1}_A] \leq \frac{1}{n} \sum_{k=1}^n \mathbb{E} [|X_k| \mathbb{1}_A] \leq \varepsilon,$$

which shows that the sequence of random variables  $(S_n)_{n \geq 1}$  is uniformly integrable.

- (2) Let  $X$  belong to the closure of  $\{X_n : n \geq 1\}$  in the  $L^1$  norm. There exists a sequence  $(Z_j)_{j \geq 1}$  of elements of this set such that  $\mathbb{E} [|Z_j - X|] \rightarrow 0$ . For every event  $A$  with  $\mathbb{P}(A) \leq \delta$ ,

$$\mathbb{E} [|X|] \leq \mathbb{E} [|X - Z_j|] + C, \quad \mathbb{E} [|X| \mathbb{1}_A] \leq \mathbb{E} [|X - Z_j|] + \varepsilon.$$

Letting  $j \rightarrow \infty$ , we obtain  $\mathbb{E} [|X|] \leq C$  and  $\mathbb{E} [|X| \mathbb{1}_A] \leq \varepsilon$ . The same constants  $C$  and  $\delta$  work for the entire closure, which is therefore uniformly integrable.

\* \* \*

**Solution to Exercise 2.8.** [statement p. 111] Write  $\mu = \mathcal{L}(X)$  and, for  $N \geq 1$ ,

$$F_N = \overline{\{x_n : n \geq N\}}.$$

The Portmanteau theorem gives  $\mu(F_N) = 1$ , since  $\delta_{x_n}(F_N) = 1$  for  $n \geq N$ . Thus  $\mu(\bigcap_N F_N) = 1$ : this intersection is nonempty. Choose  $x$  in this intersection. By definition of the  $F_N$ , there exists a subsequence  $x_{n_j} \rightarrow x$ . We then have  $\delta_{x_{n_j}} \Rightarrow \delta_x$ , hence  $\mu = \delta_x$  by uniqueness of the weak limit. Thus  $X = x$  almost surely. Finally, for every  $\varepsilon > 0$ , the Portmanteau theorem gives  $\liminf_n \mathbb{1}_{\{x_n \in B(x, \varepsilon)\}} \geq 1$ , so  $x_n \rightarrow x$ .

\* \* \*

**Solution to Exercise 2.9.** [statement p. 112] This is true for modes of convergence defined through a metric: thus it holds for  $L^p$  convergence, convergence in distribution and convergence in probability;  $X_n \rightarrow X$  in probability iff  $\mathbb{E} [\min(d(X_n, X), 1)] \rightarrow 0$ .

However, it is false for almost sure convergence. For example, if  $(X_n)_{n \geq 1}$  is a sequence of independent Bernoulli random variables such that  $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = 1/n$ , the sequence  $(X_n)_{n \geq 1}$  almost surely takes the values 0 and 1 infinitely often by Borel–Cantelli. But clearly, from every subsequence indexed by  $\phi$ , extracting a further subsequence indexed by  $\psi$  such that  $\phi \circ \psi(n) \geq n^2$  yields a sequence  $(X_{\phi \circ \psi(n)})_{n \geq 1}$  converging almost surely to 0.

**Solution to Exercise 2.10.** [statement p. 112] Since  $Y_n$  converges in  $L^1$ , the family  $(Y_n)$  is uniformly integrable. For every  $A > 0$ , the domination  $0 \leq X_n \leq Y_n$  gives

$$\sup_n \mathbb{E} [X_n \mathbb{1}_{\{X_n \geq A\}}] \leq \sup_n \mathbb{E} [Y_n \mathbb{1}_{\{Y_n \geq A\}}] \rightarrow 0 \quad (A \rightarrow \infty).$$

The family  $(X_n)$  is therefore uniformly integrable. Its convergence in probability implies its convergence in  $L^1$  by Proposition 2.8.

\* \* \*

**Solution to Exercise 2.11.** [statement p. 112] *First solution.* We use the  $\varepsilon - \delta$  characterization of uniform integrability. First, the family  $(XY_i)_{i \in I}$  is bounded in  $L^1$ , since by Hölder's inequality:

$$\mathbb{E} [|XY_i|] \leq \mathbb{E} [|X|^q]^{1/q} \mathbb{E} [|Y_i|^p]^{1/p}$$

and  $(Y_i)_{i \in I}$  is bounded in  $L^p$ .

Next, let  $\varepsilon > 0$  and  $C = \sup_{i \in I} \mathbb{E} [|Y_i|^p]^{1/p}$ . If  $C = 0$ , all products  $XY_i$  vanish almost surely. Otherwise, since  $|X|^q \in L^1$ , there exists  $\delta > 0$  such that  $\mathbb{P}(A) \leq \delta$  implies  $\mathbb{E} [|X|^q \mathbb{1}_A] \leq \varepsilon^q / C^q$ . By Hölder's inequality, we then have for every  $i \in I$ :

$$\mathbb{E} [|XY_i| \mathbb{1}_A] \leq \mathbb{E} [|X|^q \mathbb{1}_A]^{1/q} \mathbb{E} [|Y_i|^p]^{1/p} \leq \varepsilon,$$

which proves the result.

*Second solution.* Let  $i \in I$ . We first use the inequality  $ab \leq a^p/p + b^q/q$  for  $a, b \geq 0$ :

$$\begin{aligned} \mathbb{E} [|XY_i| \mathbb{1}_{|XY_i| > A}] &\leq \mathbb{E} [|XY_i| \mathbb{1}_{|X|^q/q + |Y_i|^p/p > A}] \\ &\leq \mathbb{E} [|XY_i| (\mathbb{1}_{|X|^q/q > A/2} + \mathbb{1}_{|Y_i|^p/p > A/2})] \\ &= \mathbb{E} [|Y_i| |X| \mathbb{1}_{|X|^q/q > A/2}] + \mathbb{E} [|X| |Y_i| \mathbb{1}_{|Y_i|^p/p > A/2}] \end{aligned}$$

To control the first term, we use Hölder's inequality:

$$\mathbb{E} [|Y_i| |X| \mathbb{1}_{|X|^q/q > A/2}] \leq \mathbb{E} [|Y_i|^p]^{1/p} \mathbb{E} [|X|^q \mathbb{1}_{|X|^q/q > A/2}]^{1/q}.$$

The first term is bounded uniformly in  $i$ , and the second term tends to 0 by dominated convergence, since  $X \in L^q(\Omega, \mathcal{A}, \mathbb{P})$ .

Controlling  $\mathbb{E} [|X| |Y_i| \mathbb{1}_{|Y_i|^p/p > A/2}]$  is slightly more delicate. For  $R > 0$ , using Hölder's inequality, we write

$$\begin{aligned} \mathbb{E} (|XY_i| \mathbb{1}_{|Y_i|^p/p > A/2}) &\leq R \mathbb{E} (|Y_i| \mathbb{1}_{|Y_i|^p/p > A/2}) + \mathbb{E} (|XY_i| \mathbb{1}_{|X| > R}) \\ &\leq R \mathbb{E} (|Y_i| \mathbb{1}_{|Y_i|^p/p > A/2}) + \|X \mathbb{1}_{|X| > R}\|_q \|Y_i\|_p. \end{aligned}$$

Since the family  $(Y_i)_{i \in I}$  is uniformly integrable and bounded in  $L^p$ , we have

$$\sup_{i \in I} \mathbb{E} (|XY_i| \mathbb{1}_{|Y_i|^p/p > A/2}) \leq \underbrace{R \sup_{i \in I} \mathbb{E} (|Y_i| \mathbb{1}_{|Y_i|^p/p > A/2})}_{\rightarrow 0 \text{ as } A \rightarrow \infty} + \|X \mathbb{1}_{|X| > R}\|_q \underbrace{\sup_{i \in I} \|Y_i\|_p}_{< \infty}.$$

Finally, as above,

$$\|X \mathbb{1}_{|X| > R}\|_q \xrightarrow{R \rightarrow \infty} 0$$

by dominated convergence. This completes the proof.

**Solution to Exercise 2.12.** [statement p. 112] By independence,  $\mathbb{E}[T_n] = \mathbb{E}[X^{1/n}]^n$ , and computing the moment gives

$$\mathbb{E}[T_n] = \begin{cases} \left(\frac{\lambda n}{\lambda n - 1}\right)^n, & \lambda n > 1, \\ +\infty, & \lambda n \leq 1. \end{cases}$$

Thus,  $T_n \in L^1$  precisely when  $n > 1/\lambda$ . The requested convergence holds for every  $\lambda > 0$ ; it therefore concerns a tail of the sequence.

By computing its distribution function, for example, we verify that  $\ln(X)$  has an exponential distribution with parameter  $\lambda$ , so that  $T_n$  converges almost surely to  $e^{1/\lambda}$  by the strong law of large numbers.

The formula already obtained gives

$$\mathbb{E}[T_n] = \left(1 - \frac{1}{\lambda n}\right)^{-n} \rightarrow e^{1/\lambda}.$$

On the tail  $n > 1/\lambda$ , the random variables  $T_n$  are nonnegative and integrable. Since  $T_n \rightarrow e^{1/\lambda}$  almost surely, Scheffé's lemma gives directly

$$\mathbb{E}[|T_n - e^{1/\lambda}|] \rightarrow 0.$$

\* \* \*

**Solution to Exercise 2.13.** [statement p. 112]

- (1) Since the sequence of random variables  $(Y_n)_{n \geq 1}$  is bounded in  $L^1$ , it is tight. The sequence of random variables  $(X_n)_{n \geq 1}$  is also tight (by Prokhorov's theorem, since it converges in distribution). The sequence of random variables  $((X_n, Y_n))_{n \geq 1}$  is therefore tight. Thus there exist a subsequence indexed by  $\varphi$  and a random variable  $(U, V)$  such that  $(X_{\varphi(n)}, Y_{\varphi(n)}) \rightarrow (U, V)$  in distribution, with  $U = X$  in distribution. By Skorokhod's representation theorem, we may assume that this convergence holds almost surely. We also have  $X_{\varphi(n)} \leq Y_{\varphi(n)}$  almost surely (this property depends only on the distribution of  $(X_{\varphi(n)}, Y_{\varphi(n)})$ ). In particular,  $U \leq V$  almost surely. By Fatou's lemma, we have  $\mathbb{E}[V] \leq 1$ , and since  $\mathbb{E}[U] = 1$ , we deduce that  $\mathbb{E}[V] = 1$ . Thus  $\mathbb{E}[U] = \mathbb{E}[V]$  with  $U \leq V$  almost surely, which implies  $U = V$  almost surely. The same argument applies to every convergent subsequence: every subsequential limit therefore has the distribution of  $(X, X)$ . Tightness then implies convergence of the entire sequence to this distribution.

*Alternative solution.* Convergence in distribution and nonnegativity give  $1 = \mathbb{E}[X] \leq \liminf_n \mathbb{E}[X_n]$ , while  $\mathbb{E}[X_n] \leq \mathbb{E}[Y_n] = 1$ . Hence  $\mathbb{E}[X_n] \rightarrow 1$ . By Markov's inequality, for every  $\varepsilon > 0$ ,

$$\mathbb{P}(|Y_n - X_n| > \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}[Y_n - X_n] \rightarrow 0.$$

By Slutsky's lemma,  $(Y_n - X_n, X_n) \Rightarrow (0, X)$ . Continuity of  $(a, b) \mapsto (b, a + b)$  then gives  $(X_n, Y_n) \Rightarrow (X, X)$ .

- (2) By Skorokhod's representation theorem, we may assume that  $Y_n$  converges almost surely to  $X$ . Since  $\mathbb{E}[Y_n] \rightarrow 1 = \mathbb{E}[X]$  (by the proof of (1)), Scheffé's lemma yields  $Y_n \rightarrow X$  in  $L^1$ , which implies, by a result from the lectures, that the sequence  $(Y_n)_{n \geq 1}$  is uniformly integrable.

\* \* \*

**Solution to Exercise 2.14.** [statement p. 112] In both cases, first show that  $X_n$  converges in probability to  $\alpha$  (setting  $\alpha = 1$  in the first question).

Case (1): we have

$$\mathbb{P}(|X_n - 1| \geq \varepsilon) = \mathbb{P}(1 - X_n \geq \varepsilon) \leq \frac{1}{\varepsilon}(1 - \mathbb{E}[X_n]) \xrightarrow[n \rightarrow \infty]{} 0.$$

Case (2): we have

$$\begin{aligned} \mathbb{P}(|X_n - \alpha| > \varepsilon) &= \mathbb{P}((X_n - \alpha)^2 > \varepsilon^2) \leq \frac{\mathbb{E}[(X_n - \alpha)^2]}{\varepsilon^2} \\ &= \frac{\mathbb{E}[X_n^2] - 2\alpha\mathbb{E}[X_n] + \alpha^2}{\varepsilon^2} \rightarrow \frac{\alpha^2 - 2\alpha^2 + \alpha^2}{\varepsilon^2} = 0. \end{aligned}$$

To conclude, observe that the sequence  $(\mathbb{E}[|X_n - \alpha|^p])_{n \geq 1}$  is bounded; it suffices to show that 0 is its only subsequential limit. Let  $\phi$  index a subsequence such that  $(\mathbb{E}[|X_{\phi(n)} - \alpha|^p])_{n \geq 1}$  converges. Since  $X_n$  converges in probability to  $\alpha$ , there exists a further subsequence indexed by  $\psi$  such that  $X_{\phi \circ \psi(n)}$  converges almost surely to  $\alpha$ . By dominated convergence,  $(\mathbb{E}[|X_{\phi \circ \psi(n)} - \alpha|^p])_{n \geq 1}$  converges to 0, which completes the proof. Alternatively, one may use the Vitali convergence theorem, observing that  $|X_n - \alpha|^p$  is uniformly integrable because it is bounded.

\* \* \*

**Solution to Exercise 2.15.** [statement p. 113]

- (1) True by Cesàro's lemma.  
 (2) True, for example by writing

$$\mathbb{E}[|S_n - X|] \leq \frac{\sum_{k=1}^n \mathbb{E}[|X_k - X|]}{n}$$

and using Cesàro's lemma.

- (3) False, for example if  $(X_n)_{n \geq 1}$  are independent random variables with  $\mathbb{P}(X_n = n) = \frac{1}{n}$  and  $\mathbb{P}(X_n = 0) = 1 - 1/n$ . Then  $X_n \rightarrow 0$  in probability. Let us verify that  $S_n \not\rightarrow 0$  in probability. For  $\varepsilon \in ]0, 1/2[$ , observe that

$$\mathbb{P}(S_{2n} \geq \varepsilon) \geq \mathbb{P}(\exists k \in \{n, n+1, \dots, 2n\} : X_k = k).$$

For  $n \geq 2$ , by independence,

$$\mathbb{P}(\exists k \in \{n, \dots, 2n\} : X_k = k) = 1 - \prod_{k=n}^{2n} \left(1 - \frac{1}{k}\right) = 1 - \frac{n-1}{2n} = \frac{n+1}{2n} \rightarrow \frac{1}{2}.$$

Thus  $\mathbb{P}(S_{2n} \geq \varepsilon) \not\rightarrow 0$ .

- (4) False, for example by taking  $X_{2n} = -X_{2n+1} = N$  where  $N$  has distribution  $\mathcal{N}(0, 1)$ .

**Solution to Exercise 2.16.** [statement p. 113]

- (1) Let  $F : \mathbb{R} \rightarrow \mathbb{R}$  be bounded and continuous, and set  $g_n = f_n(Y_n)$ . We have  $g_n \geq 0$  and  $\mathbb{E}[g_n] = 1$ . For every  $\varepsilon > 0$ ,

$$\mathbb{E}[|g_n - 1|] = 2\mathbb{E}[(1 - g_n)_+] \leq 2\varepsilon + 2\mathbb{P}(|g_n - 1| \geq \varepsilon).$$

Convergence in probability therefore gives  $\mathbb{E}[|g_n - 1|] \rightarrow 0$ . By the change-of-measure identity,

$$|\mathbb{E}[F(X_n)] - \mathbb{E}[F(Y_n)]| = |\mathbb{E}[(g_n - 1)F(Y_n)]| \leq \|F\|_\infty \mathbb{E}[|g_n - 1|] \rightarrow 0.$$

Since  $Y_n \Rightarrow Y$ , we conclude that  $X_n \Rightarrow Y$ .

*Alternative solution.* Let  $F$  be closed,  $0 < \varepsilon < 1$ ,  $g_n = f_n(Y_n)$  and  $B_n = \{|g_n - 1| > \varepsilon\}$ . The change-of-measure identity gives

$$\mathbb{P}(X_n \in F) = \mathbb{E}[g_n \mathbb{1}_{\{Y_n \in F\}}] \leq (1 + \varepsilon)\mathbb{P}(Y_n \in F) + \mathbb{E}[g_n \mathbb{1}_{B_n}].$$

Since  $\mathbb{E}[g_n] = 1$  and  $g_n \geq 1 - \varepsilon$  on  $B_n^c$ ,

$$\mathbb{E}[g_n \mathbb{1}_{B_n}] = 1 - \mathbb{E}[g_n \mathbb{1}_{B_n^c}] \leq \varepsilon + \mathbb{P}(B_n).$$

Thus, by convergence in probability and the Portmanteau theorem,

$$\limsup_n \mathbb{P}(X_n \in F) \leq (1 + \varepsilon)\mathbb{P}(Y \in F) + \varepsilon.$$

Let  $\varepsilon$  tend to 0 to conclude.

*Alternative solution.* Let  $F : \mathbb{R} \rightarrow \mathbb{R}$  be bounded and continuous. Let  $\varepsilon > 0$ . Since  $\mathbb{E}[F(Y_n)] \rightarrow \mathbb{E}[F(Y)]$ , it suffices to show that  $|\mathbb{E}[F(X_n)] - \mathbb{E}[F(Y_n)]| \rightarrow 0$ . To do so, we write

$$|\mathbb{E}[F(X_n)] - \mathbb{E}[F(Y_n)]| = |\mathbb{E}[(f_n(Y_n) - 1)F(Y_n)]| \leq \|F\|_\infty \mathbb{E}[|f_n(Y_n) - 1|].$$

It therefore suffices to prove that  $f_n(Y_n) \rightarrow 1$  in  $L^1$ . This follows from Skorokhod's representation theorem combined with Scheffé's lemma, since  $\mathbb{E}[f_n(Y_n)] = 1$ .

- (2) The function  $f$  is continuous as a uniform limit of continuous functions. Set  $g_n = f_n(Y_n)$  and  $g = f(Y)$ . Since  $Y_n \rightarrow Y$  almost surely,

$$|g_n - g| \leq \|f_n - f\|_\infty + |f(Y_n) - f(Y)| \rightarrow 0 \quad \text{almost surely.}$$

Moreover,  $g_n, g \geq 0$  and  $\mathbb{E}[g_n] = \mathbb{E}[g] = 1$ . Scheffé's lemma therefore gives  $\mathbb{E}[|g_n - g|] \rightarrow 0$ . For every bounded continuous function  $F : \mathbb{R} \rightarrow \mathbb{R}$ , the change-of-measure identity yields

$$\begin{aligned} |\mathbb{E}[F(X_n)] - \mathbb{E}[gF(Y)]| &= |\mathbb{E}[g_n F(Y_n)] - \mathbb{E}[gF(Y)]| \\ &\leq \|F\|_\infty \mathbb{E}[|g_n - g|] + \mathbb{E}[|g(F(Y_n) - F(Y))|] \rightarrow 0. \end{aligned}$$

The last term tends to zero by dominated convergence, with integrable bound  $2\|F\|_\infty g$ . Thus,  $X_n$  converges in distribution to the probability measure with density  $f$  with respect to the distribution of  $Y$ .

**Solution to Exercise 2.17.** [statement p. 113] By assumption, for every  $k \geq 1$

$$\mathbb{P} \left( \bigcup_{n_0 \geq 1} \bigcap_{n \geq n_0} \{X_k^n = X_k\} \right) = 1.$$

Since the union over  $n_0$  is increasing, we therefore have  $\mathbb{P}(X_k^n = X_k) \rightarrow 1$  as  $n \rightarrow \infty$ , for every  $k \geq 1$ .

There therefore exists a sequence  $(i_M)_{M \geq 1}$  of integers, which we may assume to be strictly increasing, such that for every  $M \geq 1$  and  $n \geq i_M$  we have

$$\sum_{k=1}^M \mathbb{P}(X_k^n \neq X_k) \leq \frac{1}{M}.$$

For  $1 \leq n < i_1$ , set  $A_n = 1$ . For  $n \geq i_1$ , set  $A_n = M$  if  $i_M \leq n < i_{M+1}$ . Let us verify that this sequence works. First, clearly  $A_n \rightarrow \infty$ . We write

$$\begin{aligned} \mathbb{P}(\{\text{for every } k \in \{1, 2, \dots, A_n\} \text{ we have } X_k^n = X_k\}^c) &\leq \sum_{k=1}^{A_n} \mathbb{P}(X_k^n \neq X_k) \\ &\leq \frac{1}{A_n}, \end{aligned}$$

which tends to 0.

### 3 Probability measures on a metric space

Related lecture notes: [Chapter 3](#), p. 18.

#### 3.1 Exercises

**Exercise 3.1.** [solution p. 133] Let  $(x_n)_{n \geq 1}$  be a sequence of real numbers and  $\mu \in \mathcal{M}_1(\mathbb{R})$  such that  $\delta_{x_n} \Rightarrow \mu$ . Show that there exists  $x \in \mathbb{R}$  such that  $\mu = \delta_x$ .

\* \* \*

**Exercise 3.2.** [solution p. 133] Let  $(E, d)$  be a metric space,  $x \in E$  a fixed element and  $(X_n)_{n \geq 1}$  random variables taking values in  $E$ . Show that  $X_n \Rightarrow x$  if and only if  $X_n \rightarrow x$  in probability.

\* \* \*

**Exercise 3.3.** [solution p. 133] (**Slutsky's lemma**) Let  $E$  and  $F$  be two separable metric spaces,  $y \in F$  a fixed element, and  $X, X_1, X_2, \dots$  and  $(Y_i)_{i \geq 1}$  random variables taking values in  $E$  and  $F$ , respectively, such that  $X_n \Rightarrow X$  and  $Y_n \Rightarrow y$ . Show that  $(X_n, Y_n) \Rightarrow (X, y)$ .

\* \* \*

**Exercise 3.4.** [solution p. 134] (**Countable products of metric spaces**) Let  $((E_i, d_i))_{i \geq 1}$  be a sequence of separable metric spaces. Set  $E = E_1 \times E_2 \times \dots$  and

$$\forall x = (x_i)_{i \geq 1}, y = (y_i)_{i \geq 1} \in E, \quad d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)}.$$

- (1) Show that  $(E, d)$  is separable.
- (2) Show that a sequence  $(x^n)_{n \geq 1} \in E$  converges to  $x \in E$  if and only if for every  $k \geq 1$ ,  $x_k^n \rightarrow x_k$  as  $n \rightarrow \infty$ .
- (3) Show that  $(E, d)$  is complete if all the  $E_i$  are complete.
- (4) Show that  $(E, d)$  is compact if all the  $E_i$  are compact.

\* \* \*

**Exercise 3.5.** [solution p. 134] (**An example of a nonseparable metric space**) Show that  $\mathcal{C}_b(\mathbb{R})$ , the space of bounded continuous functions from  $\mathbb{R}$  to  $\mathbb{R}$  equipped with the uniform norm, is not separable.

\* \* \*

**Exercise 3.6.** [solution p. 134] (**Converging together theorem**) On the same probability space, consider random variables  $(Y_{n,k})_{n,k \geq 1}$  and  $(X_n)_{n \geq 1}$  taking values in a separable metric space  $(E, d)$ .

Assume that for every  $k \geq 1$ , the sequence  $(Y_{n,k}, n \geq 1)$  converges in distribution to a limit  $Y_k$ . Finally, assume that for every  $\varepsilon > 0$ ,

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

- (1) Assume that  $Y_k \Rightarrow Y$  as  $k \rightarrow \infty$ . Show that  $X_n \Rightarrow Y$ .
- (2) Assume in addition that  $(E, d)$  is complete. Show that  $(Y_k)_{k \geq 1}$  converges in distribution.

*You may find the Lévy–Prokhorov metric useful.*

\* \* \*

**Exercise 3.7.** [solution p. 135] Let  $(E, d)$  be a complete metric space and  $A \subset E$ . Show that  $A$  is totally bounded if and only if  $A$  is relatively compact.

\* \* \*

**Exercise 3.8.** [solution p. 136] Let  $X, X_1, X_2, \dots$  be real-valued random variables such that  $X_n \Rightarrow X$  and  $X_n \geq 0$ . Show that  $\mathbb{E}[X] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n]$ .

\* \* \*

**Exercise 3.9.** [solution p. 136] Let  $(E, d)$  be a separable metric space.

- (1) Show that the map  $F : x \mapsto \delta_x$  is a homeomorphism from  $E$  onto  $E_0 := \{\delta_y : y \in E\}$ , where the latter is viewed as a subspace of  $\mathcal{M}_1(E)$  equipped with the topology of weak convergence.

Let  $\mu_n = \delta_{x_n}$  be a sequence in  $E_0$  converging weakly to  $\mu$ . Suppose, for a contradiction, that  $(x_n)_{n \geq 1}$  has no convergent subsequence.

- (2) Show that  $\{x_n, n \geq 1\}$  is closed, as are all its subsets (finite or infinite). Show that this contradicts weak convergence of  $\mu_n$ . Deduce that  $E_0$  is closed in  $\mathcal{M}_1(E)$ .
- (3) Deduce that if  $\mathcal{M}_1(E)$  is compact, then  $E$  is compact.

\* \* \*

**Exercise 3.10.** [solution p. 137] Let  $(E_i)_{i \geq 1}$  be a sequence of Polish spaces. Set  $E = E_1 \times E_2 \times \dots$ , equipped with the same metric  $d$  as in [Exercise 3.4](#):

$$\forall x, y \in E, \quad d(x, y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)}.$$

Show that a family  $\Gamma \subset \mathcal{M}_1(E)$  is tight if and only if for every  $i \geq 1$ ,  $\{\pi_i \mu : \mu \in \Gamma\}$  is tight, where  $\pi_i : E \rightarrow E_i$  is the canonical projection and  $\pi_i \mu$  is the pushforward of  $\mu$  under  $\pi_i$ .

**Exercise 3.11.** [solution p. 137] Let  $(E, d)$  be a metric space and  $A \subset E$ . View  $A$  as a metric space equipped with the metric  $d$ .

- (1) Show that the open subsets of  $(A, d)$  are of the form  $A \cap O$  with  $O$  open in  $(E, d)$ .
- (2) Show that the elements of  $\mathcal{B}(A)$  are of the form  $A \cap B$  with  $B \in \mathcal{B}(E)$ .
- (3) Let  $K \subset A$  be compact in  $(A, d)$ . Show that  $K$  is closed in  $E$ .

\* \* \*

**Exercise 3.12.** [solution p. 138] Show that no metric on  $\mathbb{R}^{[0,1]}$  metrizes almost everywhere convergence (you may show the existence of a sequence of functions that does not converge almost surely, but such that every subsequence has a further subsequence converging almost surely to 0).

\* \* \*

**Exercise 3.13.** [solution p. 138] Let  $(E, d)$  be a metric space.

- (1) Assume that  $E$  is separable. Show that every open subset of  $E$  is a countable union of open balls.
- (2→\*) Is the converse true?

\* \* \*

**Exercise 3.14.** [solution p. 139] Let  $(E, d)$  be a separable complete metric space. Let  $I : E \rightarrow [0, \infty]$  be a function such that for every  $c \in [0, \infty]$ ,  $\{x \in E : I(x) \leq c\}$  is compact. Let  $(\mathbb{P}_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(E)$ . Consider the following properties (A), (B) and (C):

(A) For every closed subset  $F$  of  $E$ , we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}_n(F) \leq -\inf_{x \in F} I(x).$$

(B) For every compact subset  $K$  of  $E$ , we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}_n(K) \leq -\inf_{x \in K} I(x).$$

(C) For every  $M \geq 0$ , there exists a compact subset  $K_M$  of  $E$  such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}_n(E \setminus K_M) \leq -M.$$

- (1) Prove that (C) and (B) imply (A).
- (2) Prove that (A) implies (C).

**Exercise 3.15.** [solution p. 140] Let  $d \geq 1$  be an integer. Let  $\mu, \nu \in \mathcal{M}_1(\mathbb{R}^d)$ .

- (1) Prove that the family  $\mathcal{F}$  of probability measures in  $\mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$  of the form  $P_{(X,Y)}$ , where  $X$  and  $Y$  are random variables defined on the same probability space, with  $X$  having distribution  $\mu$  and  $Y$  having distribution  $\nu$ , is compact in  $\mathcal{M}_1(\mathbb{R}^d \times \mathbb{R}^d)$  equipped with the Lévy–Prokhorov metric  $d_{LP}$  (here  $P_U$  denotes the distribution of a random variable  $U$ ).
- (2) Assume that  $\int_{\mathbb{R}^d} |x| \mu(dx) < \infty$  and  $\int_{\mathbb{R}^d} |x| \nu(dx) < \infty$ . Prove that the infimum

$$I = \inf\{\mathbb{E}[|X - Y|] : X \text{ has distribution } \mu \text{ and } Y \text{ has distribution } \nu\}$$

is attained.

\* \* \*

**Exercise 3.16.** [solution p. 141] Let  $(E, d)$  be a separable metric space. Prove that there exists a countable set  $H$  of bounded continuous functions from  $E$  to  $\mathbb{R}$  such that for every  $\mu \in \mathcal{M}_1(E)$  and every sequence  $(\mu_n)_{n \geq 1}$  of elements of  $\mathcal{M}_1(E)$ ,  $\mu_n$  converges weakly to  $\mu$  if and only if for every  $f \in H$  we have  $\mu_n(f) \rightarrow \mu(f)$ .

\* \* \*

**Exercise 3.17.** [solution p. 141] For  $x = (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}^*}$ , write  $\sup(x) = \sup_{i \geq 1} x_i$ . This supremum takes values in  $\mathbb{R} \cup \{+\infty\}$ , equipped with its usual topology.

- (1) Let  $X, X^1, X^2, \dots$  be random variables taking values in  $\mathbb{R}^{\mathbb{N}}$  (equipped with the sigma-algebra and metric defined in Section 3.3) such that  $X^n \Rightarrow X$ . Alexandra says: we then have  $\sup(X^n) \Rightarrow \sup(X)$ . Is she right? Justify your answer.
- (2) Consider  $\ell^1 = \{x \in \mathbb{R}^{\mathbb{N}} : \sum_{i=1}^{\infty} |x_i| < \infty\}$ , equipped with the metric  $d(x, y) = \sum_{i=1}^{\infty} |x_i - y_i|$  and its associated Borel sigma-algebra. Let  $X, X^1, X^2, \dots$  be random variables taking values in  $\ell^1$  such that  $X^n \Rightarrow X$ . Béatrice says: we then have  $\sup(X^n) \Rightarrow \sup(X)$ . Is she right? Justify your answer.

\* \* \*

**Exercise 3.18.** [solution p. 142] Let  $(E, d)$  be a metric space and write  $d_{LP}$  for the Lévy–Prokhorov metric on  $\mathcal{M}_1(E)$ .

- (1) Show that

$$d_{LP}(\mu, \nu) = \inf\{r > 0 : \forall F \text{ closed in } E, \mu(F) \leq \nu(F^r) + r\}$$

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

- (2) What happens if “closed” is replaced by “open”? In other words, do we have

$$d_{LP}(\mu, \nu) = \inf\{r > 0 : \forall O \text{ open in } E, \mu(O) \leq \nu(O^r) + r\}?$$

**Exercise 3.19.** [solution p. 142] Let  $\mu$  be a tight probability measure on a metric space  $(E, d)$ . Show that there exists a Borel set  $A \subset E$  such that  $(A, d)$  is separable and  $\mu(A) = 1$ .

\* \* \*

**Exercise 3.20.** [solution p. 143] Consider a sequence  $(E_n, d_n)_{n \geq 1}$  of metric spaces equipped with their Borel sigma-algebras. For every  $n \geq 1$ , let  $\mu_n$  be a probability measure on  $E_n$ , and  $(X_n(k))_{k \geq 1}$  a sequence of i.i.d. random variables taking values in  $E_n$  with distribution  $\mu_n$ . Assume that for every  $n \geq 1$ , the map  $(x, y) \mapsto d_n(x, y)$  is measurable with respect to the product sigma-algebra  $\mathcal{B}(E_n) \otimes \mathcal{B}(E_n)$  (this holds in particular if  $E_n$  is separable).

Write  $\mathcal{M}_\infty(\mathbb{R}) = \mathbb{R}^{\mathbb{N}^* \times \mathbb{N}^*} = \{(a_{i,j})_{i,j \geq 1} : a_{i,j} \in \mathbb{R} \text{ for } i, j \geq 1\}$  for the set of “infinite” real matrices, equipped with the metric

$$d((a_{i,j})_{i,j \geq 1}, (b_{i,j})_{i,j \geq 1}) = \sum_{i,j=1}^{\infty} \frac{|a_{i,j} - b_{i,j}|}{|a_{i,j} - b_{i,j}| + 1} \times \frac{1}{2^{i+j}}$$

and its associated Borel sigma-algebra.

Show that the following three assertions are equivalent:

- (i) the sequence  $((d_n(X_n(i), X_n(j)))_{i,j \geq 1})_{n \geq 1}$  is tight in  $\mathcal{M}_\infty(\mathbb{R})$ ;
- (ii) the sequence  $(d_n(X_n(1), X_n(2)))_{n \geq 1}$  is tight in  $\mathbb{R}$ ;
- (iii) for every  $n \geq 1$ , there exists  $x_n \in E_n$  such that the sequence  $(d_n(x_n, X_n(1)))_{n \geq 1}$  is tight in  $\mathbb{R}$ .

\* \* \*

**Exercise 3.21.** [solution p. 144] Let  $(E, d)$  be a metric space and  $(\mu_n)$  a sequence of probability measures on  $E$ . Assume that for every bounded continuous function  $f : E \rightarrow \mathbb{R}$ , the sequence  $(\int f d\mu_n)_{n \geq 1}$  converges.

- (1) Assume that  $E$  is compact. Show that  $(\mu_n)_{n \geq 1}$  converges weakly.
- (2) (\*) Assume that  $E$  is Polish. Show that  $(\mu_n)_{n \geq 1}$  converges weakly.

\* \* \*

**Exercise 3.22.** [solution p. 145] Let  $E$  be a separable normed vector space,  $n \geq 2$  an integer and  $(X_i)_{1 \leq i \leq n}$  i.i.d. random variables.

- (1) Assume that the random variable  $X_1$  is tight. Show that the random variable  $X_1 + \cdots + X_n$  is tight.
- (2) Is the converse true? Justify your answer.

**Exercise 3.23.** [solution p. 146] Let  $(E, d)$  be a Polish metric space. Equip the set  $\mathcal{M}_1(E)$  of probability measures on  $E$  with the Lévy–Prokhorov metric  $d_{LP}$ . Let  $(M_n)_{n \geq 1}$  be a sequence of random variables taking values in  $\mathcal{M}_1(E)$ .

- (1) Let  $M$  be a random variable taking values in  $\mathcal{M}_1(E)$ . Show that the following properties are equivalent:
  - (a)  $d_{LP}(M_n, M) \rightarrow 0$  in probability.
  - (b) for every bounded continuous function  $f : E \rightarrow \mathbb{R}$  we have  $\int f dM_n \rightarrow \int f dM$  in probability.
  - (c) for every subsequence indexed by  $\varphi$ , there exists a further subsequence indexed by  $\psi$  such that almost surely  $M_{\varphi \circ \psi(n)}$  converges to  $M$  in  $(\mathcal{M}_1(E), d_{LP})$ .

You may use the following fact (a consequence of the construction in the [finite control lemma](#), taking radii  $1/N$ ): there exists a countable set  $\mathcal{F} \subset \mathcal{C}_b(E, \mathbb{R})$  such that for every  $\mu \in \mathcal{M}_1(E)$  and every  $\varepsilon > 0$ , there exist  $f_1, \dots, f_K \in \mathcal{F}$  and  $\delta > 0$  such that

$$\left\{ \nu \in \mathcal{M}_1(E) : \left| \int f_i d\nu - \int f_i d\mu \right| < \delta \text{ for every } 1 \leq i \leq K \right\} \subset B(\mu, \varepsilon),$$

where  $B(\mu, \varepsilon)$  denotes the open ball in  $\mathcal{M}_1(E)$  centered at  $\mu$  with radius  $\varepsilon$  for  $d_{LP}$ .

- (2) Let  $\mu \in \mathcal{M}_1(E)$  be a (deterministic) probability measure. Conditionally on  $M_n$ , let  $\Pi_n^1$  and  $\Pi_n^2$  be two independent random variables with distribution  $M_n$  (that is,  $\mathbb{E}[f(\Pi_n^1)|M_n] = \int_E f dM_n$  for every nonnegative measurable function  $f$ ). Assume that  $(\Pi_n^1, \Pi_n^2)$  converges in distribution as  $n \rightarrow \infty$  to  $(\Pi^1, \Pi^2)$ , where  $\Pi^1$  and  $\Pi^2$  are two independent random variables with distribution  $\mu$ . Prove that  $d_{LP}(M_n, \mu) \rightarrow 0$  in probability.

\* \* \*

**Exercise 3.24.** [solution p. 147] Let  $(E, d)$  be a separable metric space and  $X_n, Y_n$  random variables taking values in  $E$ . Assume that  $X_n$  converges almost surely and that  $\forall \varepsilon > 0$ ,  $\mathbb{P}(d(X_n, Y_n) \geq \varepsilon) \rightarrow 0$ .

- (1) Show that  $Y_n$  converges in probability.
- (2) Is it true that  $Y_n$  converges almost surely? Justify your answer (if the answer is yes, give a proof; if it is no, give a counterexample with a space  $E$  of your choice).

\* \* \*

**Problem 3.25.** [solution p. 147]

- (1) Prove that there exists a countable set  $D$  of compactly supported Lipschitz functions from  $\mathbb{R}$  to  $\mathbb{R}$  such that for every sequence  $(\mu_n)_{n \geq 1}$  of elements of  $\mathcal{M}_1(\mathbb{R})$  and  $\mu \in \mathcal{M}_1(\mathbb{R})$ ,  $\mu_n$  converges weakly to  $\mu$  if and only if for every  $f \in D$  we have  $\mu_n(f) \rightarrow \mu(f)$ .

- (2) Prove that every compactly supported Lipschitz function  $f : \mathbb{R} \rightarrow \mathbb{R}$  can be written as the difference of two bounded nondecreasing Lipschitz functions.

*Hint.* Choose  $a < b$  such that  $\text{supp}(f) \subset ]a, b[$  and, if  $f$  is  $L$ -Lipschitz, consider the truncated ramp  $H(x) = L \min(\max(x - a, 0), b - a)$ .

- (3) Let  $Z$  be a real-valued random variable with a density. Let  $(Z_i)_{i \geq 1}$  be i.i.d. real-valued random variables with the same distribution as  $Z$ , and  $P_n$  the polynomial  $P_n(X) = \prod_{i=1}^n (X - Z_i) \in \mathbb{R}[X]$ . Write  $(Y_i^n)_{1 \leq i \leq n-1}$  for the roots of  $P'_n$ . Prove that almost surely,  $\frac{1}{n-1} \sum_{i=1}^{n-1} \delta_{Y_i^n}$  converges weakly as  $n$  tends to infinity.
- (4) Let  $\mathcal{E} = \{a_1, \dots, a_K\} \subset \mathbb{C}$  be a **finite** set. Let  $(Z_i)_{i \geq 1}$  be i.i.d. complex-valued random variables taking values in  $\mathcal{E}$ . Let  $P_n$  be the polynomial  $P_n(X) = \prod_{i=1}^n (X - Z_i) \in \mathbb{C}[X]$ . Write  $(Y_i^n)_{1 \leq i \leq n-1}$  for the complex roots of  $P'_n$ . Prove that almost surely,  $\frac{1}{n-1} \sum_{i=1}^{n-1} \delta_{Y_i^n}$  converges weakly as  $n$  tends to infinity.

\* \* \*

**Problem 3.26.** [solution p. 148] Consider  $\mathbb{U} = \cup_{n=0}^{\infty} (\mathbb{N}^*)^n$ , where  $(\mathbb{N}^*)^0 = \{\emptyset\}$  by convention. An element  $u \in \mathbb{U}$  is thus a finite sequence of strictly positive integers, also called a word. If  $u \in (\mathbb{N}^*)^n$ , write  $|u| = n$  for its length. Define  $\partial\mathbb{U} = (\mathbb{N}^*)^{\mathbb{N}^*}$  and set  $\overline{\mathbb{U}} = \mathbb{U} \cup \partial\mathbb{U}$ . For  $u \in \mathbb{U}$  and  $v \in \overline{\mathbb{U}}$ , write  $uv$  for the concatenation of  $u$  and  $v$  (in particular  $\emptyset u = u\emptyset$  for every word  $u$ ), and write  $u \preceq uv$ .

For  $u, v \in \overline{\mathbb{U}}$  with  $u \neq v$ , write  $u \wedge v \in \mathbb{U}$  for the longest word for which there exist  $x, y \in \overline{\mathbb{U}}$  such that  $u = (u \wedge v)x$  and  $v = (u \wedge v)y$ . Finally, set  $u \wedge u = u$ .

- (1) For  $u, v \in \overline{\mathbb{U}}$ , define:

$$d(u, v) = \begin{cases} 0 & \text{if } u = v, \\ e^{-|u \wedge v|} & \text{otherwise.} \end{cases}$$

Show that  $d$  is a metric on  $\overline{\mathbb{U}}$  satisfying  $d(u, w) \leq \max(d(u, v), d(v, w))$  for all  $u, v, w \in \overline{\mathbb{U}}$  and making this space Polish.

- (2) Show that open balls in  $\overline{\mathbb{U}}$  are closed.
- (3) For  $u \in \mathbb{U}$ , write  $T(u) = \{uv : v \in \overline{\mathbb{U}}\}$ . Let  $(\mu_n)_{n \geq 1}$  be a sequence of probability measures on  $\overline{\mathbb{U}}$  equipped with its Borel sigma-algebra. Show that  $(\mu_n)_{n \geq 1}$  converges weakly to a probability measure  $\mu$  on  $\overline{\mathbb{U}}$  if and only if for every  $u \in \mathbb{U}$ ,

$$\mu_n(\{u\}) \rightarrow \mu(\{u\}) \quad \text{and} \quad \mu_n(T(u)) \rightarrow \mu(T(u)).$$

**Problem 3.27.** [solution p. 149] Let  $(E, d)$  be a separable metric space. Equip  $\mathcal{M}_1(E)$  with the Lévy–Prokhorov metric  $d_{LP}$ .

You may use the following fact (a consequence of the [finite control lemma](#)): for every  $\mu \in \mathcal{M}_1(E)$  and every  $\varepsilon > 0$ , there exist  $f_1, \dots, f_K \in \mathcal{C}_b(E, \mathbb{R})$  and  $\delta > 0$  such that

$$\left\{ \nu \in \mathcal{M}_1(E) : \left| \int f_i d\nu - \int f_i d\mu \right| < \delta \text{ for every } 1 \leq i \leq K \right\} \subset B(\mu, \varepsilon),$$

where  $B(\mu, \varepsilon)$  denotes the open ball in  $\mathcal{M}_1(E)$  centered at  $\mu$  with radius  $\varepsilon$  for  $d_{LP}$ .

Let  $\mu \in \mathcal{M}_1(E)$  be a fixed (deterministic) probability measure, and  $(M_n)_{n \geq 1}$  a sequence of random variables taking values in  $\mathcal{M}_1(E)$ .

### Part One.

(1) Prove that the following assertions are equivalent:

- (a) for every  $f \in \mathcal{C}_b(E, \mathbb{R})$ ,  $\int f dM_n \rightarrow \int f d\mu$  in  $L^1$ .
- (b)  $M_n$  converges in distribution to  $\mu$  (viewed as a random variable taking values in  $\mathcal{M}_1(E)$ );
- (c)  $M_n$  converges in probability to  $\mu$  (viewed as a random variable taking values in  $\mathcal{M}_1(E)$ );
- (d) for every  $f \in \mathcal{C}_b(E, \mathbb{R})$ ,  $\int f dM_n$  converges in probability to  $\int f d\mu$ ;
- (e) when  $E = \mathbb{R}$ , writing  $\Phi_n(t) = \int e^{itx} M_n(dx)$  and  $\Phi(t) = \int e^{itx} \mu(dx)$ , we have  $\Phi_n(t) \rightarrow \Phi(t)$  in probability for every  $t \in \mathbb{R}$ .

(2) In this question, assume that  $E = \mathbb{R}$  and that  $\mu$  is determined by its moments.

- (a) Assume that for every  $k \geq 1$ ,  $\int x^k dM_n$  converges in probability to  $\int x^k d\mu$ . Show that  $M_n$  converges in distribution to  $\mu$ .
- (b) Assume that for every  $k \geq 1$ ,  $\mathbb{E}[\int x^k dM_n]$  converges to  $\int x^k d\mu$  and  $\text{Var}(\int x^k dM_n) \rightarrow 0$ . Show that  $M_n$  converges in distribution to  $\mu$ .

**Part Two.** In what follows, assume that the random variables  $(M_n)_{n \geq 1}$  are defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

(3) Prove that the following four assertions are equivalent:

- (a) for almost every  $\omega \in \Omega$ ,  $M_n$  converges weakly to  $\mu$ ;
- (b) for almost every  $\omega \in \Omega$ , for every  $f \in \mathcal{C}_b(E, \mathbb{R})$ ,  $\int f dM_n \rightarrow \int f d\mu$ ;
- (c) for every  $f \in \mathcal{C}_b(E, \mathbb{R})$ , for almost every  $\omega \in \Omega$ ,  $\int f dM_n \rightarrow \int f d\mu$ ;
- (d) when  $E = \mathbb{R}$ , for every  $t \in \mathbb{R}$ , almost surely  $\int e^{itx} dM_n(dx) \rightarrow \int e^{itx} \mu(dx)$ .

(4) In this question, assume that  $E = \mathbb{R}$  and that  $\mu$  is determined by its moments. Assume that for every  $k \geq 1$ ,  $\int x^k dM_n$  converges for almost every  $\omega \in \Omega$  to  $\int x^k d\mu$ . Show that for almost every  $\omega \in \Omega$ ,  $M_n$  converges weakly to  $\mu$ .

- (5) Let  $(X_n)_{n \geq 1}$  be a sequence of i.i.d. random variables with distribution  $\mu$ . What can be said about the behavior, as  $n \rightarrow \infty$ , of  $M_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$ ?

\* \* \*

**Problem 3.28.** [solution p. 152] Let  $(E, d)$  be a **compact** metric space equipped with its Borel sigma-algebra. Write  $\mathcal{M}_1(E)$  for the set of probability measures on  $E$ . For  $\mu \in \mathcal{M}_1(E)$  and  $\varepsilon > 0$ , set

$$I_\mu(\varepsilon) = \inf_{x \in E} \mu(\overline{B}(x, \varepsilon)),$$

where  $\overline{B}(x, \varepsilon) = \{y \in E : d(x, y) \leq \varepsilon\}$  denotes the closed ball of radius  $\varepsilon$  centered at  $x$ .

**Part One.**

- (1) Compute  $\lim_{M \rightarrow \infty} I_\mu(M)$ .
- (2) Show that  $I_\mu$  is càdlàg.
- (3) Show that if for every  $x \in E$  we have  $\mu(\{x\}) = 0$ , then  $\lim_{\varepsilon \downarrow 0} I_\mu(\varepsilon) = 0$ .
- (4) Camille says: “The converse of the statement in question (3) is also true.” Is he right? Justify your answer.

**Part Two.** Let  $(\mu_n)_{n \geq 1}$  be a sequence in  $\mathcal{M}_1(E)$  and  $\mu \in \mathcal{M}_1(E)$ .

- (5) Assume that  $\mu_n$  converges weakly to  $\mu$ . Show that for every  $\varepsilon > 0$  we have

$$\limsup_{n \rightarrow \infty} I_{\mu_n}(\varepsilon) \leq I_\mu(\varepsilon).$$

A probability measure on  $E$  is said to have *full support* if it assigns strictly positive mass to every nonempty open set.

- (6) Assume that  $\mu_n$  converges weakly to  $\mu$ . Show that the following three assertions are equivalent:
- (a)  $\mu$  has full support;
  - (b)  $I_\mu(\varepsilon) > 0$  for every  $\varepsilon > 0$ ;
  - (c)  $\liminf_{n \rightarrow \infty} I_{\mu_n}(\varepsilon) > 0$  for every  $\varepsilon > 0$ .

**Part Three.** In what follows, assume that  $(M_n)_{n \geq 1}$  is a sequence of random variables taking values in  $(\mathcal{M}_1(E), d_{LP})$  and that  $M$  is a random variable taking values in  $(\mathcal{M}_1(E), d_{LP})$ .

- (7) Assume that  $M_n$  converges in distribution to  $M$ . Prove that the following two assertions are equivalent:
- (a)  $M$  almost surely has full support;

(b) for all  $\varepsilon, \varepsilon' > 0$ , there exist  $\delta, N > 0$  such that

$$\forall n \geq N, \quad \mathbb{P}(I_{M_n}(\varepsilon) < \delta) < \varepsilon'.$$

- (8) Dominique says: “ $M_n$  converges in distribution to  $M$  if and only if, on every compact interval of  $\mathbb{R}_+^*$ ,  $I_{M_n}$  converges in distribution to  $I_M$  for the Skorokhod topology.” Is she right? Justify your answer.

### 3.2 Solutions

**Solution to Exercise 3.1.** [statement p. 124] First, the sequence  $(x_n)$  is bounded. Indeed, otherwise there exists a subsequence indexed by  $\phi$  such that  $|x_{\phi(n)}| \rightarrow \infty$ . By the Portmanteau theorem, for every  $A > 0$  we then have  $0 = \liminf_{n \rightarrow \infty} \delta_{x_{\phi(n)}}([ -A, A ]) \geq \mu([ -A, A ])$ , hence  $\mu([ -A, A ]) = 0$  for every  $A > 0$ , a contradiction.

Let us verify that  $(x_n)$  has only one subsequential limit. If  $x_{\phi(n)} \rightarrow x$ , then by the Portmanteau theorem, for every  $\varepsilon > 0$ ,  $1 = \limsup_{n \rightarrow \infty} \delta_{x_{\phi(n)}}([x - \varepsilon, x + \varepsilon]) \leq \mu([x - \varepsilon, x + \varepsilon])$ . Since this holds for every  $\varepsilon > 0$ , we deduce that  $\mu(\{x\}) = 1$ . Since there cannot be  $x \neq y$  such that  $\mu(\{x\}) = \mu(\{y\}) = 1$ , this shows that  $\mu = \delta_x$ , where  $x$  is the unique subsequential limit of  $(x_n)_{n \geq 1}$ .

\* \* \*

**Solution to Exercise 3.2.** [statement p. 124]

$\Leftarrow$  Suppose that  $X_n \rightarrow x$  in probability. Let  $f \in \mathcal{C}_b(E)$  and  $\varepsilon > 0$ . By continuity of  $f$  at  $x$ , there exists  $\delta > 0$  such that  $d(y, x) < \delta$  implies  $|f(y) - f(x)| < \varepsilon$ . Then

$$|\mathbb{E}[f(X_n)] - f(x)| \leq \varepsilon + 2\|f\|_\infty \mathbb{P}(d(X_n, x) \geq \delta).$$

Letting  $n \rightarrow \infty$  and then  $\varepsilon \downarrow 0$ , we obtain  $\mathbb{E}[f(X_n)] \rightarrow f(x)$ , hence  $X_n \Rightarrow x$ . Since distance to a fixed point is continuous and hence measurable, this argument is valid on every metric space.

$\Rightarrow$  For  $\varepsilon > 0$ , writing  $B(x, \varepsilon)$  for the open ball centered at  $x$  with radius  $\varepsilon$ , so that its complement is closed, the Portmanteau theorem gives:

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \notin B(x, \varepsilon)) \leq \mathbb{P}(x \notin B(x, \varepsilon)) = 0.$$

Thus for every  $\varepsilon > 0$ ,  $\mathbb{P}(d(X_n, x) \geq \varepsilon) \rightarrow 0$ , hence  $X_n \rightarrow x$  in probability.

\* \* \*

**Solution to Exercise 3.3.** [statement p. 124] By assumption,  $(X_n, y) \Rightarrow (X, y)$ . Equip the product with the metric

$$d((x_1, x_2), (y_1, y_2)) = d_E(x_1, y_1) + d_F(x_2, y_2).$$

By assumption,

$$d((X_n, y), (X_n, Y_n)) = d_F(Y_n, y) \rightarrow 0 \quad \text{in probability.}$$

It follows that  $(X_n, Y_n) \Rightarrow (X, y)$  by a result from the lectures.

**Solution to Exercise 3.4.** [statement p. 124] First, one readily verifies that  $d$  is a metric and that for all  $N \geq 1, x, y \in E$ ,

$$\sum_{n=N}^{\infty} \frac{1}{2^n} \frac{d_n(x_n, y_n)}{1 + d_n(x_n, y_n)} \leq \frac{1}{2^{N-1}}.$$

- (1) Let  $D_i$  be a countable dense subset of  $E_i$  and  $a_i \in E_i$  a fixed element of  $E_i$ . Write  $D$  for the set of sequences of the form  $(x_1, \dots, x_K, a_{K+1}, \dots)$ , where  $K \geq 1$  is an integer,  $x_1 \in D_1, \dots, x_K \in D_K$ . Let  $\varepsilon > 0$  and  $y = (y_1, y_2, \dots) \in E$ . Choose  $N$  such that  $1/2^{N-1} < \varepsilon$  and an element  $x = (x_1, \dots, x_N, a_{N+1}, \dots) \in D$  such that  $d_i(x_i, y_i) \leq \varepsilon/N$  for every  $1 \leq i \leq N$ . We then have  $d(x, y) \leq 2\varepsilon$ .

- (2)  $\Rightarrow$  For every  $i \geq 1$ , we have

$$\frac{1}{2^i} \frac{d_i(x_i, y_i)}{1 + d_i(x_i, y_i)} \leq d(x, y).$$

Thus  $d(x^n, x) \rightarrow 0$  implies  $d_i(x_i^n, x_i) \rightarrow 0$  for every  $i \geq 1$ .

$\Leftarrow$  Suppose that  $x_k^n \rightarrow x_k$  for every  $k \geq 1$  as  $n \rightarrow \infty$ . Choose  $N$  such that  $1/2^{N-1} < \varepsilon$  and an integer  $n_0$  such that  $n \geq n_0$  implies  $d_i(x_i^n, x_i) \leq \varepsilon/N$  for every  $1 \leq i \leq N$ . Then  $d(x^n, x) \leq 2\varepsilon$  for  $n \geq n_0$ .

- (3) Let  $(x^n)_{n \geq 1}$  be a Cauchy sequence in  $(E, d)$ . Then for every  $i \geq 1$ ,  $(x_i^n)_{n \geq 1}$  is a Cauchy sequence in  $(E_i, d_i)$  and hence converges to an element denoted by  $x_i$ . We deduce that  $x^n \rightarrow (x_1, x_2, \dots)$  by the previous question.
- (4) Let  $(x^n)_{n \geq 1}$  be a sequence in  $(E, d)$ . For  $i \geq 1$ , each sequence  $(x_i^n)_{n \geq 1}$  takes values in  $E_i$  and therefore has a subsequential limit. By a diagonal argument, there exists a subsequence indexed by  $\phi$  such that  $(x_i^{\phi(n)})_{n \geq 1}$  converges for every  $i \geq 1$  to a value denoted by  $x_i$ . By question (2), we deduce that  $x^{\phi(n)} \rightarrow (x_1, x_2, \dots)$ .

\* \* \*

**Solution to Exercise 3.5.** [statement p. 124] Consider the subset  $K$  of  $\mathcal{C}_b(\mathbb{R})$  consisting of functions that are affine on every interval  $[m, m+1]$ ,  $m \in \mathbb{Z}$ , and take values 0 or 1 on  $\mathbb{Z}$ . Each sequence in  $\{0, 1\}^{\mathbb{Z}}$  defines exactly one such function, so  $K$  is uncountable. If  $f, g \in K$  are distinct, their values differ at an integer and  $\|f - g\|_{\infty} = 1$ . The open balls  $(B(f, 1/2))_{f \in K}$  are therefore pairwise disjoint. Every countable subset of  $\mathcal{C}_b(\mathbb{R})$  misses one of these balls: this space is not separable.

\* \* \*

**Solution to Exercise 3.6.** [statement p. 124]

- (1) We apply the Portmanteau theorem. Let  $F$  be closed and  $\varepsilon > 0$ . Write  $F_{\varepsilon} = \{x \in E : d(x, F) \leq \varepsilon\}$ , which is closed. Then

$$\mathbb{P}(X_n \in F) \leq \mathbb{P}(Y_{n,k} \in F_{\varepsilon}) + \mathbb{P}(d(X_n, Y_{n,k}) \geq \varepsilon).$$

Since  $Y_{n,k} \Rightarrow Y_k$ , the Portmanteau theorem gives

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(Y_k \in F_\varepsilon) + \limsup_{n \rightarrow \infty} \mathbb{P}(d(X_n, Y_{n,k}) \geq \varepsilon).$$

Now  $Y_k \Rightarrow Y$ , so taking the lim sup as  $k \rightarrow \infty$ , the Portmanteau theorem gives:

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(Y \in F_\varepsilon).$$

Letting  $\varepsilon \downarrow 0$ , we conclude that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) \leq \mathbb{P}(Y \in F),$$

and hence  $X_n \Rightarrow Y$  by the Portmanteau theorem.

- (2) Since  $(E, d)$  is separable and complete,  $(\mathcal{M}_1(E), d_{LP})$  is separable and complete. It therefore suffices to show that  $(Y_k)_{k \geq 1}$  is Cauchy in  $(\mathcal{M}_1(E), d_{LP})$ . To this end, write

$$d_{LP}(Y_p, Y_q) \leq d_{LP}(Y_p, Y_{n,p}) + d_{LP}(Y_{n,p}, Y_{n,q}) + d_{LP}(Y_{n,q}, Y_q),$$

so that

$$d_{LP}(Y_p, Y_q) \leq \limsup_{n \rightarrow \infty} d_{LP}(Y_{n,p}, Y_{n,q}).$$

Next, write

$$d_{LP}(Y_{n,p}, Y_{n,q}) \leq d_{LP}(Y_{n,p}, X_n) + d_{LP}(Y_{n,q}, X_n).$$

For every Borel set  $A$ ,

$$\mathbb{P}(X_n \in A) \leq \mathbb{P}(Y_{n,k} \in A^\varepsilon) + \mathbb{P}(d(X_n, Y_{n,k}) \geq \varepsilon).$$

Choosing  $K$  such that

$$k \geq K \implies \limsup_{n \rightarrow \infty} \mathbb{P}(d(X_n, Y_{n,k}) \geq \varepsilon) \leq \varepsilon/2,$$

we deduce that for  $k \geq K$ ,  $\limsup_{n \rightarrow \infty} d_{LP}(X_n, Y_{n,k}) \leq \varepsilon$ . Thus, for all  $p, q \geq K$ ,

$$\limsup_{n \rightarrow \infty} d_{LP}(Y_{n,p}, Y_{n,q}) \leq 2\varepsilon,$$

which completes the proof.

\* \* \*

### Solution to Exercise 3.7. [statement p. 125]

$\Leftarrow$  If  $\bar{A}$  is compact, for every  $\varepsilon > 0$  it can be covered by finitely many balls of radius  $\varepsilon$ , and this is therefore also true of  $A$ .

$\Rightarrow$  First, observe that  $\bar{A}$  is also totally bounded (if  $A$  is contained in a union of balls of radius  $\varepsilon$ , then  $\bar{A}$  is contained in a union of balls of radius  $2\varepsilon$ ). By the sequential characterization of compactness in a metric space, it suffices to show that

every sequence  $(x_n)_{n \geq 1}$  in  $\bar{A}$  has a subsequential limit. By total boundedness, we can inductively construct subsequences indexed by  $\psi_1, \psi_2, \dots$  and elements  $z_1, z_2, \dots$  such that for every  $n \geq 1$ ,

$$x_{\psi_1(n)} \in B(z_1, \varepsilon), \dots, x_{\psi_1 \circ \dots \circ \psi_k(n)} \in B(z_k, \varepsilon/2^k), \dots$$

and  $B(z_{k+1}, \varepsilon/2^{k+1}) \subset B(z_k, \varepsilon/2^k)$ . By a diagonal argument, set  $\phi(n) = \psi_1 \circ \dots \circ \psi_n(n)$  and verify that the sequence  $(x_{\phi(n)})_{n \geq 1}$  is Cauchy, hence convergent since  $E$  is complete.

\* \* \*

**Solution to Exercise 3.8.** [statement p. 125] By Skorokhod's representation theorem, after changing the probability space if necessary, we may assume that the convergence  $X_n \rightarrow X$  holds almost surely. In particular,  $X \geq 0$  almost surely, and the desired result follows from Fatou's lemma.

\* \* \*

**Solution to Exercise 3.9.** [statement p. 125]

- (1) Clearly, if  $x_n \rightarrow x$ , then  $f(x_n) \rightarrow f(x)$  for every bounded continuous function  $f : E \rightarrow \mathbb{R}$ , so  $F$  is continuous. Moreover,  $F$  is clearly injective: if  $\delta_x = \delta_y$  with  $x \neq y$ , then  $1 = \delta_x(\{x\}) = \delta_y(\{x\}) = 0$ . Finally, if  $\delta_{x_n} \rightarrow \delta_x$ , considering the functions  $f_K(u) = \max(1 - Kd(x, u), 0)$  and observing that  $\delta_z(f_K) = 0$  if  $d(z, x) > 1/K$ , we see that  $x_n \rightarrow x$ . This completes the proof.
- (2) Every subset of  $\{x_n : n \geq 1\}$  is closed: a point in its closure that did not belong to it would yield a convergent subsequence of  $(x_n)$ . Moreover, each value is taken only finitely many times, since otherwise there would be a constant subsequence. We can therefore extract a sequence  $(x_{\phi(n)})$  of pairwise distinct values. The closed sets

$$F_0 = \{x_{\phi(2n)} : n \geq 1\}, \quad F_1 = \{x_{\phi(2n-1)} : n \geq 1\}$$

are disjoint. By the Portmanteau theorem, for  $j = 0, 1$ ,  $1 = \limsup_n \delta_{x_{\phi(n)}}(F_j) \leq \mu(F_j)$ , which is impossible.

To show that  $E_0$  is closed in  $\mathcal{M}_1(E)$ , consider a sequence  $\mu_n = \delta_{x_n}$  in  $E_0$  converging weakly to  $\mu$ . By the preceding argument, there exists a subsequence such that  $(x_{\phi(n)})_{n \geq 1}$  converges to a limit denoted by  $x$ . Then  $\delta_{x_{\phi(n)}} \Rightarrow \delta_x$ , and hence  $\mu = \delta_x$ . Thus,  $E_0$  is closed.

- (3) Let  $(x_n)_{n \geq 1}$  be a sequence in  $E$ . Then  $(\delta_{x_n})_{n \geq 1}$  is a sequence in  $E_0$ , which is compact (being closed in the compact space  $\mathcal{M}_1(E)$ ). There therefore exists a subsequence indexed by  $\phi$  such that  $(\delta_{x_{\phi(n)}})$  converges weakly, and by the preceding argument there exists  $x \in E$  such that  $x_{\phi(n)} \rightarrow x$ , which proves the result.

**Solution to Exercise 3.10.** [statement p. 125]

$\Leftarrow$  Suppose that for every  $i \geq 1$ ,  $\{\pi_i \mu : \mu \in \Gamma\}$  is tight. Let  $\varepsilon > 0$ , and for every  $i \geq 1$  choose a compact set  $K_i \subset E_i$  such that for every  $\mu \in \Gamma$  we have  $\pi_i \mu({}^c K_i) \leq \varepsilon/2^i$ . Set

$$K = \prod_{i=1}^{\infty} K_i = \bigcap_{i=1}^{\infty} \pi_i^{-1}(K_i).$$

A diagonal argument shows that  $K$  is compact in  $E$ . Moreover, for every  $\mu \in \Gamma$

$$\mu({}^c K) \leq \sum_{i=1}^{\infty} (\mu(\pi_i^{-1}({}^c K_i))) = \sum_{i=1}^{\infty} (1 - \pi_i \mu(K_i)) \leq \varepsilon.$$

$\Rightarrow$  Let  $\varepsilon > 0$  and let  $K$  be a compact subset of  $E$  such that  $\mu(K) \geq 1 - \varepsilon$  for every  $\mu \in \Gamma$ . By continuity of the canonical projections, the sets  $K_i = \pi_i(K)$  are compact. Moreover, for  $i \geq 1$  and  $\mu \in \Gamma$ , since  $K \subset \pi_i^{-1}(K_i)$ , we have

$$\pi_i \mu(K_i) \geq \mu(K) \geq 1 - \varepsilon.$$

\* \* \*

**Solution to Exercise 3.11.** [statement p. 126]

- (1) Since the inclusion  $I : (A, d) \rightarrow (E, d)$  is continuous, the inverse image of every open set is open, so sets of the form  $A \cap O$  with  $O$  open in  $(E, d)$  are open in  $(A, d)$ .

Conversely, let  $U$  be open in  $(A, d)$ . For every  $x \in U$ , there then exists  $r_x > 0$  such that  $B_A(x, r_x) \subset U$ . Then

$$U = \bigcup_{x \in U} B_A(x, r_x).$$

Now  $B_A(x, r_x) = B_E(x, r_x) \cap A$ . Hence

$$U = \bigcup_{x \in U} (B_E(x, r_x) \cap A) = \left( \bigcup_{x \in U} B_E(x, r_x) \right) \cap A,$$

the result follows by taking  $O = \bigcup_{x \in U} B_E(x, r_x)$ , which is open in  $E$ .

- (2) Since the inclusion  $I : (A, d) \rightarrow (E, d)$  is continuous and hence measurable, sets of the form  $A \cap B$  with  $B \in \mathcal{B}(E)$  belong to  $\mathcal{B}(A)$ .

Conversely, sets of the form  $A \cap B$  with  $B \in \mathcal{B}(E)$  include the open subsets of  $A$  (by the previous question) and form a sigma-algebra, so they include all elements of  $\mathcal{B}(A)$ .

- (3) Let  $(x_n)_{n \geq 1}$  be a sequence of elements of  $K$  converging to  $x \in E$ . By compactness of  $K$ , there exist a subsequence indexed by  $\phi$  and  $y \in K$  such that  $x_{\phi(n)} \rightarrow y$ . Thus  $y = x \in K$ , so  $K$  is closed.

**Solution to Exercise 3.12.** [statement p. 126] For  $n \geq 1$ , write  $n = 2^N + k$  with  $0 \leq k < 2^N$ , and set

$$f_n(x) = \mathbb{1}_{[k2^{-N}, (k+1)2^{-N}]}(x).$$

At each level  $N$ , every point of  $[0, 1]$  belongs to at least one interval in the sweep and, for  $N \geq 2$ , does not belong to all of them. The sequence  $(f_n(x))$  therefore keeps taking both values 0 and 1: it does not converge at any point. However,  $\text{Leb}(\{f_n = 1\}) = 2^{-N} \rightarrow 0$ . From every subsequence, we can therefore extract  $(f_{n_j})$  such that  $\sum_j \text{Leb}(\{f_{n_j} = 1\}) < \infty$ . By Borel–Cantelli,  $f_{n_j} \rightarrow 0$  almost everywhere. The subsequence criterion for convergence in a metric space therefore rules out any metric that metrizes almost everywhere convergence. This argument remains valid on the quotient of measurable functions modulo almost everywhere equality.

\* \* \*

**Solution to Exercise 3.13.** [statement p. 126]

(1) Let  $(x_i)_{i \geq 1}$  be a countable dense sequence in  $E$ . If  $U$  is open, write

$$U = \bigcup_{i, m \geq 1: B(x_i, 1/m) \subset U} B(x_i, 1/m).$$

Indeed, the inclusion  $\supset$  is clear. For the inclusion  $\subset$ , let  $x \in U$  and choose  $\varepsilon > 0$  such that  $B(x, \varepsilon) \subset U$ . Choose an integer  $m > 2/\varepsilon$ , then  $x_i$  such that  $d(x, x_i) < 1/(2m)$ . Then  $x \in B(x_i, 1/m)$  and, for  $z \in B(x_i, 1/m)$ ,  $d(z, x) < 3/(2m) < \varepsilon$ , so  $B(x_i, 1/m) \subset U$ .

(2) Yes! Suppose that every open subset of  $E$  is a countable union of open balls. Argue by contradiction, assuming that  $(E, d)$  is nonseparable.

First verify that there exist  $r > 0$  and an uncountable subset  $E_1 \subset E$  such that for all  $x \neq y$  in  $E_1$  we have  $d(x, y) > r$ . For each  $r = 1/n$ , consider a subset of  $E$  maximal under inclusion with this property (which exists by Zorn's lemma). If all these subsets are countable, then clearly  $E$  is separable.

Now consider the set  $E_2 \subset E_1$  of points  $x \in E_1$  for which there exists a point  $y_x \in E$  such that  $0 < d(x, y_x) < r/10$ .

Case 1.  $E_2$  is uncountable. Set

$$U = \bigcup_{x \in E_2} B(x, d(x, y_x)),$$

which is a disjoint union. Consider an open ball  $B(z, a) \subset U$ . Since  $z \in U$ , there exists  $x \in E$  such that  $d(z, x) < d(x, y_x)$ . But

$$a \leq d(z, y_x) \leq d(z, x) + d(x, y_x) < 2d(x, y_x) < r/5$$

because  $y_x \notin U$  and hence  $y_x \notin B(z, a)$ . Consequently,  $B(z, a)$  is contained in a unique ball  $B(x, d(x, y_x))$ , so uncountably many open balls are needed to cover  $U$ , a contradiction.

Case 2.  $E_3 = E_1 \setminus E_2$  is uncountable. Every point of  $E_3$  is isolated in  $E$ , so every subset of  $E_3$  is open in  $E$ . For  $x \in E_3$ , set

$$R(x) = \sup\{a > 0 : B(x, a) \cap E_3 \text{ is countable}\}.$$

Then  $0 < R(x) < \infty$  and  $B(x, R(x)) \cap E_3$  is countable. For every countable subset  $S \subset E_3$  and every  $x \in E_3 \setminus S$ , one can choose distinct points  $z_j \in E_3 \setminus (S \cup \{x\})$  such that  $R(x) \leq d(x, z_j) < R(x) + 1/j$ : every ball of radius strictly greater than  $R(x)$  is uncountable, while the ball of radius  $R(x)$  is countable. Call a set  $\{x\} \cup \{z_j : j \geq 1\}$  constructed in this way a star centered at  $x$ . By Zorn's lemma, there exists a maximal family of pairwise disjoint stars. It is uncountable: otherwise their union would be countable and the preceding construction would allow us to add a disjoint star. Let  $U$  be the set of their centers. If  $U = \bigcup_{i \geq 1} B(x_i, r_i)$ , the centers  $x_i$  belong to  $U$ . For at least one  $i$ ,  $r_i > R(x_i)$ , since otherwise each ball, being contained in  $U \subset E_3$ , would be countable. But this ball then contains points  $z_j$  of the star centered at  $x_i$ , whereas that star meets  $U$  only at  $x_i$ . A contradiction.

\* \* \*

**Solution to Exercise 3.14.** [statement p. 126]

- (1) Assume (B) and (C). Let  $F$  be closed and  $a = \inf_{x \in F} I(x)$ . If  $a = 0$ , the required bound follows from  $\mathbb{P}_n(F) \leq 1$ . Otherwise, fix  $0 \leq c < a$  and a compact set  $K_c$  supplied by (C). Since  $F \cap K_c$  is compact, (B) implies

$$\limsup_n \frac{1}{n} \ln \mathbb{P}_n(F \cap K_c) \leq -a \leq -c.$$

Moreover,  $\limsup_n n^{-1} \ln \mathbb{P}_n(E \setminus K_c) \leq -c$ . The bound

$$\mathbb{P}_n(F) \leq \mathbb{P}_n(F \cap K_c) + \mathbb{P}_n(E \setminus K_c)$$

therefore gives  $\limsup_n n^{-1} \ln \mathbb{P}_n(F) \leq -c$ . Letting  $c$  tend to  $a$  if  $a < \infty$ , or to  $+\infty$  if  $a = +\infty$ , completes the proof.

- (2) Let  $(x_i)_{i \geq 1}$  be a dense sequence in  $E$  and fix  $M > 0$ . For  $k, \ell \geq 1$ , set

$$O_{k,\ell} = \bigcup_{i=1}^{\ell} B(x_i, 1/k).$$

Fix  $M' > M$ . By compactness of  $\{I \leq kM'\}$ , there exists  $j_k$  such that  $\{I \leq kM'\} \subset O_{k,j_k}$ . By (A),

$$\limsup_n \frac{1}{n} \ln \mathbb{P}_n(E \setminus O_{k,j_k}) \leq -kM'.$$

There therefore exists  $n_k$  such that  $\mathbb{P}_n(E \setminus O_{k,j_k}) \leq e^{-nkM'}$  for  $n \geq n_k$ . Since  $O_{k,\ell} \uparrow E$  as  $\ell \rightarrow \infty$ , we may increase  $j_k$  to  $\ell_k$  to obtain

$$\mathbb{P}_n(E \setminus O_{k,\ell_k}) \leq e^{-nkM} \quad \text{for every } n \geq 1.$$

Set  $K_M = \bigcap_{k \geq 1} \overline{O_{k, \ell_k}}$ . This set is closed and totally bounded, since for each  $k$  it is contained in a finite union of closed balls of radius  $1/k$ . It is therefore compact in the complete space  $E$ . Finally,

$$\mathbb{P}_n(E \setminus K_M) \leq \sum_{k \geq 1} e^{-nkM} = \frac{e^{-nM}}{1 - e^{-nM}},$$

hence  $\limsup_n n^{-1} \ln \mathbb{P}_n(E \setminus K_M) \leq -M$ . For the level  $M = 0$ , the compact set obtained at level 1 works.

**Remark.** Property (C) is called exponential tightness. This exercise is inspired by Lemmas 2.5 and 2.6 in James Lynch and Jayaram Sethuraman, "Large Deviations for Processes with Independent Increments." Ann. Probab. 15 (2) 610 - 627, April, 1987. <https://doi.org/10.1214/aop/1176992161> and is now a basic building block of large deviation theory.

\* \* \*

### Solution to Exercise 3.15. [statement p. 127]

- (1) Let  $(\mathbb{P}_{(X_n, Y_n)})_{n \geq 1}$  be a sequence of elements of  $\mathcal{F}$ . We show that this sequence has a subsequence converging weakly to an element of  $\mathcal{F}$ . First show that this sequence is tight.

Let  $\varepsilon > 0$ . Let  $K_1$  and  $K_2$  be two compact subsets of  $\mathbb{R}^d$  such that  $\mu(K_1) \geq 1 - \varepsilon$  and  $\nu(K_2) \geq 1 - \varepsilon$ . Then for every  $n \geq 1$ :

$$\mathbb{P}((X_n, Y_n) \in K_1 \times K_2) \geq 1 - 2\varepsilon.$$

Thus, by Prokhorov's theorem, there exist a subsequence indexed by  $\varphi$  and a random variable  $(U, V)$  such that  $(X_{\varphi(n)}, Y_{\varphi(n)})$  converges in distribution to  $(U, V)$ . It suffices to verify that  $U$  and  $X$  have the same distribution, as do  $V$  and  $Y$ . To this end, continuity of  $(x, y) \mapsto x$  and  $(x, y) \mapsto y$  implies that  $X_{\varphi(n)}$  converges in distribution to  $U$  and  $Y_{\varphi(n)}$  converges in distribution to  $V$ . Since  $X_{\varphi(n)}$  has the same distribution as  $X$  and  $Y_{\varphi(n)}$  has the same distribution as  $Y$ , the result follows.

- (2) Let  $P_{(X_n, Y_n)}$  be a sequence of elements of  $\mathcal{F}$  such that

$$\mathbb{E}[|X_n - Y_n|] \rightarrow I.$$

By question (1), there exist a subsequence indexed by  $\varphi$  and an element  $P_{(X, Y)}$  of  $\mathcal{F}$  such that  $(X_{\varphi(n)}, Y_{\varphi(n)})$  converges in distribution to  $(X, Y)$ . Reindexing this subsequence if necessary, assume that  $(X_n, Y_n)$  converges in distribution to  $(X, Y)$ . Let us show that

$$\mathbb{E}[|X - Y|] = I.$$

For every  $R \geq 0$ , since the function  $(x, y) \mapsto |x - y| \wedge R$  is bounded and continuous,

$$\mathbb{E}[|X_n - Y_n| \wedge R] \xrightarrow{n \rightarrow \infty} \mathbb{E}[|X - Y| \wedge R].$$

Since  $\mathbb{E} [|X_n - Y_n| \wedge R] \leq \mathbb{E} [|X_n - Y_n|]$ , we deduce that

$$\mathbb{E} [|X - Y| \wedge R] \leq I.$$

Letting  $R \rightarrow \infty$ , we obtain  $\mathbb{E} [|X - Y|] \leq I$ . Clearly  $I \leq \mathbb{E} [|X - Y|]$  by definition of  $I$ , which proves the result.

**Remark.** The family  $\mathcal{F}$  is called the set of couplings of  $X$  and  $Y$ , and the coupling in question (2) is called an optimal coupling (or transport) for the  $L^1$  cost. This exercise is inspired by Section 5.4.3 of [this document](#).

\* \* \*

**Solution to Exercise 3.16.** [statement p. 127] Let  $(x_i)_{i \geq 1}$  be a dense sequence in  $E$ . For  $K, N \geq 1$  and  $\emptyset \neq I \subset \{1, \dots, K\}$ , set

$$f_I^{K,N}(x) = \max \left( 1 - Nd \left( x, \bigcup_{i \in I} B(x_i, 1/N) \right), 0 \right),$$

and set  $f_{\emptyset}^{K,N} = 0$ . The family  $H$  of these functions is countable and contained in  $\mathcal{C}_b(E)$ . Weak convergence implies convergence of the integrals of its elements. Conversely, suppose that  $\mu_n(f) \rightarrow \mu(f)$  for every  $f \in H$ . Fix  $\varepsilon > 0$ , choose an integer  $N > 3/\varepsilon$ , then  $K$  such that

$$\mu \left( \bigcup_{i=1}^K B(x_i, 1/N) \right) > 1 - 1/N.$$

The functions  $f_I^{K,N}$  are exactly those constructed in the [finite control lemma](#), with  $\eta = 1/N$ . For sufficiently large  $n$ , the  $2^K$  differences  $|\mu_n(f_I^{K,N}) - \mu(f_I^{K,N})|$  are all less than  $1/N$ . The lemma therefore gives

$$d_{LP}(\mu_n, \mu) \leq 3/N < \varepsilon.$$

Thus  $d_{LP}(\mu_n, \mu) \rightarrow 0$ , hence  $\mu_n \xrightarrow{\text{weak}} \mu$ .

\* \* \*

**Solution to Exercise 3.17.** [statement p. 127]

(1) No. Take deterministic  $X^n = e_n$ , where  $(e_n)_k = \mathbb{1}_{\{k=n\}}$ . Every coordinate of  $e_n$  tends to 0, so  $e_n \rightarrow 0$  in  $\mathbb{R}^N$ , and hence  $X^n \Rightarrow 0$ . Yet  $\sup(X^n) = 1$  for every  $n$ , whereas  $\sup(0) = 0$ .

(2) Yes! One verifies that the function

$$\begin{aligned} \sup &: \ell^1 \rightarrow \mathbb{R} \\ (x_i)_{i \geq 1} &\mapsto \sup_{i \geq 1} x_i \end{aligned}$$

is continuous (this is a deterministic argument), and the result follows from preservation of convergence in distribution under composition with a continuous function.

**Solution to Exercise 3.18.** [statement p. 127]

(1) Write

$$d_F(\mu, \nu) = \inf\{r > 0 : \forall F \text{ closed in } E, \mu(F) \leq \nu(F^r) + r\}.$$

First, we have  $d_{LP}(\mu, \nu) \geq d_F(\mu, \nu)$ . Indeed, if we choose  $r > d_{LP}(\mu, \nu)$ , then

$$\mu(A) \leq \nu(A^r) + r$$

for every Borel set  $A$ , and hence for every closed set  $A$ , which implies that  $d_F(\mu, \nu) \leq r$ . Thus  $d_{LP}(\mu, \nu) \geq d_F(\mu, \nu)$ .

Now choose  $r > d_F(\mu, \nu)$ , so that

$$\mu(F) \leq \nu(F^r) + r$$

for every closed set  $F$ . For a Borel set  $A$ , we then have

$$\mu(A) \leq \mu(\overline{A}) \leq \nu(\overline{A}^r) + r = \nu(A^r) + r$$

because  $\overline{A}^r = A^r$  (indeed, the inclusion  $A^r \subset \overline{A}^r$  is clear, and for the other inclusion, if  $x \in \overline{A}^r$ , we find  $y \in \overline{A}$  such that  $d(x, y) < r$ , then  $z \in A$  such that  $d(y, z) < r - d(x, y)$ , so that  $d(x, z) < r$ ). This yields  $d_{LP}(\mu, \nu) \leq r$ , which completes the proof.

(2) Yes! As in the previous question, we have  $d_{LP}(\mu, \nu) \geq d_O(\mu, \nu)$ . For the reverse inequality, let  $r > d_O(\mu, \nu)$ , so that

$$\mu(O) \leq \nu(O^r) + r$$

for every open set  $O$ . Let  $A$  be a Borel set; we will verify that  $\mu(A) \leq \nu(A^{r+\eta}) + r + \eta$  for every  $\eta > 0$ , which will imply  $d_{LP}(\mu, \nu) \leq r + \eta$  and complete the proof. To this end, write

$$\mu(A) \leq \mu(A^\eta) \leq \nu((A^\eta)^r) + r \leq \nu(A^{\eta+r}) + r + \eta.$$

\* \* \*

**Solution to Exercise 3.19.** [statement p. 128] By tightness, for every  $n \geq 1$ , there exists a compact set  $K_n \subset E$  such that  $\mu(K_n) \geq 1 - 1/n$ . Set

$$A = \bigcup_{n=1}^{\infty} K_n.$$

Then for every  $n \geq 1$ ,  $\mu(A) \geq 1 - 1/n$ , so letting  $n \rightarrow \infty$  gives  $\mu(A) = 1$ . Moreover,  $A$  is separable, being a countable union of separable sets (since they are compact).

**Solution to Exercise 3.20.** [statement p. 128]

(i)  $\implies$  (ii) This follows from continuity of the map

$$\begin{aligned} \mathcal{M}_\infty(\mathbb{R}) &\rightarrow \mathbb{R} \\ (a_{i,j})_{i,j \geq 1} &\mapsto a_{1,2} \end{aligned}$$

(ii)  $\implies$  (i) Let  $\varepsilon > 0$ . For  $i, j \geq 1$ , let  $C_{i,j} > 0$  be such that for every  $n \geq 1$  we have

$$\mathbb{P}(d_n(X_n(1), X_n(2)) \geq C_{i,j}) \leq \frac{\varepsilon}{2^{i+j}}.$$

We verify (using a diagonal subsequence argument) that

$$K = \{(a_{i,j})_{i,j \geq 1} \in \mathcal{M}_\infty(\mathbb{R}) : |a_{i,j}| \leq C_{i,j} \text{ for all } i, j \geq 1\}$$

is a compact subset of  $\mathcal{M}_\infty(\mathbb{R})$ , and that

$$\begin{aligned} \mathbb{P}((d_n(X_n(i), X_n(j)))_{i,j \geq 1} \notin K) &\leq \sum_{i,j=1}^{\infty} \mathbb{P}(d_n(X_n(i), X_n(j)) \geq C_{i,j}) \\ &\leq \sum_{i,j=1}^{\infty} \mathbb{P}(d_n(X_n(1), X_n(2)) \geq C_{i,j}) \\ &\leq \sum_{i,j=1}^{\infty} \frac{\varepsilon}{2^{i+j}} \\ &= \varepsilon. \end{aligned}$$

(iii)  $\implies$  (ii) Let  $x_n \in E_n$  be such that the sequence  $(d_n(x_n, X_n(1)))_{n \geq 1}$  is tight in  $\mathbb{R}$ . Fix  $\varepsilon > 0$  and  $C > 0$  such that  $\mathbb{P}(d_n(x_n, X_n(1)) \geq C/2) \leq \varepsilon/2$ . Then by the triangle inequality

$$\begin{aligned} \mathbb{P}(d_n(X_n(1), X_n(2)) > C) &\leq \mathbb{P}(d_n(x_n, X_n(1)) > C/2) + \mathbb{P}(d_n(x_n, X_n(2)) > C/2) \\ &= 2\mathbb{P}(d_n(x_n, X_n(1)) > C/2) \\ &\leq \varepsilon \end{aligned}$$

(ii)  $\implies$  (iii) For every  $\varepsilon > 0$ , let  $C_\varepsilon > 0$  be such that  $\mathbb{P}(d_n(X_n(1), X_n(2)) > C_\varepsilon) \leq \varepsilon$  for every  $n \geq 1$

Since Fubini's theorem gives

$$\mathbb{P}(d_n(X_n(1), X_n(2)) > C) = \int_{E_n} \mu_n(dx) \mathbb{P}(d_n(x, X_n(1)) > C),$$

there exists  $x_n^\varepsilon \in E_n$  such that  $\mathbb{P}(d_n(x_n^\varepsilon, X_n(1)) > C_\varepsilon) \leq \varepsilon$  for every  $n \geq 1$ .

Let us verify that  $x_n = x_n^{1/4}$  works. For  $\varepsilon < 1/2$ , the event

$$\{d_n(x_n^{1/4}, X_n(1)) \leq C_{1/4}\} \cap \{d_n(x_n^\varepsilon, X_n(1)) \leq C_\varepsilon\}$$

is nonempty because it has probability at least  $1 - \varepsilon - 1/4$ . We deduce that  $d_n(x_n^{1/4}, x_n^\varepsilon) \leq C_\varepsilon + C_{1/4}$ . Thus,

$$\mathbb{P}(d_n(x_n^{1/4}, X_n(1)) > 2C_\varepsilon + C_{1/4}) \leq \mathbb{P}(d_n(x_n^\varepsilon, X_n(1)) > C_\varepsilon) \leq \varepsilon,$$

which completes the proof.

**Solution to Exercise 3.21.** [statement p. 128]

- (1) For bounded continuous  $f$ , write  $\ell(f)$  for the limit of  $(\int f d\mu_n)_{n \geq 1}$ . Since  $E$  is compact,  $(\mu_n)_{n \geq 1}$  is tight. Let  $\psi$  index a subsequence and let  $\mu$  be a probability measure on  $E$  such that  $\mu_{\psi(n)} \Rightarrow \mu$ . In particular,  $\ell(f) = \int f d\mu$ .

Let us verify that  $\mu_n \Rightarrow \mu$ . By Prokhorov's theorem, it suffices to verify that if  $\mu_{\phi(n)} \Rightarrow \nu$ , then  $\nu = \mu$ . By assumption,  $\mu_{\phi(n)}(f) \rightarrow \ell(f)$ , hence  $\int f d\nu = \ell(f) = \int f d\mu$ . We conclude that  $\nu = \mu$ .

- (2) Set  $\ell(f) = \lim_n \mu_n(f)$  for  $f \in \mathcal{C}_b(E)$ . We will show that  $(\mu_n)$  is tight. The conclusion will then follow exactly as in (1), by Prokhorov's theorem and uniqueness of the limit.

Suppose, for a contradiction, that  $(\mu_n)$  is not tight. There then exist  $r > 0$  and  $\eta > 0$  such that, for every finite subset  $F \subset E$ , there exists  $n$  satisfying

$$\mu_n \left( E \setminus \bigcup_{x \in F} B(x, r) \right) > \eta.$$

Indeed, if this property were false, for every  $\varepsilon > 0$  and every  $k \geq 1$ , one could choose a finite union of balls of radius  $2^{-k}$  whose mass is at least  $1 - \varepsilon 2^{-k}$  for all  $\mu_n$ . The intersection of the corresponding unions of closed balls would be closed and totally bounded, hence compact since  $E$  is complete, and would have mass at least  $1 - \varepsilon$  for all  $\mu_n$ . Moreover, the index  $n$  witnessing this property may be chosen arbitrarily large: the finitely many measures corresponding to the initial indices are tight and can be absorbed by adding finitely many balls to the union.

Set  $n_0 = 0$  and inductively construct a subsequence indexed by  $(n_j)$ , compact sets  $K_j, H_j$  and functions  $h_j : E \rightarrow [0, 1]$ . Suppose the objects with indices strictly less than  $j$  have been constructed. Choose a finite union  $A_j$  of open balls of radius  $r$  whose centers form an  $r/4$ -net of  $K_1 \cup \dots \cup K_{j-1}$ ; for  $j = 1$ , set  $A_1 = \emptyset$ . Then choose  $n_j > n_{j-1}$  large enough that

$$\mu_{n_j}(A_j^c) > \eta, \quad \sum_{i < j} |\mu_{n_j}(h_i) - \ell(h_i)| < \eta/8.$$

This is possible by the preceding property and convergence of the integrals of the functions  $h_i$  already constructed. Take a compact set  $K_j$  such that  $\mu_{n_j}(K_j^c) < \eta/8$ , and set

$$H_j = K_j \setminus A_j, \quad h_j(x) = \left( 1 - \frac{4}{r} d(x, H_j) \right)_+.$$

Then  $H_j$  is compact and  $\mu_{n_j}(h_j) > 7\eta/8$ . By construction,  $H_j$  is at distance at least  $3r/4$  from the  $K_i$ ,  $i < j$ . The sets  $\{h_j > 0\}$  are therefore pairwise at distance at least  $r/4$ , and, for  $j > i$ ,  $\{h_j > 0\} \cap K_i = \emptyset$ .

The functions  $h_j$  are nonnegative and the sets where they are positive are disjoint. Thus  $\sum_{j=1}^m h_j \leq 1$  for every  $m$ , and then  $\sum_{j \geq 1} \ell(h_j) \leq 1$  by positivity of  $\ell$ . Choose  $J$  such that  $\sum_{j \geq J} \ell(h_j) < \eta/8$  and set

$$f = \sum_{\substack{j \geq J \\ j \text{ even}}} h_j.$$

Since the sets where the functions are positive are disjoint, this sum is also the supremum of the functions considered. It is therefore  $4/r$ -Lipschitz and takes values between 0 and 1. For even  $j \geq J$ ,  $\mu_{n_j}(f) > 7\eta/8$ . For odd  $j \geq J$ , on the other hand, we have

$$\begin{aligned} \mu_{n_j}(f) &\leq \sum_{\substack{J \leq i < j \\ i \text{ even}}} \mu_{n_j}(h_i) + \mu_{n_j}(K_j^c) \\ &\leq \sum_{i \geq J} \ell(h_i) + \sum_{i < j} |\mu_{n_j}(h_i) - \ell(h_i)| + \mu_{n_j}(K_j^c) < 3\eta/8. \end{aligned}$$

The sequence  $(\mu_n(f))$  therefore does not converge, a contradiction. The family  $(\mu_n)$  is tight, which completes the proof.

This result is due to Alexandrov. See Ramamoorthi, R.V., Rao, B.V. & Sethuraman, J., *A note on weak convergence*, Sankhya A 74, 269–276 (2012), <https://doi.org/10.1007/s13171-012-0016-6>.

\* \* \*

### Solution to Exercise 3.22. [statement p. 128]

- (1) Let  $\varepsilon > 0$  and let  $K$  be a compact subset of  $E$  such that  $\mathbb{P}(X_1 \in K) \geq 1 - \varepsilon/n$ . Let

$$\begin{aligned} \Phi : E^n &\rightarrow E \\ (x_1, \dots, x_n) &\mapsto x_1 + \dots + x_n \end{aligned}$$

which is continuous. Then, writing  $K_n = \Phi(K \times K \times \dots \times K)$ , which is compact as the continuous image of a compact set, we have

$$\mathbb{P}(X_1 + \dots + X_n \notin K_n) \leq \mathbb{P}(\exists 1 \leq i \leq n : X_i \notin K) \leq \sum_{i=1}^n \mathbb{P}(X_i \notin K) \leq \sum_{i=1}^n \frac{\varepsilon}{n} = \varepsilon,$$

so that  $\mathbb{P}(X_1 + \dots + X_n \in K_n) \geq 1 - \varepsilon$ , which proves the result.

- (2) The answer is yes. It suffices to prove that if  $X$  and  $Y$  are two independent random variables such that  $X + Y$  is tight, then  $X$  and  $Y$  are tight. The desired result then follows by induction.

To this end, let  $\varepsilon > 0$  and let  $K$  be compact such that  $\mathbb{P}(X + Y \in K) > 1 - \varepsilon$ . Write  $\mathbb{P}_X$  and  $\mathbb{P}_Y$  for the distributions of  $X$  and  $Y$ , respectively, so that Fubini's theorem gives

$$1 - \varepsilon < \int_{E \times E} \mathbb{1}_K(x + y) \mathbb{P}_X(dx) \mathbb{P}_Y(dy) = \int_E \mathbb{P}_Y(K - x) \mathbb{P}_X(dx).$$

It follows that there exists  $x \in E$  such that  $\mathbb{P}_Y(K - x) > 1 - \varepsilon$ . Since  $K - x$  is compact as the image of a compact set under the map  $z \mapsto z - x$ , this shows that  $Y$  is tight, and by symmetry  $X$  is tight as well. This completes the proof.

\* \* \*

**Solution to Exercise 3.23.** [statement p. 129]

- (1) The equivalence between (a) and (c) follows from the general fact that  $X_n \rightarrow X$  in probability if and only if every subsequence has a further subsequence converging almost surely to  $X$ .

Let us show that (c) implies (b). Let  $f : E \rightarrow \mathbb{R}$  be bounded and continuous. It suffices to show that for every subsequence indexed by  $\phi$ , there exists a further subsequence indexed by  $\psi$  such that  $\int f dM_{\phi \circ \psi(n)} \rightarrow \int f dM$  almost surely. Let  $\phi$  index a subsequence. By (c), let  $\psi$  index a further subsequence such that almost surely  $M_{\phi \circ \psi(n)}$  converges to  $M$  in  $(\mathcal{M}_1(E), d_{LP})$ . Since the metric space  $E$  is Polish, convergence in  $d_{LP}$  implies weak convergence; thus almost surely  $M_{\phi \circ \psi(n)}$  converges weakly to  $M$ , and hence  $\int f dM_{\phi \circ \psi(n)} \rightarrow \int f dM$  almost surely.

Finally, let us show that (b) implies (a). Let  $\phi$  index a subsequence. We show that there exists a further subsequence indexed by  $\psi$  such that almost surely  $d_{LP}(M_{\phi \circ \psi(n)}, M) \rightarrow 0$ . First, for every function  $f \in \mathcal{F}$ , we have  $\int f dM_{\phi(n)} \rightarrow \int f dM$  in probability, so there exists a further subsequence indexed by  $\psi$  such that  $\int f dM_{\phi \circ \psi(n)} \rightarrow \int f dM$  almost surely. Since  $\mathcal{F}$  is countable, a diagonal argument yields a further subsequence indexed by  $\psi$  such that

$$\text{almost surely, } \forall f \in \mathcal{F}, \int f dM_{\phi \circ \psi(n)} \xrightarrow{n \rightarrow \infty} \int f dM. \quad (4)$$

Let us verify that almost surely  $d_{LP}(M_{\phi \circ \psi(n)}, M) \rightarrow 0$ . Let  $\varepsilon > 0$ . By the recalled fact, there exist  $f_1, \dots, f_K \in \mathcal{F}$  and  $\delta > 0$  such that

$$\left\{ \nu \in \mathcal{M}_1(E) : \left| \int f_i d\nu - \int f_i dM \right| < \delta \text{ for every } 1 \leq i \leq K \right\} \subset B(M, \varepsilon).$$

**Caution.** The functions  $f_1, \dots, f_K$ , as well as  $K$  and  $\delta$ , may depend on the realization of  $M$ . We argue pathwise on the event of probability 1 in (4); no measurable selection is needed.

By (4), almost surely, for sufficiently large  $n$  we have

$$\left| \int f_i dM_{\phi \circ \psi(n)} - \int f_i dM \right| < \delta$$

for every  $1 \leq i \leq K$ . It follows that almost surely, for sufficiently large  $n$  we have  $d_{LP}(M_{\phi \circ \psi(n)}, M) \leq \varepsilon$ , and the result follows.

- (2) Let  $f \in \mathcal{C}_b(E, \mathbb{R})$  and set  $m = \int_E f d\mu$  and  $M_n(f) = \int_E f dM_n$ . Conditional independence of the two samples gives directly

$$\mathbb{E} \left[ (M_n(f) - m)^2 \right] = \mathbb{E} \left[ (f(\Pi_n^1) - m)(f(\Pi_n^2) - m) \right]. \quad (5)$$

The function  $(x, y) \mapsto (f(x) - m)(f(y) - m)$  is bounded and continuous. By convergence in distribution of the pair, the right-hand side therefore converges to

$$\mathbb{E} \left[ (f(\Pi^1) - m)(f(\Pi^2) - m) \right] = \mathbb{E} \left[ f(\Pi^1) - m \right] \mathbb{E} \left[ f(\Pi^2) - m \right] = 0,$$

because the limiting random variables are independent with distribution  $\mu$ . Thus  $M_n(f) \rightarrow \mu(f)$  in  $L^2$ , hence in probability. Question (1) gives  $d_{LP}(M_n, \mu) \rightarrow 0$  in probability.

**Remark.** The first question is inspired by Proposition 2.2 of [this article](#), and the second question comes from Lemma 3.1 in [this article](#).

\* \* \*

**Solution to Exercise 3.24.** [\[statement p. 129\]](#)

- (1) Write  $X$  for the almost sure limit of  $X_n$ . For  $\varepsilon > 0$ , write:

$$\mathbb{P}(d(Y_n, X) \geq 2\varepsilon) \leq \mathbb{P}(d(Y_n, X_n) \geq \varepsilon) + \mathbb{P}(d(X_n, X) \geq \varepsilon).$$

The first term of the sum converges to 0 by assumption, and the second also tends to 0 because almost sure convergence implies convergence in probability.

- (2) No: taking  $X_n = X$ , it suffices to choose a case in which convergence in probability holds without almost sure convergence. For example, take  $E = \mathbb{R}$ ,  $X_n = X = 0$  and  $(Y_n)_{n \geq 1}$  a sequence of independent random variables such that  $\mathbb{P}(Y_n = 1) = 1/n$  and  $\mathbb{P}(Y_n = 0) = 1 - 1/n$ .

\* \* \*

**Solution to Problem 3.25.** [\[statement p. 129\]](#)

- (1) One may use polynomials with rational coefficients multiplied by smooth cutoff functions, in the spirit of the families of test functions in Section 1.2.
- (2) Let  $L$  be a Lipschitz constant of  $f$ , and choose  $a < b$  such that  $\text{supp}(f) \subset ]a, b[$ . Set

$$c(x) = \max(a, \min(x, b)), \quad H(x) = L(c(x) - a),$$

$$F = \frac{H + f}{2}, \quad G = \frac{H - f}{2}.$$

Since  $f(x) = f(c(x))$ , for all  $x < y$ ,

$$|f(y) - f(x)| \leq L(c(y) - c(x)) = H(y) - H(x).$$

Thus  $F$  and  $G$  are nondecreasing. They are bounded and  $L$ -Lipschitz, since  $H$  and  $f$  are, and  $f = F - G$ .

- (3) Since  $Z$  has a density, the roots of  $P_n$  are almost surely distinct. By Rolle's theorem, the roots of  $P'_n$  are real and lie between the zeros of  $P_n$ . Thus write  $Z_1^n < Z_2^n < \dots < Z_n^n$  for the roots of  $P_n$ ; reindexing the  $Y_i^n$  if necessary, we may assume that

$$Z_1^n < Y_1^n < Z_2^n < \dots < Y_{n-1}^n < Z_n^n.$$

By questions (1) and (2), it suffices to consider each bounded nondecreasing Lipschitz function  $f$  appearing in the decompositions of the elements of the countable set  $D$ . By interlacing,

$$\frac{\sum_{k=1}^n f(Z_k) - f(Z_1^n)}{n-1} \leq \frac{1}{n-1} \sum_{k=1}^{n-1} f(Y_k^n) \leq \frac{\sum_{k=1}^n f(Z_k) - f(Z_1^n)}{n-1}.$$

Both bounds converge almost surely to  $\mathbb{E}[f(Z)]$  by the strong law of large numbers and boundedness of  $f$ . Taking the intersection of the probability-one events for this countable family shows that the empirical measures of the roots of  $P'_n$  converge almost surely to the distribution of  $Z$ .

- (4) Set  $N_i^n = \sum_{j=1}^n \mathbb{1}_{\{Z_j = \alpha_i\}}$ ,  $J_n = \{i : N_i^n > 0\}$  and  $K_n = |J_n|$ . The polynomial

$$P_n(X) = \prod_{i \in J_n} (X - \alpha_i)^{N_i^n}$$

has a derivative with a root of multiplicity  $N_i^n - 1$  at each  $\alpha_i$ ,  $i \in J_n$ , as well as  $K_n - 1$  other roots, counted with multiplicity, denoted by  $W_j^n$ . For bounded continuous  $f : \mathbb{C} \rightarrow \mathbb{R}$ ,

$$\frac{1}{n-1} \sum_{j=1}^{n-1} f(Y_j^n) = \frac{1}{n-1} \left( \sum_{i \in J_n} (N_i^n - 1) f(\alpha_i) + \sum_{j=1}^{K_n-1} f(W_j^n) \right).$$

Consequently, for  $n \geq 2$ ,

$$\left| \frac{1}{n-1} \sum_{j=1}^{n-1} f(Y_j^n) - \frac{1}{n} \sum_{j=1}^n f(Z_j) \right| \leq \frac{2K \|f\|_\infty}{n-1}.$$

By the strong law of large numbers, almost surely, simultaneously for  $1 \leq i \leq K$ ,  $N_i^n/n \rightarrow \mathbb{P}(Z_1 = \alpha_i)$ . On this event, the empirical measure of the  $Z_j$  converges weakly to the distribution of  $Z_1$ , and the preceding estimate shows the same convergence for the empirical measure of the roots of  $P'_n$ .

\* \* \*

### Solution to Problem 3.26. [statement p. 130]

- (1) The identity of indiscernibles and symmetry of  $d$  are clear. For the triangle inequality, observe that for  $u, v, w \in \overline{\mathcal{U}}$  we have  $d(u, w) \leq \max(d(u, v), d(v, w))$ . Indeed, if  $d(u, w) > d(u, v)$ , then  $u \wedge w \preceq u \wedge v$ , so that  $v \wedge w = u \wedge w$  and  $d(u, w) = d(v, w)$ .

For separability,  $\mathcal{U}$  is dense in  $\overline{\mathcal{U}}$ . Indeed, if  $u \in \partial\mathcal{U}$  and  $[u]_k \in \mathcal{U}$  denotes the first  $k$  integers of  $u$ , clearly  $d([u]_k, u) \rightarrow 0$  as  $k \rightarrow \infty$ .

For every  $u \in \overline{\mathcal{U}}$  and  $k \geq 1$ , write  $[u]_k$  for its prefix of length  $k$  if  $u$  has at least  $k$  letters, and for  $u$  itself otherwise. Let  $(u_n)$  be a Cauchy sequence. Truncation

is 1-Lipschitz; two distinct words of length at most  $k$  are at distance at least  $e^{-k}$ . Thus, for each  $k$ ,  $[u_n]_k$  is eventually constant, with value  $v(k)$ . These values are compatible:  $[v(\ell)]_k = v(k)$  for  $k \leq \ell$ . They define a unique finite or infinite word  $v$  of which they are the truncations, and  $u_n \rightarrow v$ . The space is therefore complete and, being separable, Polish.

- (2) Let  $u \in \overline{\mathbb{U}}$  and  $r > 0$ . Let us verify that  $B(u, r) = \{v \in \overline{\mathbb{U}} : d(u, v) < r\}$  is closed. To this end, let  $v \in \overline{\mathbb{U}} \setminus B(u, r)$ , so that  $d(u, v) \geq r$ , and verify that  $B(v, r) \cap B(u, r) = \emptyset$ . Indeed, arguing by contradiction, if  $w \in B(v, r) \cap B(u, r)$ , we have  $d(w, v) < r$  and  $d(w, u) < r$ , and then  $d(u, v) \leq \max(d(u, w), d(w, v)) < r$ , a contradiction.
- (3) First verify that the sets of the form  $\{u\}$  and  $T(u)$  for  $u \in \mathbb{U}$  are both open and closed. This follows from the previous question, observing that for  $u \in (\mathbb{N}^*)^n$  we have  $\{u\} = B(u, e^{-n-1})$  and  $T(u) = B(u, e^{-n+1/2})$ .

The forward implication then follows immediately from the Portmanteau theorem.

For the converse, by separability, every open set can be written as a countable union of open balls. Observe that the open balls in  $\overline{\mathbb{U}}$  are precisely of the form  $\{u\}$  and  $T(u)$  for  $u \in \mathbb{U}$ , and that two open balls are either disjoint or one contains the other. Consequently, every open subset  $O$  of  $\overline{\mathbb{U}}$  can be written as a countable disjoint union of sets of the preceding form (open and closed). Thus, writing  $O = \coprod_{k \geq 1} O_k$ , Fatou's lemma gives:

$$\mu(O) = \sum_{k=1}^{\infty} \mu(O_k) = \sum_{k=1}^{\infty} \liminf_{n \rightarrow \infty} \mu_n(O_k) \leq \liminf_{n \rightarrow \infty} \sum_{k=1}^{\infty} \mu_n(O_k) = \liminf_{n \rightarrow \infty} \mu_n(O),$$

which completes the proof.

\* \* \*

### Solution to Problem 3.27. [statement p. 131]

- (1) The equivalence between (b) and (c) follows from the fact that convergence in distribution and convergence in probability to a constant are equivalent.

Let us show that (c) implies (d). Let  $f \in \mathcal{C}_b(E, \mathbb{R})$ . To show that  $\int f dM_n$  converges in probability to  $\int f d\mu$ , we will show that for every subsequence indexed by  $\phi$ , there exists a further subsequence indexed by  $\psi$  such that  $\int f dM_{\phi \circ \psi(n)}$  converges almost surely to  $\int f d\mu$ . Let  $\phi$  index a subsequence. Since  $M_n$  converges in probability to  $\mu$ , there exists a further subsequence indexed by  $\psi$  such that almost surely  $M_{\phi \circ \psi(n)}$  converges (in  $d_{LP}$ ) to  $\mu$ . Then almost surely  $M_{\phi \circ \psi(n)}$  converges weakly to  $\mu$ , so almost surely  $\int f dM_{\phi \circ \psi(n)}$  converges to  $\int f d\mu$ .

Alternatively, to show that (c) implies (d), one may observe that for  $f \in \mathcal{C}_b(E, \mathbb{R})$ , the map  $\nu \mapsto \int f d\nu$  is continuous, and that convergence in probability is preserved under composition with a continuous function.

Let us show that (d) implies (c). Let  $\varepsilon > 0$ . We need to show that

$$\mathbb{P}(d_{LP}(M_n, \mu) \geq \varepsilon) \rightarrow 0,$$

that is,  $\mathbb{P}(M_n \notin B(\mu, \varepsilon)) \rightarrow 0$ . By the fact given at the beginning of the exercise, there exist  $f_1, \dots, f_K \in \mathcal{C}_b(E, \mathbb{R})$  and  $\delta > 0$  such that

$$\left\{ \nu \in \mathcal{M}_1(E) : \left| \int f_i d\nu - \int f_i d\mu \right| < \delta \text{ for every } 1 \leq i \leq K \right\} \subset B(\mu, \varepsilon).$$

Then

$$\mathbb{P}(M_n \notin B(\mu, \varepsilon)) \leq \sum_{i=1}^K \mathbb{P}\left(\left| \int f_i dM_n - \int f_i d\mu \right| \geq \delta\right)$$

which tends to 0 by assumption.

The fact that (a) implies (d) follows immediately from Markov's inequality.

The fact that (d) implies (a) follows from the fact that the sequence  $(\int f dM_n)_{n \geq 1}$  is uniformly integrable (being bounded by  $\|f\|_\infty$ ).

The implication (c)  $\Rightarrow$  (e) follows from continuity of  $\nu \mapsto \int e^{itx} \nu(dx)$ , considering the real and imaginary parts separately.

Let us show (e)  $\Rightarrow$  (c). As explained in the following remark, we may work on a common probability space. For every  $t$ , convergence in probability and the bound  $|\Phi_n(t) - \Phi(t)| \leq 2$  imply  $\mathbb{E}[|\Phi_n(t) - \Phi(t)|] \rightarrow 0$ . By Fubini's theorem and dominated convergence,

$$\mathbb{E} \left[ \int_{\mathbb{R}} e^{-|t|} |\Phi_n(t) - \Phi(t)| dt \right] \rightarrow 0.$$

The integrands are jointly measurable because they are measurable in  $\omega$  and continuous in  $t$ . From every subsequence, we can therefore extract  $(n_j)$  such that

$$\sum_j \mathbb{E} \left[ \int_{\mathbb{R}} e^{-|t|} |\Phi_{n_j}(t) - \Phi(t)| dt \right] < \infty.$$

By Tonelli's theorem, almost surely,  $\Phi_{n_j}(t) \rightarrow \Phi(t)$  for Lebesgue-almost every  $t$ . Exercise 1.11(1) then gives  $M_{n_j} \xrightarrow{\text{weak}} \mu$  almost surely, hence  $d_{LP}(M_{n_j}, \mu) \rightarrow 0$  almost surely. The subsequence criterion yields (c).

In particular, for every compact set  $K \subset \mathbb{R}$ ,  $\sup_{t \in K} |\Phi_n(t) - \Phi(t)|$  converges in probability to zero: apply Exercise 1.6(2) along the further subsequences that converge almost surely.

**Remark.** Strictly speaking, care is needed because the random variables  $(M_n)_{n \geq 1}$  need not be defined on the same probability space, so one cannot speak of almost sure convergence (this point was brought to my attention by Dimitri Faure). However, without loss of generality, we may assume that the random variables

$(M_n)_{n \geq 1}$  are defined on the same probability space, for example by considering a sequence  $(M'_n)_{n \geq 1}$  of independent random variables defined on the same probability space such that for every  $n \geq 1$ ,  $M_n$  and  $M'_n$  have the same distribution (using the existence of product measures), and working with  $M'_n$  in place of  $M_n$  (the properties considered depend only on the individual distributions).

- (2) (a) Let us prove that  $M_n$  converges in probability to  $\mu$  by showing that for every subsequence indexed by  $\phi$ , there exists a further subsequence indexed by  $\psi$  such that almost surely  $M_{\phi \circ \psi(n)}$  converges to  $\mu$  (i.e., almost surely  $d_{LP}(M_{\phi \circ \psi(n)}, \mu) \rightarrow 0$ ). Let  $\phi$  index a subsequence. By a diagonal argument, there exists a further subsequence indexed by  $\psi$  such that almost surely, for every  $k \geq 1$ ,  $\int x^k dM_{\phi \circ \psi(n)}$  converges to  $\int x^k d\mu$ . By an exercise on the exercise sheet, almost surely  $M_{\phi \circ \psi(n)}$  converges weakly to  $\mu$ , hence almost surely  $d_{LP}(M_{\phi \circ \psi(n)}, \mu) \rightarrow 0$ .

Alternatively, without redefining the random variables on the same probability space, one can verify that  $(M_n)$  is tight and identify the limit using the converging together theorem, by considering the functional  $\nu \mapsto \int \max(\min(x^k, K), -K) d\nu$  with  $K \rightarrow \infty$ .

- (b) By Chebyshev's inequality, for  $\varepsilon > 0$  we have

$$\begin{aligned} & \mathbb{P} \left( \left| \int x^k dM_n - \int x^k d\mu \right| \geq 2\varepsilon \right) \\ & \leq \mathbb{P} \left( \left| \int x^k dM_n - \mathbb{E} \left[ \int x^k dM_n \right] \right| \geq \varepsilon \right) \\ & \quad + \mathbb{P} \left( \left| \mathbb{E} \left[ \int x^k dM_n \right] - \int x^k d\mu \right| \geq \varepsilon \right) \\ & \leq \frac{\text{Var}(\int x^k dM_n)}{\varepsilon^2} + \mathbb{1}_{|\mathbb{E}[\int x^k dM_n] - \int x^k d\mu| \geq \varepsilon} \end{aligned}$$

which tends to 0 by assumption, and we may apply question (a).

- (3) The equivalence between (a) and (b) is simply the definition of weak convergence.

The fact that (b) implies (c) is clear.

To show that (c) implies (b), fix  $\varepsilon > 0$  and show that almost surely, for sufficiently large  $n$ ,  $d_{LP}(M_n, \mu) \leq \varepsilon$ . By the fact given at the beginning of the exercise, there exist  $f_1, \dots, f_K \in \mathcal{C}_b(E, \mathbb{R})$  and  $\delta > 0$  such that

$$\left\{ \nu \in \mathcal{M}_1(E) : \left| \int f_i d\nu - \int f_i d\mu \right| < \delta \text{ for every } 1 \leq i \leq K \right\} \subset B(\mu, \varepsilon).$$

By assumption, almost surely, for sufficiently large  $n$

$$\left| \int f_i dM_n - \int f_i d\mu \right| < \delta \text{ for every } 1 \leq i \leq K$$

Thus, almost surely, for sufficiently large  $n$ ,  $d_{LP}(M_n, \mu) \leq \varepsilon$ .

The fact that (b) implies (d) is clear: on the event of probability 1 on which, for every  $f \in \mathcal{C}_b(E, \mathbb{R})$ ,  $\int f dM_n \rightarrow \int f d\mu$ , we have  $\int e^{itx} dM_n(dx) \rightarrow \int e^{itx} \mu(dx)$  for every  $t \in \mathbb{R}$ .

To show that (d) implies (b), the idea is to use Fubini's theorem to interchange the quantifiers. Write

$$A = \left\{ (\omega, t) \in \Omega \times \mathbb{R} : \int e^{itx} dM_n(\omega)(dx) \not\rightarrow \int e^{itx} \mu(dx) \right\}$$

and its sections

$$A_\omega = \{t \in \mathbb{R} : (\omega, t) \in A\}, \quad A^t = \{\omega \in \Omega : (\omega, t) \in A\}.$$

By assumption, for every  $t \in \mathbb{R}$  we have  $\mathbb{P}(A^t) = 0$ . Apply Fubini's theorem, writing  $\lambda$  for Lebesgue measure:

$$0 = \int_{\mathbb{R}} \mathbb{P}(A^t) \lambda(dt) = \int_{\Omega \times \mathbb{R}} \mathbb{1}_A d\mathbb{P} \otimes d\lambda = \int_{\Omega} \lambda(A_\omega) d\mathbb{P}.$$

We deduce that  $\lambda(A_\omega) = 0$  for almost every  $\omega \in \Omega$ . Thus, almost surely, for  $\lambda$ -almost every  $t \in \mathbb{R}$  we have  $\int e^{itx} dM_n(\omega)(dx) \rightarrow \int e^{itx} \mu(dx)$ . We conclude using question (1) of Exercise 1.11 (Lévy's theorem under almost everywhere convergence in frequency).

- (4) This is a direct application of the exercise on the exercise sheet mentioned above.
- (5) By the strong law of large numbers, for every  $f \in \mathcal{C}_b(E, \mathbb{R})$ , for almost every  $\omega \in \Omega$ ,

$$\int f dM_n = \frac{1}{n} \sum_{k=1}^n f(X_k) \rightarrow \mathbb{E}[f(X_1)] = \int f d\mu.$$

By question (3), we deduce that almost surely  $M_n$  converges weakly to  $\mu$  (and hence  $M_n$  converges in probability and in distribution to  $\mu$ ).

\* \* \*

### Solution to Problem 3.28. [statement p. 132]

- (1) By compactness,  $E$  is bounded. There therefore exists  $M > 0$  such that for every  $x \in E$  we have  $\overline{B}(x, M) = E$ . Thus  $\lim_{M \rightarrow \infty} I_\mu(M) = 1$ .
- (2) Clearly  $I_\mu$  is nondecreasing, which implies existence of the left-hand limits. To prove right-continuity, argue by contradiction, assuming that  $I_\mu$  is not right-continuous at  $\varepsilon_0 > 0$ . Then by monotonicity

$$I_\mu(\varepsilon_0) < \lim_{\varepsilon \downarrow \varepsilon_0} I_\mu(\varepsilon).$$

We can therefore find  $x_0 \in E$  and  $\delta > 0$  such that for every  $n \geq 1$ ,

$$\mu(\overline{B}(x_0, \varepsilon_0)) + \delta < I_\mu(\varepsilon_0 + 1/n).$$

In particular, for every  $n \geq 1$  we have

$$\mu(\overline{B}(x_0, \varepsilon_0)) + \delta < \mu(\overline{B}(x_0, \varepsilon_0 + 1/n)).$$

Since  $\mu(\overline{B}(x_0, \varepsilon_0 + 1/n)) \rightarrow \mu(\overline{B}(x_0, \varepsilon_0))$  as  $n \rightarrow \infty$ , we obtain a contradiction.

(3) For every  $x \in E$  and every  $r > 0$ ,

$$\inf_{y \in E} \mu(\{y\}) \leq I_\mu(r) \leq \mu(\overline{B}(x, r)).$$

Letting  $r \downarrow 0$  and then taking the infimum over  $x$ , we obtain

$$\lim_{r \downarrow 0} I_\mu(r) = \inf_{x \in E} \mu(\{x\}).$$

The limit is therefore zero if  $\mu$  has no atoms. In fact, it is always zero when  $E$  is infinite: otherwise infinitely many distinct points would all have mass bounded below by the same strictly positive constant.

(4) Camille is wrong: for example, if  $E = [0, 1]$  and  $\lambda$  denotes Lebesgue measure, then  $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\lambda$  satisfies  $\lim_{\varepsilon \downarrow 0} I_\mu(\varepsilon) = 0$  but  $\mu(\{0\}) > 0$ .

(5) Fix  $\varepsilon > 0$ . For every  $x \in E$ , the Portmanteau theorem applied to the closed set  $\overline{B}(x, \varepsilon)$  gives

$$\limsup_{n \rightarrow \infty} I_{\mu_n}(\varepsilon) \leq \limsup_{n \rightarrow \infty} \mu_n(\overline{B}(x, \varepsilon)) \leq \mu(\overline{B}(x, \varepsilon)).$$

Taking the infimum over  $x$ , we obtain  $\limsup_n I_{\mu_n}(\varepsilon) \leq I_\mu(\varepsilon)$ .

(6) Assume (a). If  $I_\mu(\varepsilon) = 0$ , choose  $x_j$  such that  $\mu(\overline{B}(x_j, \varepsilon)) \rightarrow 0$ . By compactness, passing to a subsequence if necessary,  $x_j \rightarrow x$ . For sufficiently large  $j$ ,

$$\overline{B}(x, \varepsilon/2) \subset \overline{B}(x_j, \varepsilon).$$

Thus  $\mu(B(x, \varepsilon/2)) = 0$ , contradicting full support. Hence (a) implies (b).

Assume (b) and fix  $\varepsilon > 0$ . Choose  $x_1, \dots, x_k$  such that  $E = \bigcup_{i=1}^k B(x_i, \varepsilon/2)$ . By the Portmanteau theorem,

$$\liminf_n \mu_n(B(x_i, \varepsilon/2)) \geq \mu(B(x_i, \varepsilon/2)) \geq \mu(\overline{B}(x_i, \varepsilon/4)) \geq I_\mu(\varepsilon/4) > 0.$$

For every  $x \in E$ , some index  $i$  satisfies  $d(x, x_i) < \varepsilon/2$ , so  $B(x_i, \varepsilon/2) \subset \overline{B}(x, \varepsilon)$ . Thus

$$I_{\mu_n}(\varepsilon) \geq \min_{1 \leq i \leq k} \mu_n(B(x_i, \varepsilon/2)),$$

and its limit inferior is strictly positive. This proves (c).

Finally, assume (c). Let  $O$  be a nonempty open set and choose  $x \in O$ ,  $r > 0$  such that  $\overline{B}(x, r) \subset O$ . By (5),

$$\mu(O) \geq I_\mu(r) \geq \limsup_n I_{\mu_n}(r) \geq \liminf_n I_{\mu_n}(r) > 0.$$

Hence (c) implies (a), which completes the proof.

- (7) Since  $E$  is separable,  $(\mathcal{M}_1(E), d_{LP})$  is also separable. We may therefore use Skorokhod's representation theorem and assume that  $M_n$  converges to  $M$  almost surely in  $(\mathcal{M}_1(E), d_{LP})$ .

Assume (a) and fix  $\varepsilon, \varepsilon' > 0$ . By (6), on the almost surely convergent representation,

$$\liminf_n I_{M_n}(\varepsilon) > 0 \quad \text{almost surely.}$$

As  $j$  and  $N$  increase, the events  $\{I_{M_n}(\varepsilon) \geq 1/j \text{ for every } n \geq N\}$  therefore cover an event of probability one. Thus there exist  $j, N$  such that this event has probability strictly greater than  $1 - \varepsilon'$ . With  $\delta = 1/j$ , we obtain  $\mathbb{P}(I_{M_n}(\varepsilon) < \delta) < \varepsilon'$  for every  $n \geq N$ .

Conversely, assume (b). Fix  $k \geq 1$  and  $\varepsilon' > 0$ , then choose  $\delta, N$  as in (b) for  $\varepsilon = 1/k$ . By (5), the map  $\nu \mapsto I_\nu(1/k)$  is upper semicontinuous, so  $\{\nu : I_\nu(1/k) < \delta\}$  is open. By the Portmanteau theorem,

$$\mathbb{P}(I_M(1/k) < \delta) \leq \liminf_n \mathbb{P}(I_{M_n}(1/k) < \delta) \leq \varepsilon'.$$

Thus  $\mathbb{P}(I_M(1/k) = 0) \leq \varepsilon'$ . Since  $\varepsilon' > 0$  is arbitrary, this probability is zero. Almost surely,  $I_M(1/k) > 0$  for every  $k \geq 1$ ; by monotonicity and (6),  $M$  has full support.

- (8) No, both implications are false. Here are counterexamples in the deterministic case. If  $E = [0, 1]$ ,  $\mu_{2n} = \delta_0$  and  $\mu_{2n+1} = \delta_1$ , the sequence  $(I_{\mu_n})_{n \geq 1}$  is constant but  $(\mu_n)_{n \geq 1}$  does not converge in distribution.

If  $E = [0, 2]$ , take, for  $n \geq 2$ ,

$$\mu_n = \frac{1}{3}\delta_0 + \frac{1}{3}\delta_{1/2} + \frac{1}{3}\delta_{1/2+1/n}.$$

Then

$$I_{\mu_n}(\varepsilon) = \begin{cases} 0 & \text{if } 0 < \varepsilon < 3/2 - 1/n, \\ 1/3 & \text{if } 3/2 - 1/n \leq \varepsilon < 3/2, \\ 2/3 & \text{if } 3/2 \leq \varepsilon < 2, \\ 1 & \text{if } 2 \leq \varepsilon. \end{cases}$$

These minimum values are attained by the ball centered at 2; for the radii in the last three lines, every ball contains at least, respectively, the last atom, the last two atoms, and then all the atoms. Now  $\mu_n$  converges weakly to  $\mu = \frac{1}{3}\delta_0 + \frac{2}{3}\delta_{1/2}$ , for which

$$I_\mu(\varepsilon) = \begin{cases} 0 & \text{if } 0 < \varepsilon < 3/2, \\ 2/3 & \text{if } 3/2 \leq \varepsilon < 2, \\ 1 & \text{if } 2 \leq \varepsilon. \end{cases}$$

On the compact interval  $[1, 7/4]$ ,  $I_{\mu_n}$  takes the value  $1/3$  for all sufficiently large  $n$ , while  $I_\mu$  takes only the values 0 and  $2/3$ . For every increasing bijective time change  $\lambda_n$  of this compact interval,  $\|I_{\mu_n} \circ \lambda_n - I_\mu\|_\infty \geq 1/3$ . Hence  $J_1$  convergence is impossible.

**Remark.** This problem is inspired by the following article:

Benedikt Stufler, Mass and radius of balls in Gromov-Hausdorff-Prokhorov convergent sequences, preprint available on arXiv, arxiv:2201.12251

<https://arxiv.org/abs/2201.12251>

This problem shows how to verify that a limiting measure has full support using estimates on the measures approximating it. However, the condition  $I_\mu(r) \rightarrow 0$  as  $r \downarrow 0$  does not characterize the absence of atoms, as illustrated by question (4).

## 4 The space of continuous functions on a compact set

Related lecture notes: [Chapter 4](#), p. 35.

### 4.1 Exercises

**Exercise 4.1.** [solution p. 162] For each  $n \geq 0$ , let  $X^n, Y^n$  be two processes in  $\mathcal{C}([0, 1], \mathbb{R})$ , defined on the same probability space (which may depend on  $n$ ). Suppose that  $X^n \rightarrow X$  and  $Y^n \rightarrow Y$  in distribution.

- (1) Alix says: the sequence  $((X^n, Y^n))_{n \geq 0}$  is tight in  $\mathcal{C}([0, 1], \mathbb{R})^2$ . Is she right?
- (2) We now view the pair  $(X^n, Y^n)$  as a single process

$$t \mapsto (X_t^n, Y_t^n), \quad t \in [0, 1]$$

in  $\mathcal{C}([0, 1], \mathbb{R}^2)$ . Billie says: the sequence  $(X^n, Y^n)_{n \geq 0}$  is tight in  $\mathcal{C}([0, 1], \mathbb{R}^2)$ . Is he right?

- (3) We now view the pair  $(X^n, Y^n)$  as a function  $(X_s^n, Y_t^n)_{s, t \in [0, 1]}$  in  $\mathcal{C}([0, 1]^2, \mathbb{R}^2)$ . Camille says: the sequence  $(X^n, Y^n)_{n \geq 0}$  is tight in  $\mathcal{C}([0, 1]^2, \mathbb{R}^2)$ . Is she right?

\* \* \*

**Exercise 4.2.** [solution p. 163] Let  $(X^n)_{n \geq 1}$  be a sequence of processes in  $\mathcal{C}([0, 1], \mathbb{R}^k)$ . Suppose that  $X^n \Rightarrow X$  in  $\mathcal{C}([0, 1], \mathbb{R}^k)$ . Let  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  be a continuous function. Set

$$Y_n = \int_0^1 f(X_s^n) ds, \quad Y = \int_0^1 f(X_s) ds.$$

Show that  $(X^n, Y_n) \Rightarrow (X, Y)$ .

\* \* \*

**Exercise 4.3.** [solution p. 163] Let  $X^n, X$  be nondecreasing continuous processes from  $[0, 1]$  to  $\mathbb{R}$ . Show that  $X^n \Rightarrow X$  if and only if  $X^n$  converges to  $X$  in finite-dimensional distributions.

\* \* \*

**Exercise 4.4.** [solution p. 164]

- (1) For  $a \geq 0$ , is the map  $f \mapsto t_a(f) = \inf\{t \geq 0 : f(t) \geq a\}$  from  $\mathcal{C} = \mathcal{C}(\mathbb{R}_+, \mathbb{R})$  to  $\mathbb{R}_+ \cup \{\infty\}$  continuous? If not, does it have any points of continuity, and if so, which ones?
- (2) Let  $(X^n, n \geq 0)$  be a sequence of processes in  $\mathcal{C}$ . Suppose that  $X^n$  converges in distribution to  $X$  for the topology of uniform convergence on compact sets. Also suppose that  $X_0^n = 0$  for every  $n \geq 1$ . For  $a \geq 0$ , write

$$T_a^n = t_a(X^n), \quad T_a = t_a(X).$$

Discuss whether  $T_\alpha^n$  converges to  $T_\alpha$ . Discuss the special case where  $X$  is standard Brownian motion (this last question is best addressed using the strong Markov property of Brownian motion; see the stochastic calculus course).

\* \* \*

**Exercise 4.5.** [solution p. 165] Show that  $\{f \in \mathbb{R}^{[0,1]} : f \text{ is continuous}\}$  is not an element of the product sigma-algebra  $\mathcal{B}(\mathbb{R})^{\otimes [0,1]}$ .

*Hint.* Consider a random variable  $U$  uniformly distributed on  $[0, 1]$  and the function  $X(t) = \mathbb{1}_{t \neq U}$  for  $0 \leq t \leq 1$ .

\* \* \*

**Exercise 4.6.** [solution p. 165] Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  equipped with a filtration, that is, a family  $(\mathcal{F}_t, t \geq 0)$  of sigma-algebras such that  $\mathcal{F}_s \subset \mathcal{F}_t$  for all  $s \leq t$ . Recall that a martingale is a family of random variables  $(X_t, t \geq 0)$  such that  $X_t \in L^1$  for every  $t \geq 0$  and such that, for all  $s \leq t$ , we have  $\mathbb{E}[X_t | \mathcal{F}_s] = X_s$ .

Let  $(X^n, n \geq 0)$  be a sequence of integrable processes in  $\mathcal{C} = \mathcal{C}([0, 1], \mathbb{R})$  adapted to filtrations  $(\mathcal{F}_t^n)$ , that is,  $X_t^n$  is measurable with respect to  $\mathcal{F}_t^n$  for every  $t \in [0, 1]$ . Suppose that  $X^n$  is an  $\mathcal{F}_t^n$ -martingale and that  $X^n$  converges in distribution to  $X$  in  $\mathcal{C}$ . Also suppose that, for every  $t \in [0, 1]$ ,  $(X_t^n)_{n \geq 1}$  is uniformly integrable.

Show that  $(X_t, 0 \leq t \leq 1)$  is an  $(\mathcal{F}_t)$ -martingale, where  $\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t)$ .

\* \* \*

**Exercise 4.7.** [solution p. 166] Let  $\mathcal{C}$  denote the set of continuous functions from  $[0, 1]$  to  $\mathbb{R}$ . Let  $(X_n(t) : 0 \leq t \leq 1)_{n \geq 1}$  be a sequence of  $\mathcal{C}$ -valued random variables. Write  $X_n = (X_n(t) : 0 \leq t \leq 1)$ .

(1) Prove that the sequence  $(X_n)_{n \geq 1}$  is tight in  $\mathcal{C}$  if and only if the following two conditions hold:

- (a) the sequence  $(X_n(0))_{n \geq 1}$  is tight in  $\mathbb{R}$ ;
- (b) for every sequence  $\delta_k \rightarrow 0$  and every sequence  $(n_k)_{k \geq 1}$ , we have

$$\omega(X_{n_k}, \delta_k) \xrightarrow[k \rightarrow \infty]{(\mathbb{P})} 0,$$

where  $\omega(f, \delta) = \sup_{|s-t| \leq \delta, 0 \leq s, t \leq 1} |f(s) - f(t)|$  denotes the modulus of continuity.

(2) Let  $(X_n^i)_{n \geq 1, 1 \leq i \leq n}$  be a sequence of  $\mathcal{C}$ -valued random variables such that, for every  $n \geq 1$ , the random variables  $(X_n^i)_{1 \leq i \leq n}$  have the same distribution as  $X_n$  (but are not assumed to be independent). Suppose that:

- (I) the sequence of random variables  $(X_n)_{n \geq 1}$  is tight in  $\mathcal{C}$ ;
- (II) the sequence of random variables  $\sup_{t \in [0, 1]} |X_n(t)|$  is uniformly integrable.

Prove that the sequence of random variables  $(Y_n)_{n \geq 1}$  is tight in  $\mathcal{C}$ , where  $Y_n$  is defined by

$$Y_n(t) = \frac{1}{n} \sum_{i=1}^n X_n^i(t), \quad t \in [0, 1].$$

(3) Does the result of question (2) remain true if (II) is replaced by (II'), defined by

(II') the sequence of random variables  $\sup_{t \in [0, 1]} |X_n(t)|$  is bounded in  $L^1$ ?

Justify your answer.

\* \* \*

**Exercise 4.8.** [solution p. 167] Let  $(X_i)_{i \geq 1}$  be i.i.d. random variables taking values in  $\mathcal{C}([0, 1])$ , the space of continuous functions from  $[0, 1]$  to  $\mathbb{R}$  equipped with the uniform norm. Suppose that there exist  $C, \varepsilon > 0$  such that

$$\begin{aligned} \forall 0 \leq s, t \leq 1, \quad \mathbb{E} \left[ (X_1(s) - X_1(t))^2 \right] &\leq C|s - t|^{1+\varepsilon}, \\ \forall 0 \leq s \leq 1, \quad \mathbb{E} [X_1(s)] &= 0 \quad \text{and} \quad \mathbb{E} [X_1(s)^2] < \infty. \end{aligned}$$

Set

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}.$$

(1) For  $0 \leq s, t \leq 1$ , compute  $\mathbb{E} [(Z_n(s) - Z_n(t))^2]$ .

(2) Show that  $(Z_n)_{n \geq 1}$  converges in distribution in  $\mathcal{C}([0, 1])$ .

\* \* \*

**Exercise 4.9.** [solution p. 168] For  $x \in \mathbb{R}$ , let  $\mathbb{P}_x$  denote the law of standard real-valued Brownian motion started at  $x$ . Let  $(x_n)_{n \geq 1}$  be a sequence of real numbers.

(1) If  $(x_n)_{n \geq 1}$  is bounded, show that the sequence  $(\mathbb{P}_{x_n})_{n \geq 1}$  is tight in  $\mathcal{C}([0, 1], \mathbb{R})$ .

(2) Does the result remain true if the sequence  $(x_n)_{n \geq 1}$  is no longer assumed to be bounded? Justify your answer.

\* \* \*

**Exercise 4.10.** [solution p. 168] Let  $X$  and  $Y$  be two random variables taking values in  $\mathcal{C}([0, 1], \mathbb{R})$ , the set of continuous functions from  $[0, 1]$  to  $\mathbb{R}$  equipped with the topology of uniform convergence. Suppose that, for every  $\varepsilon > 0$ , there exists a pair of random variables  $(X_\varepsilon, Y_\varepsilon)$  taking values in  $\mathcal{C}([0, 1], \mathbb{R})^2$  such that  $X_\varepsilon$  has the same distribution as  $X$ ,  $Y_\varepsilon$  has the same distribution as  $Y$ , and  $\mathbb{P} (\|X_\varepsilon - Y_\varepsilon\|_\infty > \varepsilon) < \varepsilon$ . Show that  $X$  and  $Y$  have the same distribution.

**Problem 4.11.** [solution p. 169] In this exercise, let  $\mathbb{R}^{[0,1]}$  denote the set of functions from  $[0, 1]$  to  $\mathbb{R}$  equipped with the product sigma-algebra  $\mathcal{B}(\mathbb{R})^{\otimes[0,1]}$ , and let  $\mathcal{C} = \mathcal{C}([0, 1], \mathbb{R})$  denote the vector space of real-valued continuous functions on  $[0, 1]$  equipped with the uniform norm and its Borel sigma-algebra. Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

- (1) Show that the Borel sigma-algebra of  $\mathcal{C}$  is the set of all  $A$  of the form  $A = B \cap \mathcal{C}$  for  $B \in \mathcal{B}(\mathbb{R})^{\otimes[0,1]}$ .

If  $X, Y : \Omega \rightarrow \mathbb{R}^{[0,1]}$  are two random variables, we say that  $Y$  is a *modification* (or *version*) of  $X$  if, for every  $t \in [0, 1]$ , we have  $\mathbb{P}(X(t) = Y(t)) = 1$ .

- (2) Let  $X, Y : \Omega \rightarrow \mathbb{R}^{[0,1]}$  be two random variables.
- (a) Check that if  $Y$  is a modification of  $X$ , then  $X$  and  $Y$  have the same distribution.
  - (b) Billie says: the converse of (a) is true. Is he right? Justify your answer.

For the rest of this exercise, fix a random variable  $Y : \Omega \rightarrow \mathcal{C}$ . (Thus  $Y_\omega$  is a continuous function for every  $\omega \in \Omega$ .)

- (3) Let  $X : \Omega \rightarrow \mathbb{R}^{[0,1]}$  be a random variable. Suppose that  $X$  and  $Y$  have the same distribution.
- (a) Let  $Y^*$  denote the random variable  $Y$  viewed as a function taking values in  $\mathbb{R}^{[0,1]}$ . Prove that there exists a measurable map  $F : \mathbb{R}^{[0,1]} \rightarrow \mathcal{C}$  such that  $Y = F \circ Y^*$ .  
*Hint.* You may use the Doob–Dynkin lemma, which states the following. If  $f, g$  are two measurable functions defined on  $(\Omega, \mathcal{F})$  and taking values in  $(S, \mathcal{S})$  and  $(T, \mathcal{T})$ , respectively, if  $f : (\Omega, \sigma(g)) \rightarrow (S, \mathcal{S})$  is measurable and if  $(S, \mathcal{S})$  is a Polish space equipped with its Borel sigma-algebra, then there exists a measurable map  $h : (T, \mathcal{T}) \rightarrow (S, \mathcal{S})$  such that  $f = h \circ g$ .
  - (b) Prove that there exists a random variable  $X' : \Omega \rightarrow \mathcal{C}$  such that  $X'$  is a modification of  $X$ .
- (4) (a) Let  $(U_n)_{n \geq 1}$  and  $(V_n)_{n \geq 1}$  be two sequences of random variables taking values in a complete separable metric space  $(E, d)$ . Suppose that the two sequences  $(U_n)_{n \geq 1}$  and  $(V_n)_{n \geq 1}$  have the same distribution. Show that  $\{(U_n)_{n \geq 1} \text{ converges}\}$  (that is, the set  $\{\omega \in \Omega : U_n(\omega) \text{ converges as } n \rightarrow \infty\}$ ) is an event (i.e. is measurable) and that

$$\mathbb{P}((U_n)_{n \geq 1} \text{ converges}) = \mathbb{P}((V_n)_{n \geq 1} \text{ converges}).$$

- (b) Consider a sequence of random variables  $(X_n)_{n \geq 1}$  taking values in  $\mathbb{R}^{[0,1]}$  such that  $X_n$  converges to  $Y$  in finite-dimensional distributions. Also suppose that, for every  $q \in [0, 1] \cap \mathbb{Q}$ , the sequence  $(X_n(q))_{n \geq 1}$  converges almost surely. Show that there exists a random variable  $X : \Omega \rightarrow \mathcal{C}$  such that, for every  $q \in [0, 1] \cap \mathbb{Q}$ , the sequence  $(X_n(q))_{n \geq 1}$  converges almost surely to  $X(q)$ .

**Problem 4.12.** [solution p. 170]

**Part one.** Let  $(E, d)$  be a separable metric space and  $(x_n)_{n \geq 1}$  a dense sequence. Let  $\mathcal{M}_1(E)$  denote the set of probability measures on  $E$  equipped with its Borel sigma-algebra, and  $\text{Lip}_1(E)$  the set of 1-Lipschitz functions from  $E$  to  $\mathbb{R}$  (that is, functions satisfying  $|f(x) - f(y)| \leq d(x, y)$  for all  $x, y \in E$ ).

Let  $\mu_n, \mu \in \mathcal{M}_1(E)$ . The aim of this part is to show that  $\mu_n \Rightarrow \mu$  if and only if  $\int f d\mu_n \rightarrow \int f d\mu$  for every function  $f$  of the form

$$f(x) = \max(c_0, c_1 - d(x, x_1), \dots, c_n - d(x, x_n))$$

with  $n \geq 1$  and  $c_0, \dots, c_n \in \mathbb{R}$ .

- (1) Justify the forward implication.
- (2) Check that, to show that  $\mu_n \Rightarrow \mu$ , it suffices to check that  $\int f d\mu_n \rightarrow \int f d\mu$  for every bounded function  $f \in \text{Lip}_1(E)$ .
- (3) Prove the converse.

*Hint.* For a bounded function  $f \in \text{Lip}_1(E)$  satisfying  $f \geq c_0$ , check that, for every  $x \in E$ , we have

$$f(x) = \sup_{n \geq 1} \max(c_0, f(x_n) - d(x, x_n)).$$

- (4) If  $E$  is bounded (that is, if  $\sup_{x, y \in E} d(x, y) < \infty$ ), show that  $\mu_n \Rightarrow \mu$  if and only if  $\int f d\mu_n \rightarrow \int f d\mu$  for every function  $f(x)$  that is a polynomial in the variables  $d(x, x_1), \dots, d(x, x_n)$ , for every  $n \geq 1$ .

**Part two.** Consider the set  $E$  of measurable functions  $[0, 1] \rightarrow \mathbb{R}$ , identified up to equality almost everywhere (that is, two functions equal almost everywhere are identified), equipped with the metric

$$d(f, g) = \int_0^1 \min(1, |f(t) - g(t)|) dt. \quad (6)$$

- (5) Writing  $\lambda$  for Lebesgue measure, show that  $f_n \rightarrow f$  in  $E$  if and only if, for every  $\varepsilon > 0$ , we have  $\lambda(\{x \in [0, 1] : |f_n(x) - f(x)| \geq \varepsilon\}) \rightarrow 0$  as  $n \rightarrow \infty$ .
- (6) Suppose that  $f_n \rightarrow f$  in  $E$ . Show that there exists a subsequence indexed by  $\phi$  such that  $f_{\phi(n)}$  converges almost everywhere to  $f$ .
- (7) Prove that  $E$  is Polish.

**Part three.** Consider the metric space  $\mathcal{C}$  of continuous functions  $[0, 1] \rightarrow \mathbb{R}$  equipped with the metric  $d$  defined by (6).

(8) Check that  $(\mathcal{C}, d)$  is separable and bounded. Is it complete? Justify your answer.

Let  $(X_n)_{n \geq 1}$  be random variables taking values in  $\mathcal{C}$  and  $X$  a random variable taking values in  $\mathcal{C}$ .

(9) Prove that if  $X_n$  converges in distribution to  $X$ , then, for every  $k \geq 1$ , if  $(U_i)_{1 \leq i \leq k}$  are i.i.d. uniform random variables on  $[0, 1]$  (also independent of  $X_n, X$ ), then  $(X_n(U_1), \dots, X_n(U_k))$  converges in distribution to  $(X(U_1), \dots, X(U_k))$ .

(10) Suppose that the finite-dimensional distributions of  $X_n$  converge weakly to those of  $X$ . Prove that  $X_n$  converges in distribution to  $X$ .

(11) [Unmarked bonus question] Prove that  $X_n$  converges in distribution to  $X$  if and only if, for every  $k \geq 1$ , writing  $V_1^k, \dots, V_k^k$  for the increasing rearrangement of  $k$  i.i.d. uniform random variables on  $[0, 1]$  (also independent of  $X_n$  and  $X$ ), we have convergence in distribution of  $(X_n(V_1^k), \dots, X_n(V_k^k))$  to  $(X(V_1^k), \dots, X(V_k^k))$ .

\* \* \*

**Problem 4.13.** [solution p. 176] Let  $(E_i)_{i \geq 1}$  be independent exponential random variables with rate 1. Set  $S_0 = 0$  and  $S_i = E_1 + \dots + E_i$  for  $i \geq 1$ . For  $t \geq 0$ , set  $P(t) = \text{Card}\{i \geq 1 : S_i \leq t\}$ , so that  $P$  is a Poisson process with rate 1.

**Part one.**

(1) Show that, for every  $T \geq 0$ , we have  $\sup_{0 \leq u \leq T} \left| \frac{P(un)}{n} - u \right| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0$ .

(2) Is it true that  $\sup_{u \geq 0} \left| \frac{P(un)}{n} - u \right| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0$ ? Justify your answer.

**Part two.** Let  $(Y_n)_{n \geq 1}$  be a sequence of random variables taking values in  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ , equipped with the topology of uniform convergence on compact sets. Suppose that there exists a random variable  $Y \in \mathcal{C}(\mathbb{R}_+, \mathbb{R})$  such that, for every  $T \geq 0$ ,  $(Y_n(t) : 0 \leq t \leq T)$  converges in distribution to  $(Y(t) : 0 \leq t \leq T)$  in  $\mathcal{C}([0, T], \mathbb{R})$  as  $n \rightarrow \infty$ .

(3) Prove that, for every  $T \geq 0$ , the following convergence holds in probability:

$$\sup_{0 \leq t \leq T} \left| Y_n \left( \frac{P(tn)}{n} \right) - Y_n(t) \right| \xrightarrow[n \rightarrow \infty]{(\mathbb{P})} 0.$$

**Part three.** Let  $T > 0$ , let  $\beta : \mathbb{R} \rightarrow \mathbb{R}_+$  be a Lipschitz function, and let  $z_0 \in \mathbb{R}_+$ .

- (4) Justify that the differential equation  $z'(t) = \beta(z(t))$  with  $z(0) = z_0$  has a unique solution on  $[0, T]$ , denoted by  $z$  below.

For  $n \geq 1$ , fix  $Z_n(0) \in \mathbb{R}$  and suppose that  $(Z_n(t))_{0 \leq t \leq T}$  satisfies, for every  $0 \leq t \leq T$ :

$$Z_n(t) = Z_n(0) + P \left( n \int_0^t \beta \left( \frac{Z_n(s)}{n} \right) ds \right),$$

Set  $\bar{Z}_n(t) = Z_n(t)/n$  and suppose that  $\bar{Z}_n(0) \rightarrow z_0$  as  $n \rightarrow \infty$ .

- (5) Show that

$$\sup_{0 \leq t \leq T} |\bar{Z}_n(t) - z(t)| \xrightarrow[n \rightarrow \infty]{a.s.} 0.$$

*Hint.* You may use Gronwall's lemma.

\* \* \*

**Exercise 4.14.** [solution p. 178] Let  $K$  be a compact metric space and  $E$  a metric space. Let  $C(K, E)$  denote the space of continuous functions from  $K$  to  $E$  equipped with the metric  $d(f, g) = \sup_{x \in K} d_E(f(x), g(x))$ . Show that  $\{B \cap C(K, E) : B \in \mathcal{B}(E)^{\otimes K}\}$  is the smallest sigma-algebra on  $C(K, E)$  making all the projections

$$\begin{aligned} \Pi_x : C(K, E) &\rightarrow E \\ f &\mapsto f(x) \end{aligned}$$

for  $x \in K$  measurable (in other words, the trace of the product sigma-algebra  $\mathcal{B}(E)^{\otimes K}$  on  $C(K, E)$  is the product sigma-algebra on  $C(K, E)$ ).

## 4.2 Solutions

**Solution to Exercise 4.1.** [statement p. 156]

- (1) Yes! Let  $\varepsilon > 0$ . By Prokhorov's theorem, there exist two compact subsets  $K_1, K_2$  of  $\mathcal{C}([0, 1], \mathbb{R})$  such that  $\mathbb{P}(X^n \in K_1) \geq 1 - \varepsilon$  and  $\mathbb{P}(Y^n \in K_2) \geq 1 - \varepsilon$  for every integer  $n \geq 0$ . Then  $K_1 \times K_2$  is a compact subset of  $\mathcal{C}([0, 1], \mathbb{R})^2$ , and  $\mathbb{P}((X^n, Y^n) \in K_1 \times K_2) \geq 1 - 2\varepsilon$ , which shows that  $((X^n, Y^n))_{n \geq 0}$  is a tight sequence in  $\mathcal{C}([0, 1], \mathbb{R})^2$ .
- (2) Yes! First, it is clear that, for every  $t \in [0, 1]$ ,  $(X_t^n, Y_t^n)_{n \geq 1}$  is tight. To control the modulus of continuity, equip  $\mathbb{R}^2$  with the norm  $\|\cdot\|_1$ . For a function  $f = (f_1, f_2) \in \mathcal{C}([0, 1], \mathbb{R}^2)$ ,

$$\omega(f, \delta) = \sup\{|f_1(s) - f_1(t)| + |f_2(s) - f_2(t)| : |s - t| \leq \delta\} \leq \omega(f_1, \delta) + \omega(f_2, \delta).$$

Thus,

$$\mathbb{P}(\omega((X^n, Y^n), \delta) \geq \eta) \leq \mathbb{P}(\omega(X^n, \delta) \geq \eta/2) + \mathbb{P}(\omega(Y^n, \delta) \geq \eta/2),$$

which implies tightness.

**Remark.** Alternatively, use the fact that the function

$$\begin{aligned} \mathcal{C}([0, 1], \mathbb{R})^2 &\rightarrow \mathcal{C}([0, 1], \mathbb{R}^2) \\ (f, g) &\mapsto (f(t), g(t))_{0 \leq t \leq 1} \end{aligned}$$

is continuous and that tightness is preserved under composition with a continuous function.

(3) Yes! It suffices to check that the function

$$\begin{aligned} \Phi : \mathcal{C}([0, 1], \mathbb{R})^2 &\rightarrow \mathcal{C}([0, 1]^2, \mathbb{R}^2) \\ (f, g) &\mapsto (f(s), g(t))_{0 \leq s \leq 1, 0 \leq t \leq 1} \end{aligned}$$

is continuous; the result then follows because tightness is preserved under composition with a continuous function. To this end, equipping  $\mathbb{R}^2$  with the norm  $\|\cdot\|_1$ , and writing  $d(F, G) = \sup_{(s,t) \in [0,1]^2} \|F(s, t) - G(s, t)\|_1$  for  $F, G \in \mathcal{C}([0, 1]^2, \mathbb{R}^2)$ , observe that

$$d(\Phi(f, g), \Phi(f', g')) = \|f - f'\|_\infty + \|g - g'\|_\infty.$$

It follows that if  $f_n \rightarrow f$  and  $g_n \rightarrow g$ , then  $\Phi(f_n, g_n) \rightarrow \Phi(f, g)$ , proving the result.

\* \* \*

**Solution to Exercise 4.2.** [statement p. 156] By the Skorokhod representation theorem, we may assume that  $X^n \rightarrow X$  holds almost surely. By continuity of the map

$$\begin{aligned} \mathcal{C}([0, 1], \mathbb{R}^k) &\rightarrow \mathbb{R} \\ g &\mapsto \int_0^1 f(g(s)) ds, \end{aligned}$$

we deduce that  $Y_n \rightarrow Y$  almost surely. Thus  $(X^n, Y_n) \rightarrow (X, Y)$  almost surely, and hence in distribution.

**Remark.** We can also directly observe that the map

$$\begin{aligned} \Phi : \mathcal{C}([0, 1], \mathbb{R}^k) &\rightarrow \mathcal{C}([0, 1], \mathbb{R}^k) \times \mathbb{R} \\ g &\mapsto (g, \int_0^1 f(g(s)) ds) \end{aligned}$$

is continuous, and hence that  $\Phi(X^n)$  converges in distribution to  $\Phi(X)$ .

\* \* \*

**Solution to Exercise 4.3.** [statement p. 156]  $\Rightarrow$  This implication always holds.

$\Leftarrow$  Since  $(X^n(0))_{n \geq 1}$  converges in distribution, it suffices to check that the sequence  $(X^n)_{n \geq 1}$  is tight. The partition used in the proof of Corollary 4.19, with  $\delta = 1/k$ , gives, for every function  $f \in \mathcal{C}([0, 1], \mathbb{R})$ ,

$$\omega(f, 1/k) \leq 3 \max_{0 \leq i \leq k-1} \sup_{r \in [0, 1/k]} |f(r + i/k) - f(i/k)|.$$

In particular, if  $f$  is nondecreasing, we have

$$\omega(f, 1/k) \leq 3 \max_{0 \leq i \leq k-1} |f((i+1)/k) - f(i/k)|.$$

Moreover, by convergence of the finite-dimensional distributions of  $X^n$  to those of  $X$ , we have

$$\max_{0 \leq i \leq k-1} |X^n((i+1)/k) - X^n(i/k)| \xrightarrow[n \rightarrow \infty]{(d)} \max_{0 \leq i \leq k-1} |X((i+1)/k) - X(i/k)|.$$

By continuity of  $X$ , for every  $\eta > 0$ , we have

$$\mathbb{P} \left( \max_{0 \leq i \leq k-1} |X((i+1)/k) - X(i/k)| \geq \eta \right) \xrightarrow[k \rightarrow \infty]{} 0.$$

We deduce that, for every  $\varepsilon, \eta > 0$ , there exists  $k \geq 1$  such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\omega(X^n, 1/k) \geq \eta) \leq \varepsilon.$$

\* \* \*

#### Solution to Exercise 4.4. [statement p. 156]

- (1) No, it is not continuous. In the example below,  $f_n \rightarrow f$  uniformly on compact sets, but  $3 = t_a(f_n) \not\rightarrow t_a(f) = 1$ . Equipping  $[0, \infty]$  with its usual topology, the points of continuity are exactly the functions  $f$  such that  $t_a(f) = \infty$ , or  $f(0) > a$ , or else  $t_a(f) < \infty$  and, for every  $\varepsilon > 0$ , there exists  $t \in [t_a(f), t_a(f) + \varepsilon]$  such that  $f(t) > a$ . When  $t_a(f) > 0$ , this last condition is equivalent to the absence of a local maximum at  $t_a(f)$ .

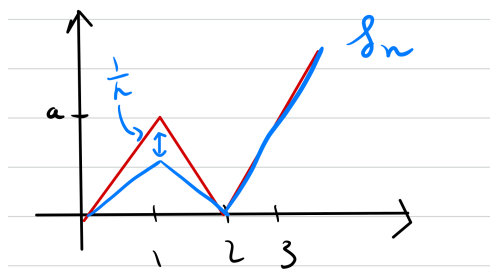


**Figure 1:** Discontinuity of the first hitting time:  $f_n$  converges to  $f$ , but  $t_a(f_n) = 3 \neq t_a(f) = 1$ .

- (2) First, it is clear that  $0 = T_0^n$  converges to  $T_0 = 0$ . We therefore assume that  $a > 0$ . By the previous question, if almost surely  $X$  has no local maximum at  $T_a$  when  $T_a < \infty$ , then  $T_a^n$  converges in distribution to  $T_a$  (use, for instance, the Skorokhod representation theorem). When  $X$  is standard Brownian motion, we know that  $T_a < \infty$  almost surely, and since, for every  $\varepsilon > 0$ ,  $\mathbb{P} \left( \sup_{[0, \varepsilon]} X > 0 \right) = 1$ , the

strong Markov property applied at the stopping time  $T_a$  shows that almost surely  $X$  has no local maximum at  $T_a$ . Thus, in this case,  $T_a^n$  converges in distribution to  $T_a$ .

Now, if with strictly positive probability  $T_a < \infty$  and  $X$  has a local maximum at  $T_a$ , then we can find  $X^n$  such that  $T_a^n$  does not converge in distribution to  $T_a$ . Since  $a > 0$  and  $X_0 = 0$ , we have  $T_a > 0$  and

$$\begin{aligned} & \{T_a < \infty \text{ and } X \text{ has a local maximum at } T_a\} \\ &= \bigcup_{k \geq 1} \{T_a < k, X_t \leq a \text{ for every } t \in [T_a, T_a + 1/k]\}. \end{aligned}$$

Choose  $k$  such that the event  $A$  appearing in this union has strictly positive probability, and set

$$X_t^n = X_t - \frac{\min(t, 1)}{n} \mathbb{1}_A.$$

Then  $X_0^n = 0$  and  $X^n \rightarrow X$  uniformly on compact sets. On  $A$ ,  $T_a^n \geq T_a + 1/k$ , and outside  $A$ ,  $T_a^n = T_a$ . For  $M > 2k$ , we therefore obtain

$$\mathbb{E}[T_a^n \wedge M] - \mathbb{E}[T_a \wedge M] \geq \mathbb{P}(A)/k > 0.$$

Since the function  $u \mapsto u \wedge M$  is bounded and continuous on  $[0, \infty]$ , this rules out convergence in distribution of  $T_a^n$  to  $T_a$ .

\* \* \*

**Solution to Exercise 4.5.** [statement p. 157] We argue by contradiction, assuming that  $A = \{f \in \mathbb{R}^{[0,1]} : f \text{ is continuous}\} \in \mathcal{B}(\mathbb{R})^{\otimes [0,1]}$ . The random variable  $X = (X(t) : 0 \leq t \leq 1)$  has the same finite-dimensional distributions as the constant function equal to 1. It therefore has the same distribution as the constant function equal to 1. Thus

$$\mathbb{P}(X \in A) = \mathbb{P}(1 \in A) = 1.$$

This is a contradiction, since  $X$  is not continuous.

\* \* \*

**Solution to Exercise 4.6.** [statement p. 157] First, observe that uniform integrability implies that, for every  $t \in [0, 1]$ ,  $X_t$  is integrable and  $\mathbb{E}[X_t] = \mathbb{E}[X_0]$ .

Fix  $0 \leq s \leq t \leq 1$ . For every integer  $m \geq 1$ , all  $0 \leq t_1 \leq \dots \leq t_m \leq s$  and every bounded continuous function  $h : \mathbb{R}^m \rightarrow \mathbb{R}$ , the martingale property gives

$$\mathbb{E}[X_s^n h(X_{t_1}^n, \dots, X_{t_m}^n)] = \mathbb{E}[X_t^n h(X_{t_1}^n, \dots, X_{t_m}^n)],$$

because  $h(X_{t_1}^n, \dots, X_{t_m}^n)$  is  $\mathcal{F}_s^n$ -measurable. By the Skorokhod representation theorem, we may assume that  $X^n \rightarrow X$  almost surely in  $\mathcal{C}$ . Both products then converge almost surely to the corresponding products for  $X$ . They are uniformly integrable, since  $h$  is bounded and the sequences  $(X_s^n)$  and  $(X_t^n)$  are uniformly integrable. Thus,

$$\mathbb{E}[X_s h(X_{t_1}, \dots, X_{t_m})] = \mathbb{E}[X_t h(X_{t_1}, \dots, X_{t_m})].$$

The functional monotone class theorem extends this identity to every bounded  $\mathcal{F}_s$ -measurable random variable, since bounded continuous cylinder functions generate  $\mathcal{F}_s$ . In particular, for every  $A \in \mathcal{F}_s$ ,

$$\mathbb{E}[X_s \mathbb{1}_A] = \mathbb{E}[X_t \mathbb{1}_A].$$

Since  $X$  is adapted and integrable, it is a martingale with respect to its natural filtration.

\* \* \*

**Solution to Exercise 4.7.** [statement p. 157]

- (1) By the lecture notes, the sequence  $(X_n)_{n \geq 1}$  is tight in  $\mathcal{C}$  if and only if the sequence  $(X_n(0))_{n \geq 1}$  is tight in  $\mathbb{R}$  and

$$\forall \varepsilon > 0, \forall \eta > 0, \exists \delta > 0, \quad \limsup_{n \rightarrow \infty} \mathbb{P}(\omega(X_n, \delta) \geq \eta) \leq \varepsilon. \quad (7)$$

We therefore need to prove that (7) is equivalent to (b). We prove that their negations are equivalent:

- The negation of (7) is: there exist  $\varepsilon, \eta > 0$  such that, for every  $\delta > 0$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\omega(X_n, \delta) \geq \eta) > \varepsilon.$$

- The negation of (b) is: there exist a sequence  $\delta_k \rightarrow 0$  and a sequence  $n_k \rightarrow \infty$  such that  $\omega(X_{n_k}, \delta_k)$  does not converge in probability to 0.

Assuming the negation of (7), there exist  $\varepsilon, \eta > 0$  and a sequence  $(n_k)$  such that, for every  $k \geq 1$ :

$$\mathbb{P}(\omega(X_{n_k}, 1/k) \geq \eta) > \varepsilon.$$

Thus, taking  $\delta_k = 1/k$ , we see that  $\omega(X_{n_k}, \delta_k)$  does not converge in probability to 0.

Assuming the negation of (b), there exist a sequence  $\delta_k \rightarrow 0$ ,  $\eta, \varepsilon > 0$  and a sequence  $(n'_k)$  tending to infinity (after passing to a subsequence; for a finite set of indices, the moduli converge in probability to zero) such that, for every  $k \geq 1$ ,

$$\mathbb{P}(\omega(X_{n'_k}, \delta_k) \geq \eta) > \varepsilon.$$

Thus, for every  $\delta > 0$  and every sufficiently large  $k$  such that  $\delta_k < \delta$ , we have

$$\mathbb{P}(\omega(X_{n'_k}, \delta) \geq \eta) > \varepsilon$$

We therefore obtain a limit superior at least equal to  $\varepsilon$ , hence strictly greater than  $\varepsilon/2$ , proving the negation of (7).

- (2) We use the result of question (1). We first prove that  $(Y_n(0))_{n \geq 1}$  is tight by showing that it is bounded in  $L^1$ . This follows because, by linearity,

$$\mathbb{E} [\|Y_n(0)\|] \leq \mathbb{E} [\|X_n(0)\|] \leq \mathbb{E} \left[ \sup_{t \in [0,1]} |X_n(t)| \right],$$

and, since the sequence of random variables  $\sup_{t \in [0,1]} |X_n(t)|$  is uniformly integrable, it is bounded in  $L^1$ .

For tightness, let  $\delta_k \rightarrow 0$  and  $(n_k)_{k \geq 1}$  be sequences. Let us check that  $\omega(Y_{n_k}, \delta_k)$  converges in probability to 0. First, since the sequence  $(X_n)_{n \geq 1}$  is tight in  $\mathcal{C}$ ,  $\omega(X_{n_k}, \delta_k)$  converges in probability to 0. By assumption (II), this sequence is also uniformly integrable, and hence  $\mathbb{E} [\omega(X_{n_k}, \delta_k)] \rightarrow 0$ . Thus:

$$\omega(Y_{n_k}, \delta_k) \leq \frac{1}{n_k} \sum_{i=1}^{n_k} \omega(X_{n_k}^i, \delta_k),$$

so that

$$\mathbb{E} [\omega(Y_{n_k}, \delta_k)] \leq \mathbb{E} [\omega(X_{n_k}, \delta_k)] \xrightarrow[k \rightarrow \infty]{} 0.$$

It follows that  $\omega(Y_{n_k}, \delta_k)$  converges in  $L^1$  to 0, and hence in probability.

- (3) No. For example, take a sequence  $(B_n)_{n \geq 1}$  of Bernoulli random variables with respective parameters  $1/n$ , and take  $(X_n^i)_{1 \leq i \leq n}$  i.i.d. with the same distribution as  $X_n$  defined by

$$X_n(t) = \begin{cases} n^2 B_n t & \text{if } t \leq 1/n \\ n B_n & \text{if } t \geq 1/n. \end{cases}$$

We have  $\mathbb{P} (\|X_n\|_\infty > \varepsilon) \leq 1/n$  for every  $\varepsilon > 0$ . Thus  $X_n \rightarrow 0$  in probability in  $\mathcal{C}$ , and the sequence  $(X_n)$  is tight: condition (I) holds. Moreover,  $\mathbb{E} [\|X_n\|_\infty] = 1$ , which gives (II'). However, the probability that exactly one of the  $X_n^i$  is nonzero is  $(1 - 1/n)^{n-1} \rightarrow e^{-1}$ ; on this event,  $\omega(Y_n, 1/n) = 1$ . Condition (1b) therefore fails, and  $(Y_n)$  is not tight in  $\mathcal{C}$ .

**Remark.** This exercise is inspired by Appendix A of the paper Berzunza Ojeda G, Janson S. The distance profile of rooted and unrooted simply generated trees. *Combinatorics, Probability and Computing*. 2022;31(3):368-410. doi:10.1017/S0963548321000304.

\* \* \*

#### Solution to Exercise 4.8. [statement p. 158]

- (1) We find that  $\mathbb{E} [(Z_n(s) - Z_n(t))^2] = \mathbb{E} [(X_1(s) - X_1(t))^2]$ .
- (2) We first check tightness using Kolmogorov's tightness criterion. On the one hand, for every  $s \in [0, 1]$ ,

$$\mathbb{E} [Z_n(s)^2] = \mathbb{E} [X_1(s)^2] < \infty,$$

so the sequence  $(Z_n(s))_{n \geq 1}$  is bounded in  $L^2$ , and hence tight. On the other hand, by the previous question, for all  $s, t \in [0, 1]$ , we indeed have

$$\mathbb{E} \left[ (Z_n(s) - Z_n(t))^2 \right] \leq C|s - t|^{1+\varepsilon}.$$

To identify the limit, let us check convergence of the finite-dimensional distributions of  $Z_n$ . For  $0 \leq t_1 < \dots < t_k \leq 1$ , we can write

$$(Z_n(t_1), \dots, Z_n(t_k)) = \frac{Y_1 + \dots + Y_n}{\sqrt{n}},$$

where  $Y_1 = (X_1(t_1), \dots, X_1(t_k))$  and the  $(Y_i)_{1 \leq i \leq n}$  are i.i.d. random variables taking values in  $\mathbb{R}^k$ . By the multivariate central limit theorem,  $(Z_n(t_1), \dots, Z_n(t_k))$  converges in distribution, as  $n \rightarrow \infty$ , to a centered Gaussian vector in  $\mathbb{R}^k$  with covariance matrix  $(\mathbb{E} [X_1(t_i)X_1(t_j)])_{1 \leq i, j \leq k}$ .

This completes the proof.

\* \* \*

**Solution to Exercise 4.9.** [statement p. 158]

- (1) Set  $M = \sup_n |x_n| < \infty$  and, for  $\varepsilon > 0$ , choose a compact set  $K \subset \mathcal{C}([0, 1], \mathbb{R})$  such that  $\mathbb{P}_0(K) \geq 1 - \varepsilon$ . The map

$$[-M, M] \times K \longrightarrow \mathcal{C}([0, 1], \mathbb{R}), \quad (x, f) \longmapsto x + f$$

is continuous. Its image  $K'$  is therefore compact, and  $\mathbb{P}_{x_n}(K') \geq \mathbb{P}_0(K) \geq 1 - \varepsilon$  for every  $n$ . The family  $(\mathbb{P}_{x_n})$  is tight.

- (2) No. Arguing by contradiction, if  $(\mathbb{P}_{x_n})_{n \geq 1}$  were tight, taking the continuous image under evaluation at time 0 would imply that the sequence of measures  $(\delta_{x_n})_{n \geq 1}$  is tight in  $\mathbb{R}$ , which fails upon taking an unbounded subsequence.

\* \* \*

**Solution to Exercise 4.10.** [statement p. 158] Let  $F : \mathcal{C}([0, 1], \mathbb{R}) \rightarrow \mathbb{R}$  be a bounded Lipschitz function, with Lipschitz constant  $L$  for the uniform norm. For  $\varepsilon > 0$ , use the pair  $(X_\varepsilon, Y_\varepsilon)$  given in the statement. Since its marginals have the laws of  $X$  and  $Y$ , we have

$$\begin{aligned} |\mathbb{E} [F(X)] - \mathbb{E} [F(Y)]| &\leq \mathbb{E} [|F(X_\varepsilon) - F(Y_\varepsilon)|] \\ &\leq L\varepsilon + 2\|F\|_\infty \mathbb{P} (\|X_\varepsilon - Y_\varepsilon\|_\infty > \varepsilon) \\ &\leq (L + 2\|F\|_\infty)\varepsilon. \end{aligned}$$

The second inequality follows by splitting over the event  $\{\|X_\varepsilon - Y_\varepsilon\|_\infty \leq \varepsilon\}$  and its complement. Letting  $\varepsilon$  tend to zero gives  $\mathbb{E} [F(X)] = \mathbb{E} [F(Y)]$  for every bounded Lipschitz function  $F$ . Such functions determine probability measures by the portmanteau theorem, so  $X$  and  $Y$  have the same distribution.

**Solution to Problem 4.11.** [statement p. 159]

- (1) We saw in the lecture notes that the Borel sigma-algebra of  $\mathcal{C}$  equals its product sigma-algebra, i.e. the smallest sigma-algebra making the projections  $\pi_t : \mathcal{C} \rightarrow \mathbb{R}$  measurable for  $t \in [0, 1]$ . Let us check that it equals  $\mathcal{F} = \{B \cap \mathcal{C} : B \in \mathcal{B}(\mathbb{R})^{\otimes [0,1]}\}$ .

First,  $\mathcal{F}$  is indeed a sigma-algebra. Next, for every  $t \in [0, 1]$  and  $A \in \mathcal{B}(\mathbb{R})$ ,  $\{f \in \mathcal{C} : f(t) \in A\} = \{f \in \mathbb{R}^{[0,1]} : f(t) \in A\} \cap \mathcal{C}$ , so the sigma-algebra  $\mathcal{F}$  makes all the projections measurable.

Finally, let  $\mathcal{G}$  be a sigma-algebra on  $\mathcal{C}$  making all the projections measurable, and consider

$$\{B \in \mathcal{B}(\mathbb{R})^{\otimes [0,1]} : B \cap \mathcal{C} \in \mathcal{G}\}.$$

This is a sigma-algebra on  $\mathbb{R}^{[0,1]}$  containing the cylinder sets of  $\mathbb{R}^{[0,1]}$ , so it equals  $\mathcal{B}(\mathbb{R})^{\otimes [0,1]}$ , which completes the proof.

- (2) (a) If  $Y$  is a modification of  $X$ , then, for  $0 \leq t_1 < t_2 < \dots < t_n \leq 1$ , we have

$$(X(t_1), \dots, X(t_n)) = (Y(t_1), \dots, Y(t_n)) \quad \text{almost surely,}$$

so these two random vectors indeed have the same distribution.

- (b) No, he is wrong: for example, take  $\varepsilon$  to be a Bernoulli random variable with parameter  $1/2$  and set  $X(t) = \varepsilon$  and  $Y(t) = 1 - \varepsilon$  for  $0 \leq t \leq 1$ .

- (3) (a) Writing  $I : \mathcal{C} \rightarrow \mathbb{R}^{[0,1]}$  for the inclusion map (measurable by (1)), observe that  $Y^* = I \circ Y$  (in particular,  $Y^*$  is measurable as a composition of measurable maps). Let us check that  $Y : (\Omega, \sigma(Y^*)) \rightarrow \mathcal{C}$  is measurable. Let  $A$  be a measurable subset of  $\mathcal{C}$ . By (1), it can be written as  $A = B \cap \mathcal{C} = I^{-1}(B)$  for  $B \in \mathcal{B}(\mathbb{R})^{\otimes [0,1]}$ , so that

$$Y^{-1}(A) = (Y^*)^{-1}(B)$$

indeed belongs to  $\sigma(Y^*)$ .

By the Doob–Dynkin lemma, since  $\mathcal{C}$  is Polish, there exists a measurable map  $F : \mathbb{R}^{[0,1]} \rightarrow \mathcal{C}$  such that  $Y = F \circ Y^*$ .

- (b) Set  $X' = F \circ X$ , which is indeed a random variable taking values in  $\mathcal{C}$ . Let us check that, for every  $t \in [0, 1]$ ,  $\mathbb{P}(X(t) = X'(t)) = 1$ . Observe that, since  $X$  and  $Y$  have the same distribution,  $X$  and  $Y^*$  have the same distribution. Hence  $(X, X') = (X, F(X))$  and  $(Y^*, Y) = (Y^*, F(Y^*))$  have the same distribution. Thus, for every  $t \in [0, 1]$ ,

$$\mathbb{P}(X(t) = X'(t)) = \mathbb{P}(Y^*(t) = Y(t)) = 1.$$

- (4) (a) Since  $E$  is complete, a sequence of elements of  $E$  converges if and only if it is Cauchy. Thus,

$$\{(u_n)_{n \geq 1} \text{ converges}\} = \bigcap_{n \geq 1} \bigcup_{N \geq 1} \bigcap_{p, q \geq N} \left\{ d(u_p, u_q) \leq \frac{1}{n} \right\},$$

and the desired results follow.

- (b) For  $q \in [0, 1] \cap \mathbb{Q}$ , let  $Z(q)$  denote the almost sure limit of  $(X_n(q))$ , defined arbitrarily on the null event where it does not exist. The finite-dimensional distributions of  $Z$  at rational times coincide with those of  $Y$ . For each  $t \in [0, 1] \setminus \mathbb{Q}$ , fix a rational sequence  $(q_j(t))$  converging to  $t$ . By (4a) and continuity of  $Y$ ,  $(Z(q_j(t)))_j$  converges almost surely; define  $Z(t)$  to be this limit, and zero if the limit does not exist. Each  $Z(t)$  is measurable. For every finite collection of times, simultaneous almost sure convergence of these rational approximations shows that the corresponding joint distributions of  $Z$  and  $Y$  coincide. Thus  $Z$  is a random variable in  $\mathbb{R}^{[0,1]}$  with the same distribution as  $Y$ .

By (3b), there exists a random variable  $X : \Omega \rightarrow \mathcal{C}$  that is a modification of  $Z$ . For every rational  $q$ ,  $X_n(q) \rightarrow Z(q) = X(q)$  almost surely, which completes the proof.

**Remark.** The result remains true if  $Y$  is not defined on the same probability space as the sequence  $(X_n)_{n \geq 1}$ : we can still extend  $Z$  to  $[0, 1]$  as a random variable in  $\mathbb{R}^{[0,1]}$  with the same finite-dimensional distributions as  $Y$ , using question (4a).

Indeed, for every  $t \in [0, 1] \setminus \mathbb{Q}$ , consider a sequence of rational numbers  $q_n \rightarrow t$ . By continuity of  $Y$ ,  $Y(q_n)$  converges a.s. to  $Y(t)$ . Since  $Z$  and  $Y$  have the same finite-dimensional distributions, the two sequences  $(Y(q_n))_{n \geq 1}$  and  $(Z(q_n))_{n \geq 1}$  have the same distribution. By (4a), there therefore exists a random variable denoted by  $Z(t)$ , the almost sure limit of  $(Z(q_n))_{n \geq 1}$ . It follows that  $Z$  has the same finite-dimensional distributions as  $Y$ .

**Remark.** This exercise is inspired by Theorem 3.24 in O. Kallenberg, Foundations of Modern Probability (2nd Edition), Springer.

\* \* \*

### Solution to Problem 4.12. [statement p. 160]

#### Part one.

- (1) This follows simply from the fact that a function of the form

$$f(x) = \max(c_0, c_1 - d(x, x_1), \dots, c_n - d(x, x_n))$$

is bounded and continuous.

- (2) By linearity of the integral, we then have  $\int f d\mu_n \rightarrow \int f d\mu$  for every bounded Lipschitz function  $f$ , which implies that  $\mu_n \Rightarrow \mu$  by the portmanteau theorem.
- (3) We start by checking the hint. Let  $f \in \text{Lip}_1(E)$  be bounded and satisfy  $f \geq c_0$ .

If  $n \geq 1$ , we have  $f(x) \geq c_0$  and  $f(x) \geq f(x_n) - d(x, x_n)$  because  $f$  is 1-Lipschitz. Thus  $f(x) \geq \sup_{n \geq 1} \max(c_0, f(x_n) - d(x, x_n))$ .

Next, fix  $\varepsilon > 0$  and check that we can find  $n \geq 1$  such that  $f(x) \leq \max(c_0, f(x_n) - d(x, x_n)) + \varepsilon$ . By density, let  $x_n$  satisfy  $d(x, x_n) \leq \varepsilon/2$ . Then

$$\begin{aligned} f(x) - \varepsilon &= f(x_n) + f(x) - f(x_n) - \varepsilon \\ &\leq f(x_n) + d(x, x_n) - 2d(x, x_n) \\ &= f(x_n) - d(x, x_n), \end{aligned}$$

which proves the result.

To prove the converse, let  $f \in \text{Lip}_1(E)$  be bounded and let  $c_0$  satisfy  $f \geq c_0$ . For  $i \geq 1$ , set

$$f_i(x) = \max(c_0, f(x_1) - d(x, x_1), \dots, f(x_i) - d(x, x_i)).$$

The sequence  $(f_i)$  increases pointwise to  $f$ . Each  $f_i$  is exactly of the form of the test functions in the assumption; hence, for fixed  $i$ , the assumption gives

$$\int f_i d\mu_n \rightarrow \int f_i d\mu.$$

Since  $\int f_i d\mu_n \leq \int f d\mu_n$ , we deduce that

$$\int f_i d\mu \leq \liminf_{n \rightarrow \infty} \int f d\mu_n.$$

By monotone convergence, letting  $i \rightarrow \infty$  gives

$$\int f d\mu \leq \liminf_{n \rightarrow \infty} \int f d\mu_n.$$

Replacing  $f$  by  $-f$  yields the desired result.

- (4) As in (1), the forward implication is immediate. For the converse, let  $A > 0$  be such that  $d(x, y) \leq A$  for every  $x, y \in E$ . Observe that the map

$$\begin{aligned} \Phi : [0, A]^n &\longrightarrow \mathbb{R} \\ (t_1, \dots, t_n) &\mapsto \max(c_0, c_1 - t_1, \dots, c_n - t_n) \end{aligned}$$

can be approximated uniformly by polynomials in  $t_1, \dots, t_n$  by the Stone–Weierstrass theorem, which completes the proof.

### Part two.

- (5) For the forward implication, for  $\varepsilon \in ]0, 1[$ , Markov's inequality gives:

$$\begin{aligned} \lambda\{x \in [0, 1] : |f_n(x) - f(x)| \geq \varepsilon\} \\ &= \lambda\{x \in [0, 1] : \min(1, |f_n(x) - f(x)|) \geq \varepsilon\} \\ &\leq \frac{1}{\varepsilon} \int_0^1 \min(1, |f_n(t) - f(t)|) dt, \end{aligned}$$

which tends to 0.

For the converse, suppose that  $\lambda(\{x \in [0, 1] : |f_n(x) - f(x)| \geq \varepsilon\}) \rightarrow 0$  for every  $\varepsilon > 0$ . Let  $\varepsilon \in ]0, 1[$  and write  $A = \{x \in [0, 1] : |f_n(x) - f(x)| \geq \varepsilon\}$ . Then

$$\int_0^1 \min(1, |f_n(t) - f(t)|) dt \leq \int_A 1 dt + \int_{A^c} \varepsilon dt \leq \lambda(A) + \varepsilon,$$

so that  $d(f_n, f) \leq 2\varepsilon$  for sufficiently large  $n$ .

- (6) Suppose that  $f_n \rightarrow f$  in  $E$ . Choose a subsequence indexed by  $\phi$  such that, for every  $n \geq 1$ ,

$$\lambda\left(\left\{x \in [0, 1] : |f_{\phi(n)}(x) - f(x)| \geq \frac{1}{2^n}\right\}\right) \leq \frac{1}{2^n}.$$

By the Borel–Cantelli lemma,  $f_{\phi(n)}$  converges almost everywhere to  $f$ .

- (7) The space  $L^1([0, 1], dt)$  is separable, and  $d(f, g) \leq \|f - g\|_1$ . It is dense in  $E$ : the truncations  $f^{(m)} = (-m) \vee (f \wedge m)$  belong to  $L^1$  and converge to  $f$  for  $d$  by dominated convergence. A countable dense subset of  $L^1$  is therefore dense in  $E$ .

For completeness, let  $(f_n)$  be a Cauchy sequence in  $E$ . Choose a subsequence indexed by  $\phi$  such that  $d(f_{\phi(j+1)}, f_{\phi(j)}) \leq 2^{-2j}$ . Then

$$\lambda\{|f_{\phi(j+1)} - f_{\phi(j)}| > 2^{-j}\} \leq 2^{-j}.$$

By Borel–Cantelli, the successive differences are eventually bounded by  $2^{-j}$  almost everywhere. The subsequence is therefore Cauchy almost everywhere and converges to a measurable real-valued function  $f$ , defined arbitrarily on the exceptional set. By dominated convergence,  $d(f_{\phi(j)}, f) \rightarrow 0$ . The Cauchy property then implies  $d(f_n, f) \rightarrow 0$ .

**Part three.** For  $t \in [0, 1]$  and  $\varepsilon > 0$ , set  $I_{t,\varepsilon} = [0, 1] \cap [t - \varepsilon, t + \varepsilon]$  and

$$A_{t,\varepsilon}(f) = \frac{1}{\lambda(I_{t,\varepsilon})} \int_{I_{t,\varepsilon}} \arctan(f(u)) du.$$

Since  $|\arctan x - \arctan y| \leq \pi \min(1, |x - y|)$ , the map  $A_{t,\varepsilon}$  is continuous for  $d$ . For every  $f \in \mathcal{C}$ , continuity of  $f$  gives  $A_{t,1/m}(f) \rightarrow \arctan(f(t))$ . Thus the evaluation map  $\Pi_t(f) = f(t) = \tan(\lim_m A_{t,1/m}(f))$  is measurable. The map  $(f, t) \mapsto f(t)$  is then jointly measurable, since the paths are continuous. This also justifies evaluations at random times and the applications of Fubini below.

- (8) Separability follows, for example, from separability of the metric space of continuous functions from  $[0, 1]$  to  $\mathbb{R}$  equipped with the uniform norm, together with the fact that  $d(f, g) \leq \|f - g\|_\infty$  for every  $f, g \in \mathcal{C}$ .

Next, clearly  $d(f, g) \leq 1$  for all  $f, g \in \mathcal{C}$ .

Finally,  $(\mathcal{C}, d)$  is not complete. Indeed, if we set

$$f_n(x) = \begin{cases} 0 & \text{if } x \leq \frac{1}{2}, \\ n(x - \frac{1}{2}) & \text{if } \frac{1}{2} < x < \frac{1}{2} + \frac{1}{n}, \\ 1 & \text{if } x \geq \frac{1}{2} + \frac{1}{n}. \end{cases}$$

the sequence  $f_n$  is Cauchy for  $\|\cdot\|_1$ , hence for  $d$ , but does not converge. Indeed, if it converged to a continuous function  $f \in \mathcal{C}$ , it would also converge to  $f$  in  $E$ . But  $f_n \rightarrow \mathbb{1}_{[1/2,1]}$  for  $\|\cdot\|_1$ , hence in  $E$ . Thus  $f = \mathbb{1}_{[1/2,1]}$  almost everywhere, which is a contradiction.

(9) Let  $F : \mathbb{R}^k \rightarrow \mathbb{R}$  be bounded and continuous. It suffices to show that

$$\mathbb{E}[F(X_n(u_1), \dots, X_n(u_k))] \xrightarrow{n \rightarrow \infty} \mathbb{E}[F(X(u_1), \dots, X(u_k))].$$

To this end, write

$$\begin{aligned} \mathbb{E}[F(X_n(u_1), \dots, X_n(u_k))] \\ &= \mathbb{E} \left[ \int_{[0,1]^k} F(X_n(t_1), \dots, X_n(t_k)) dt_1 \cdots dt_k \right] \\ &= \mathbb{E}[\Phi(X_n)], \end{aligned}$$

where

$$\begin{aligned} \Phi : \mathcal{C} &\longrightarrow \mathbb{R} \\ f &\mapsto \int_{[0,1]^k} F(f(t_1), \dots, f(t_k)) dt_1 \cdots dt_k. \end{aligned}$$

It then suffices to observe that  $\Phi$  is bounded and continuous. It is bounded because  $F$  is bounded. If  $f_n \rightarrow f$  for  $d$ , every subsequence has, by (6), a further subsequence converging almost everywhere to  $f$ . By dominated convergence on  $[0,1]^k$ , the images of this further subsequence under  $\Phi$  converge to  $\Phi(f)$ . The subsequence criterion for convergence therefore gives  $\Phi(f_n) \rightarrow \Phi(f)$ .

(10) Let  $(f_n)_{n \geq 1}$  be a dense sequence in  $\mathcal{C}$ . By question (4), it suffices to check that, for every  $k \geq 1$  and every polynomial  $P$  in  $k$  variables, we have

$$\mathbb{E}[P(d(X_n, f_1), \dots, d(X_n, f_k))] \xrightarrow{n \rightarrow \infty} \mathbb{E}[P(d(X, f_1), \dots, d(X, f_k))]. \quad (8)$$

Now

$$\begin{aligned} \mathbb{E}[P(d(X_n, f_1), \dots, d(X_n, f_k))] \\ &= \mathbb{E} \left[ P \left( \int_0^1 \min(1, |X_n(t) - f_1(t)|) dt, \dots, \int_0^1 \min(1, |X_n(t) - f_k(t)|) dt \right) \right] \end{aligned}$$

is a linear combination of terms of the form

$$\int_{[0,1]^N} \mathbb{E} \left[ \prod_{i=1}^N \min(1, |X_n(t_i) - g_i(t_i)|) \right] dt_1 \cdots dt_N$$

with  $N \geq 1$  and  $g_i \in \{f_1, \dots, f_k\}$  for every  $1 \leq i \leq N$ . For fixed  $t_1, \dots, t_N \in [0,1]$ , we have

$$\mathbb{E} \left[ \prod_{i=1}^N \min(1, |X_n(t_i) - g_i(t_i)|) \right] = \mathbb{E}[\Psi(X_n(t_1), \dots, X_n(t_N))]$$

with

$$\begin{aligned} \Psi : \quad \mathbb{R}^N &\longrightarrow \mathbb{R} \\ (x_1, \dots, x_N) &\mapsto \prod_{i=1}^N \min(1, |x_i - g_i(t_i)|) \end{aligned}$$

bounded and continuous. Thus, by dominated convergence,

$$\begin{aligned} \int_{[0,1]^N} \mathbb{E} \left[ \prod_{i=1}^N \min(1, |X_n(t_i) - g_i(t_i)|) \right] dt_1 \cdots dt_N \\ \xrightarrow{n \rightarrow \infty} \int_{[0,1]^N} \mathbb{E} \left[ \prod_{i=1}^N \min(1, |X(t_i) - g_i(t_i)|) \right] dt_1 \cdots dt_N. \end{aligned}$$

This proves (8) and completes the proof.

- (11) First suppose that  $X_n$  converges in distribution to  $X$ . For  $k \geq 1$  and a bounded continuous function  $F : \mathbb{R}^k \rightarrow \mathbb{R}$ , set

$$\Phi(f) = k! \int_{0 < t_1 < \dots < t_k < 1} F(f(t_1), \dots, f(t_k)) dt_1 \cdots dt_k.$$

As in (9), question (6) and dominated convergence show that  $\Phi$  is bounded and continuous for  $d$ . Since  $(V_1^k, \dots, V_k^k)$  has density  $k!$  on this simplex, independence gives

$$\begin{aligned} \mathbb{E} [F(X_n(V_1^k), \dots, X_n(V_k^k))] &= \mathbb{E} [\Phi(X_n)] \\ &\xrightarrow{n \rightarrow \infty} \mathbb{E} [\Phi(X)] = \mathbb{E} [F(X(V_1^k), \dots, X(V_k^k))]. \end{aligned}$$

This proves the forward implication.

For the converse, we may take the  $V_1^k, \dots, V_k^k$  to be the rearrangement of the first  $k$  terms of a single sequence  $(U_i)_{i \geq 1}$  of i.i.d. uniform random variables on  $[0, 1]$ , independent of  $(X_n)_{n \geq 1}$  and  $X$ . Write  $\rho(x, y) = \min(1, |x - y|)$ . For fixed  $g \in \mathcal{C}$ , we will approximate  $d(f, g)$  by

$$D_{k,g}(f) = \frac{1}{k} \sum_{i=1}^k \rho(f(V_i^k), g(i/k)).$$

Compare this quantity with

$$A_{k,g}(f) = \frac{1}{k} \sum_{i=1}^k \rho(f(V_i^k), g(V_i^k)) = \frac{1}{k} \sum_{i=1}^k \rho(f(U_i), g(U_i)).$$

For deterministic  $f$ , the  $k$  terms in this average are i.i.d., take values in  $[0, 1]$ , have mean  $d(f, g)$  and variance at most  $1/4$ . Thus

$$\mathbb{E} [|A_{k,g}(f) - d(f, g)|] \leq \text{Var}(A_{k,g}(f))^{1/2} \leq \frac{1}{2\sqrt{k}}.$$

Next, set

$$\Delta_k = \max_{1 \leq i \leq k} |V_i^k - i/k|, \quad \omega_g(\delta) = \sup_{\substack{s, t \in [0, 1] \\ |s - t| \leq \delta}} |g(s) - g(t)|.$$

If  $F_k$  denotes the empirical distribution function of  $U_1, \dots, U_k$ , the almost sure absence of ties gives  $F_k(V_i^k) = i/k$ , hence

$$\Delta_k \leq \sup_{0 \leq t \leq 1} |F_k(t) - t| \xrightarrow[k \rightarrow \infty]{a.s.} 0$$

by the Glivenko–Cantelli theorem ([Exercise 1.8](#)). Now

$$|D_{k,g}(f) - A_{k,g}(f)| \leq \min(1, \omega_g(\Delta_k)).$$

Uniform continuity of  $g$  and dominated convergence therefore give

$$\sup_{f \in \mathcal{C}} \mathbb{E} [|D_{k,g}(f) - d(f, g)|] \leq \frac{1}{2\sqrt{k}} + \mathbb{E} [\min(1, \omega_g(\Delta_k))] =: \varepsilon_{k,g} \rightarrow 0.$$

This bound does not depend on  $f$ . By independence, it therefore remains valid for  $f = X_n$ , uniformly in  $n$ , as well as for  $f = X$ .

Now let  $g_1, \dots, g_r \in \mathcal{C}$  and let  $P$  be a polynomial in  $r$  variables. For fixed  $k$ , the assumption gives

$$\mathbb{E} [P(D_{k,g_1}(X_n), \dots, D_{k,g_r}(X_n))] \xrightarrow[n \rightarrow \infty]{} \mathbb{E} [P(D_{k,g_1}(X), \dots, D_{k,g_r}(X))],$$

because the expression being tested is a bounded continuous function of  $(X_n(V_1^k), \dots, X_n(V_k^k))$ . The polynomial  $P$  is Lipschitz on  $[0, 1]^r$ : there exists  $L \geq 0$  such that  $|P(a) - P(b)| \leq L \sum_{j=1}^r |a_j - b_j|$  for  $a, b \in [0, 1]^r$ . By the triangle inequality, the preceding bound implies

$$\begin{aligned} \limsup_{n \rightarrow \infty} |\mathbb{E} [P(d(X_n, g_1), \dots, d(X_n, g_r))] \\ - \mathbb{E} [P(d(X, g_1), \dots, d(X, g_r))]| \leq 2L \sum_{j=1}^r \varepsilon_{k,g_j}. \end{aligned}$$

Letting  $k$  tend to infinity yields convergence of the expectations of all polynomials in these distances. We then choose the  $g_j$  from a dense sequence in  $(\mathcal{C}, d)$ , which is separable and bounded by (8), and question (4) implies that  $X_n$  converges in distribution to  $X$ .

**Remark.** Part of this problem is inspired by Aldous, D., & Pitman, J. (2002). Invariance principles for non-uniform random mappings and trees. In *Asymptotic Combinatorics with Application to Mathematical Physics* (pp. 113-147). Springer, Dordrecht.

**Solution to Problem 4.13.** [statement p. 161]

(1) We begin by showing that

$$\frac{P(t)}{t} \xrightarrow[t \rightarrow \infty]{a.s.} 1. \quad (9)$$

Clearly  $P(t) \rightarrow \infty$  as  $t \rightarrow \infty$ , because  $P$  is nondecreasing and  $P(S_n) = n$ . Observe that  $S_{P(t)} \leq t < S_{P(t)+1}$ . Let  $0 < \varepsilon < 1$ . By the law of large numbers, there exists  $N > 0$  such that, for  $n \geq N$ , we have  $|S_n/n - 1| \leq \varepsilon$ . Then, for  $t$  large enough that  $P(t) \geq N$ , we have

$$(1 - \varepsilon)P(t) \leq S_{P(t)} \leq t < S_{P(t)+1} \leq (1 + \varepsilon)(P(t) + 1).$$

It follows that

$$\frac{1}{1 + \varepsilon} - \frac{1}{t} \leq \frac{P(t)}{t} \leq \frac{1}{1 - \varepsilon}$$

and (9) follows.

Let us return to uniform convergence on  $[0, T]$ . On the event of probability 1 where (9) holds, fix  $\varepsilon > 0$  and choose  $N > 0$  such that  $|P(t)/t - 1| \leq \varepsilon$  for  $t \geq N$ . Splitting the times into  $nu \leq N$  and  $nu \geq N$ , we obtain

$$\sup_{0 \leq u \leq T} \left| \frac{P(nu)}{n} - u \right| \leq \frac{P(N) + N}{n} + T\varepsilon.$$

The first term tends to zero. We conclude by letting  $\varepsilon \downarrow 0$ .

(2) By definition, observe that, for  $u = S_K/n$ , we have

$$\left| \frac{P(un)}{n} - u \right| = \left| \frac{K}{n} - \frac{S_K}{n} \right| = \frac{|S_K - K|}{n}.$$

For every  $M > 0$ , the central limit theorem gives

$$\mathbb{P}(|S_K - K| > M) \rightarrow 1 \quad (K \rightarrow \infty).$$

Indeed,  $(S_K - K)/\sqrt{K}$  converges in distribution to a standard normal law, which has no atom at zero, and  $M/\sqrt{K} \rightarrow 0$ . Consequently,

$$\mathbb{P} \left( \sup_{K \geq 1} |S_K - K| > M \right) \geq \sup_{K \geq 1} \mathbb{P}(|S_K - K| > M) = 1.$$

Intersecting these events over integers  $M \geq 1$  gives  $\sup_K |S_K - K| = \infty$  almost surely. The preceding identity therefore implies, simultaneously for every integer  $n \geq 1$ ,

$$\sup_{u \geq 0} \left| \frac{P(un)}{n} - u \right| = \infty \quad \text{almost surely.}$$

(3) Fix  $T > 0$  and set

$$\Delta_n = \sup_{0 \leq t \leq T} \left| \frac{P(nt)}{n} - t \right|, \quad \omega_{T+1}(f, \delta) = \sup_{\substack{0 \leq s, t \leq T+1 \\ |s-t| \leq \delta}} |f(s) - f(t)|.$$

By (1),  $\Delta_n \rightarrow 0$  almost surely. For  $0 < \delta \leq 1$ , on  $\{\Delta_n \leq \delta\}$ , the times  $t$  and  $P(nt)/n$  belong to  $[0, T+1]$  and are at distance at most  $\delta$ . Thus, for every  $\eta > 0$ ,

$$\begin{aligned} \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| Y_n \left( \frac{P(nt)}{n} \right) - Y_n(t) \right| > \eta \right) \\ \leq \mathbb{P}(\Delta_n > \delta) + \mathbb{P}(\omega_{T+1}(Y_n, \delta) > \eta). \end{aligned}$$

Convergence in distribution of the restrictions of  $Y_n$  to  $[0, T+1]$  implies their tightness. The tightness criterion using moduli of continuity therefore gives

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}(\omega_{T+1}(Y_n, \delta) > \eta) = 0.$$

Letting  $n \rightarrow \infty$  and then  $\delta \downarrow 0$  in the preceding bound yields the required convergence in probability. For  $T = 0$ , the difference is identically zero since  $P(0) = 0$  almost surely. No independence between  $P$  and  $Y_n$  is needed.

(4) This follows from the global Picard–Lindelöf theorem ( $\beta$  is Lipschitz).

(5) *Step one.* First check that, almost surely, there exists  $K > 0$  such that  $|\bar{Z}_n(t)| \leq K$  for every  $n \geq 1$  and every  $0 \leq t \leq T$ .

By question (1), almost surely there exist two constants  $a, b > 0$  such that  $P(t) \leq a + bt$  for every  $t \geq 0$ . Moreover, since  $\beta$  is Lipschitz, there exist two constants  $A, B > 0$  such that  $|\beta(z)| \leq A + B|z|$  for  $z \in \mathbb{R}$ . Then, for  $0 \leq t \leq T$ , by definition of  $Z_n$ :

$$|\bar{Z}_n(t)| \leq |\bar{Z}_n(0)| + \frac{a}{n} + b \int_0^t \beta(\bar{Z}_n(s)) ds \leq |\bar{Z}_n(0)| + \frac{a}{n} + b \int_0^t (A + B|\bar{Z}_n(s)|) ds.$$

We deduce that

$$|\bar{Z}_n(t)| \leq |\bar{Z}_n(0)| + \frac{a}{n} + bAT + bB \int_0^t |\bar{Z}_n(s)| ds.$$

By Gronwall's lemma, it follows that, for every  $0 \leq t \leq T$ :

$$|\bar{Z}_n(t)| \leq \left( |\bar{Z}_n(0)| + \frac{a}{n} + bAT \right) e^{bBT}.$$

Since  $|\bar{Z}_n(0)|$  converges as  $n \rightarrow \infty$ , the right-hand side is bounded, which completes step one.

*Step two.* We show that almost surely there exist  $M > 0$  and  $L > 0$  such that, for every  $t \in [0, T]$ , we have

$$|\bar{Z}_n(t) - z(t)| \leq |\bar{Z}_n(0) - z_0| + \sup_{0 \leq s \leq t} \frac{1}{n} |\hat{P}(nMs)| + L \int_0^t |\bar{Z}_n(s) - z(s)| ds$$

where we have set  $\hat{P}(t) = P(t) - t$ .

Using the fact that  $z(t) = z_0 + \int_0^t \beta(z(s))ds$ , we write

$$\begin{aligned} \bar{Z}_n(t) - z(t) &= \bar{Z}_n(0) - z_0 + \frac{1}{n} \hat{P} \left( n \int_0^t \beta(\bar{Z}_n(s)) ds \right) \\ &\quad - \left( \int_0^t \beta(z(s)) ds - \int_0^t \beta(\bar{Z}_n(s)) ds \right). \end{aligned}$$

Let  $L > 0$  be a Lipschitz constant for  $\beta$  and let  $M > 0$  be such that  $|\beta(x)| \leq M$  for every  $-K \leq x \leq K$ . Then, for  $0 \leq t \leq T$ :

$$|\bar{Z}_n(t) - z(t)| \leq |\bar{Z}_n(0) - z_0| + \sup_{0 \leq s \leq t} \frac{1}{n} |\hat{P}(nMs)| + L \int_0^t |\bar{Z}_n(s) - z(s)| ds.$$

This completes step two.

*Step three.* By Gronwall's lemma, for every  $t \in [0, T]$ , we have

$$|\bar{Z}_n(t) - z(t)| \leq \left( |\bar{Z}_n(0) - z_0| + \sup_{0 \leq s \leq t} \frac{1}{n} |\hat{P}(nMs)| \right) e^{Lt}.$$

Thus

$$\sup_{0 \leq t \leq T} |\bar{Z}_n(t) - z(t)| \leq \left( |\bar{Z}_n(0) - z_0| + \sup_{0 \leq s \leq T} \frac{1}{n} |\hat{P}(nMs)| \right) e^{LT},$$

which tends almost surely to 0 by assumption and by question (1).

**Remark.** This problem is inspired by Theorem 5.2 in [Stochastic Epidemic Models and Their Statistical Analysis](#) (Andersson & Britton).

\* \* \*

**Solution to Exercise 4.14.** [statement p. 162] Write  $\mathcal{F} = \{B \cap C(K, E) : B \in \mathcal{B}(E)^{\otimes K}\}$ , which is indeed a sigma-algebra on  $C(K, E)$ .

Since, for every  $x \in K$  and  $A \in \mathcal{B}(E)$ ,  $\{f \in C(K, E) : f(x) \in A\} = \{f \in E^K : f(x) \in A\} \cap C(K, E) \in \mathcal{F}$ , the sigma-algebra  $\mathcal{F}$  indeed makes the projections  $\Pi_x$  measurable.

Now let  $\mathcal{G}$  be a sigma-algebra on  $C(K, E)$  making all the projections  $\Pi_x$  measurable, and consider

$$\{B \in \mathcal{B}(E)^{\otimes K} : B \cap C(K, E) \in \mathcal{G}\}.$$

This is a sigma-algebra containing the cylinder sets of  $\mathcal{B}(E)^{\otimes K}$ , so it equals  $\mathcal{B}(E)^{\otimes K}$ , which completes the proof.

## 5 Donsker's theorem, Brownian bridge

Related lecture notes: [Chapter 5](#), p. 55 and [Chapter 6](#), p. 62.

### 5.1 Exercises

**Exercise 5.1.** [solution p. 185] Let  $(B_t)_{t \geq 0}$  be standard Brownian motion taking values in  $\mathbb{R}$ . For every  $c > 0$ , set:

$$B_t^{(c)} = \frac{1}{\sqrt{c}} B_{ct}.$$

- (1) Show that the family  $(B^{(c)}, c > 0)$  of  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ -valued random variables is tight.
- (2) Suppose that  $0 < c < 1$ . Show the equality in distribution

$$((B_{c+t}, t \geq 0), (B_s^{(c)}, 0 \leq s \leq 1)) \stackrel{(d)}{=} ((\sqrt{c}B'_1 + B_t, t \geq 0), (B'_s, 0 \leq s \leq 1)),$$

where  $B'$  is a process independent of  $B$  with the same distribution.

- (3) Let  $0 < t_1 < t_2 < \dots < t_k \leq 1$  and  $0 < t'_1 < \dots < t'_k \leq 1$  be strictly positive real numbers, and retain the notation of the previous question. Show the convergence in distribution

$$(B_{t_1}, \dots, B_{t_k}, B_{t_1}^{(c)}, \dots, B_{t'_k}^{(c)}) \xrightarrow[c \rightarrow 0]{(d)} (B_{t_1}, \dots, B_{t_k}, B'_{t'_1}, \dots, B'_{t'_k}).$$

- (4) Deduce that  $(B, B^{(c)})$  converges in distribution in  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})^2$  to  $(B, B')$  as  $c \rightarrow 0$ . Explain why this result may seem surprising.
- (5) What happens in the previous question if we instead let  $c \rightarrow \infty$ ?

\* \* \*

**Exercise 5.2.** [solution p. 186] Let  $(X_i)_{i \geq 1}$  be i.i.d. centered real-valued random variables with variance 1. Set  $S_0 = 0$ ,  $S_n = X_1 + \dots + X_n$ , as well as  $A_0 = 0$ ,  $A_n = S_1 + \dots + S_n$ . Define  $(S_t)_{t \geq 0}$  and  $(A_t)_{t \geq 0}$  by linear interpolation. Finally, for  $0 \leq t \leq 1$ , write

$$S_t^{(n)} = \frac{S_{nt}}{\sqrt{n}}, \quad A_t^{(n)} = \frac{A_{nt}}{n^{3/2}}.$$

Show that  $(S^{(n)}, A^{(n)})$  converges in distribution in  $\mathcal{C}([0, 1], \mathbb{R})^2$  to a limit whose distribution you should identify.

\* \* \*

**Exercise 5.3.** [solution p. 186] Let  $(b_t)_{0 \leq t \leq 1}$  be a standard Brownian bridge.

Show that, almost surely, for all  $s, t \in ]0, 1[$ , if  $s \neq t$  and if  $s$  and  $t$  are times of local maximum of  $b$ , then  $b_s \neq b_t$ .

*You may assume that, almost surely, the values of the local maxima of Brownian motion are pairwise distinct.*

**Exercise 5.4.** [solution p. 187] **(Functional strong law of large numbers)**

- (1) Let  $(X_i)_{i \geq 1}$  be i.i.d. real-valued random variables such that  $\mathbb{E}[|X_1|] < \infty$ . Set  $S_0 = 0$ ,  $S_n = X_1 + \cdots + X_n$  and define  $S_t$  by linear interpolation for  $t \geq 0$ . Finally, write

$$S_t^{(n)} = \frac{S_{nt}}{n}, \quad t \geq 0, \quad n \geq 1.$$

Show that, almost surely, the following convergence holds for the topology of uniform convergence on compact sets:

$$(S_t^{(n)})_{t \geq 0} \xrightarrow[n \rightarrow \infty]{} (\mathbb{E}[X_1] t)_{t \geq 0}.$$

- (2) Let  $(N_t)_{t \geq 0}$  be a standard Poisson process with  $N_0 = 0$ . Show that, almost surely, the following convergence holds for the topology of uniform convergence on compact sets:

$$\left( \frac{1}{n} N_{nt} \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{} (t)_{t \geq 0}.$$

\* \* \*

**Exercise 5.5.** [solution p. 187] **(Details of the proof of Lévy's theorem)** Consider a simple symmetric random walk  $(S_n)_{n \geq 0}$  (that is,  $S_0 = 0$  and  $S_n = X_1 + \cdots + X_n$  with  $(X_i)_{i \geq 1}$  i.i.d. and  $\mathbb{P}(X_1 = \pm 1) = \frac{1}{2}$ ). Set

$$I_n = \min_{0 \leq k \leq n} S_k, \quad \theta_0 = 0,$$

and, for  $n \geq 1$ ,

$$\theta_n = \text{Card}\{1 \leq k \leq n : S_k - I_k \neq S_{k-1} - I_{k-1}\}.$$

Also define  $\theta_0^{-1} = 0$  and  $\theta_k^{-1} = \inf\{n \geq 1 : \theta_n = k\}$  for  $k \geq 1$ . Finally, set  $X_n = S_{\theta_n^{-1}} - I_{\theta_n^{-1}}$  and define  $(X_t)_{t \geq 0}$  by linear interpolation.

- (1) Justify that  $(X_t)_{t \geq 0}$  and  $(|S_t|)_{t \geq 0}$  have the same distribution.  
 (2) Check that  $(S_n - I_n)_{n \geq 0}$  is a null recurrent Markov chain and deduce that

$$\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{S_k - I_k = 0} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0.$$

- (3) Show that, for every  $T > 0$ ,

$$\max_{0 \leq k \leq \lfloor nT \rfloor} \left| \frac{\theta_k^{-1}}{n} - \frac{k}{n} \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

and deduce that

$$\left( \frac{X_{nt}}{\sqrt{n}} : t \geq 0 \right) \xrightarrow[n \rightarrow \infty]{(d)} (B_t - I_t : t \geq 0).$$

- (4) Conclude that  $(B_t - I_t : t \geq 0)$  and  $(|B_t| : t \geq 0)$  have the same distribution.

**Exercise 5.6.** [solution p. 189] Let  $X$  be a nonconstant real-valued random variable with characteristic function denoted by  $\phi$ . Recall that  $X$  is called lattice if there exist  $b \in \mathbb{R}$  and  $h > 0$  such that  $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$ , and that its span is the largest such  $h$ .

- (1) Show that  $X$  is lattice if and only if there exists  $t_0 \in \mathbb{R}_*$  such that  $|\phi(t_0)| = 1$ .
- (2) Show that  $X$  has span  $h$  if and only if  $|\phi(2\pi/h)| = 1$  and  $|\phi(t)| < 1$  for  $0 < |t| < 2\pi/h$ .

Now suppose that  $X$  takes integer values. Recall that  $X$  is called aperiodic if its span is 1.

- (3) Suppose that  $X$  is lattice with span  $h$ . Must we have  $h \in \mathbb{Z}$ ?
- (4) Is it true that if the greatest common divisor of the support  $\text{supp}(X)$  of  $X$  is 1, then  $X$  is aperiodic?
- (5) Suppose that there exists  $k \in \text{supp}(X)$  such that  $\gcd(\{i - k : i \in \text{supp}(X)\}) = 1$ . Show that  $X$  is aperiodic.

\* \* \*

**Exercise 5.7.** [solution p. 190] Give an example of a family of integer-valued random variables  $(X_i^{(n)})_{i,n \geq 1}$  such that, for every  $n \geq 1$ , the random variables  $(X_i^{(n)})_{i \geq 1}$  are i.i.d. and aperiodic, and such that  $S_n^{(n)} = X_1^{(n)} + \cdots + X_n^{(n)}$  satisfies a central limit theorem but not a local limit theorem.

\* \* \*

**Exercise 5.8.** [solution p. 190] Let  $n \geq 1$  be an integer. Consider an urn containing  $2n$  balls, of which  $n$  are labeled  $+1$  and  $n$  are labeled  $-1$ . Draw  $2n$  balls successively, uniformly from the balls present in the urn at each draw. Set  $S_0 = 0$  and, for  $1 \leq k \leq 2n$ , let  $S_k$  denote the sum of the first  $k$  balls drawn. Define  $S_t$  for  $0 \leq t \leq 2n$  by linear interpolation. Study convergence in distribution of the sequence

$$\left( \frac{S_{2nt}}{\sqrt{n}} : 0 \leq t \leq 1 \right)$$

in the space  $\mathcal{C}([0, 1], \mathbb{R})$  in the following two cases:

- (1) draws are made **with** replacement,
- (2) draws are made **without** replacement.

\* \* \*

**Exercise 5.9.** [solution p. 191] Let  $(X_i)_{i \geq 1}$  be a sequence of independent identically distributed random variables taking values in  $\mathbb{Z}$ , with finite variance  $\sigma^2$ . Suppose that  $\mathbb{E}[X_1] = 0$  and that  $\mathbb{P}(X_1 = k) > 0$  for every  $k \in \mathbb{Z}$ . For  $n \geq 1$ , set  $S_n = X_1 + \cdots + X_n$  and let  $\mu_n$  denote the conditional law of  $X_1$  given that  $S_n = 0$ .

- (1) Prove that the sequence of probability measures  $(\mu_n)_{n \geq 1}$  converges weakly.
- (2) Alix says: if  $Y_n$  is a random variable with distribution  $\mu_n$ , the sequence of random variables  $(Y_n)_{n \geq 1}$  is uniformly integrable. Is she right?

**Problem 5.10.** [solution p. 191]

**Part one.** Consider real-valued random variables  $I_n, I_n(\delta), I, I(\delta)$  for  $\delta > 0$  and  $n \geq 1$  such that:

- (i) almost surely  $I(\delta) \rightarrow I$  as  $\delta \rightarrow 0$ ,
  - (ii) for every  $\delta > 0$ ,  $I_n(\delta)$  converges in distribution to  $I(\delta)$  as  $n \rightarrow \infty$ ,
  - (iii) for every  $n \geq 1$  and  $\delta > 0$ ,  $\mathbb{E}[|I_n(\delta) - I_n|] \leq \delta$ .
- (1) Show that  $I_n \rightarrow I$  in distribution as  $n \rightarrow \infty$ .

In what follows, consider i.i.d. aperiodic random variables  $(X_i)_{i \geq 1}$  taking values in  $\mathbb{Z}$  with  $\mathbb{E}[X_1] = 0$ ,  $\mathbb{E}[X_1^2] = 1$ . Set  $S_0 = 0$  and  $S_n = X_1 + \dots + X_n$  for  $n \geq 1$ . By convention, the truncated functions  $|x|^{-\gamma} \mathbb{1}_{|x| > \delta}$  and  $|x|^{-\gamma} \mathbb{1}_{0 < |x| \leq \delta}$  are equal to 0 at  $x = 0$ . Let  $(B_t)_{t \geq 0}$  be standard real-valued Brownian motion started at 0. Also fix  $0 < \gamma < 1$ .

**Part two.**

- (2) Show that, for every  $\delta > 0$ , almost surely the Lebesgue measure of  $\{t \in [0, 1] : B_t = \delta\}$  is zero.
- (3) Show that, for every  $\delta > 0$ , the following convergence holds in distribution:

$$\frac{1}{n^{1-\gamma/2}} \sum_{i=1}^n \frac{1}{|S_i|^\gamma} \mathbb{1}_{|S_i| > \delta\sqrt{n}} \xrightarrow[n \rightarrow \infty]{(d)} \int_0^1 \frac{1}{|B_t|^\gamma} \mathbb{1}_{|B_t| > \delta} dt.$$

**Part three.**

- (4) Show that, almost surely,  $t \mapsto \frac{1}{|B_t|^\gamma}$  is integrable on  $[0, 1]$  with respect to Lebesgue measure.
- (5) Show that the following convergence holds in distribution:

$$\frac{1}{n^{1-\gamma/2}} \sum_{i=1, S_i \neq 0}^n \frac{1}{|S_i|^\gamma} \xrightarrow[n \rightarrow \infty]{(d)} \int_0^1 \frac{1}{|B_t|^\gamma} dt.$$

(\*) What happens for  $\gamma = 1$ ?

\* \* \*

**Problem 5.11.** [solution p. 194] The aim of this problem is to study the asymptotic behavior of a random walk that can sometimes make several identical steps in a row.

Let  $\mathcal{C}$  denote the set of continuous functions from  $[0, 1]$  to  $\mathbb{R}$  equipped with the uniform norm.

The first three parts can be treated independently of one another.

**Part one.** Let  $H \in [1/2, 1]$ . Consider a random variable  $W = (W_t)_{0 \leq t \leq 1}$  taking values in  $\mathcal{C}$  with the following properties:

- a)  $(W_t)_{0 \leq t \leq 1}$  is a centered Gaussian process, that is, for every  $k \geq 1$ , all  $0 \leq t_1 < \dots < t_k \leq 1$  and all  $a_1, \dots, a_k \in \mathbb{R}$ , the random variable  $a_1 W_{t_1} + \dots + a_k W_{t_k}$  is centered Gaussian.
- b) For all  $s, t \in [0, 1]$ , we have  $\mathbb{E}[W_s W_t] = \frac{1}{2}(s^{2H} + t^{2H} - |s - t|^{2H})$ .

The family  $(\mathbb{E}[W_s W_t])_{0 \leq s, t \leq 1}$  is called the covariance function of  $W$ . Recall that the law of a centered Gaussian process is determined by its covariance function. Indeed, if  $(X_1, \dots, X_n)$  is a centered Gaussian vector (that is, every linear combination of its coordinates has a normal distribution), its characteristic function, writing  $u = (u_1, \dots, u_n) \in \mathbb{R}^n$ , is

$$\phi_{(X_1, \dots, X_n)}(u) = \mathbb{E} \left[ e^{i(u_1 X_1 + \dots + u_n X_n)} \right] = e^{-\frac{1}{2} u^T K u},$$

where  $^T$  denotes transposition and  $K = (\mathbb{E}[X_i X_j])_{1 \leq i, j \leq n}$  is the covariance matrix.

Finally, let  $\mathbb{P}_H$  denote the law of  $W$  (the existence of  $\mathbb{P}_H$  will be proved later in this problem).

- (1) If  $Z$  is a random variable with centered normal distribution  $\mathcal{N}(0, \sigma^2)$  and variance  $\sigma^2$ , prove that  $\mathbb{E}[Z^4] = 3\sigma^4$ .
- (2) Assume that  $\mathbb{P}_H$  exists for every  $1/2 \leq H \leq 1$ . Prove that the map

$$\begin{array}{ccc} \Phi & : & [1/2, 1] \rightarrow (\mathcal{M}_1(\mathcal{C}), d_{LP}) \\ H & \mapsto & \mathbb{P}_H \end{array}$$

is continuous.

- (3) Prove the existence of  $\mathbb{P}_H$  when  $H = 1$ .

**Part two.** Let  $H \in ]1/2, 1[$ . In this part,  $(Y_i)_{i \geq 1}$  are centered real-valued random variables with finite variance. Set  $S_0 = 0$  and  $S_n = Y_1 + \dots + Y_n$  for integers  $n \geq 1$ . Finally, define  $S_t$  for  $t \geq 0$  by linear interpolation and, for  $n \geq 1$  and  $t \in [0, 1]$ , set:

$$Z_n(t) = \frac{S_{nt}}{n^H}.$$

- (4) In this question, suppose that the following properties hold:

- a) for every  $n \geq 1$  and  $i \geq 1$ ,  $S_{n+i} - S_n$  has the same distribution as  $S_i$ .
- b)  $\mathbb{E}[S_n^2] = O(n^{2H})$  as  $n \rightarrow \infty$ .
- c)  $(Z_n)$  converges in finite-dimensional distributions in  $\mathcal{C}$ .

Prove that  $Z_n$  converges in distribution in  $\mathcal{C}$  as  $n \rightarrow \infty$ .

(5) In this question, suppose that the following properties hold:

- a)  $(Y_i)_{i \geq 1}$  is a Gaussian family.
- b) for every  $n \geq 1$  and  $i \geq 1$ ,  $S_{n+i} - S_n$  has the same distribution as  $S_i$ .
- c)  $\mathbb{E}[S_n^2] \sim c \cdot n^{2H}$  as  $n \rightarrow \infty$ , with  $c > 0$ .

Prove that  $Z_n$  converges in distribution, as  $n \rightarrow \infty$ , to  $(\sqrt{c} \cdot W_t : 0 \leq t \leq 1)$  where the law of  $W$  was defined in part one.

**Part three.** Let  $p \in [0, 1]$ . In this part,  $(X_n)_{n \geq 1}$  are independent random variables such that  $X_1$  has a Bernoulli distribution with parameter  $1/2$  and, for  $n \geq 2$ , the random variable  $X_n$  has a Bernoulli distribution with parameter  $1 - p$ . For  $n \geq 1$ , set  $Y_n = (-1)^{X_1 + \dots + X_n}$ , then  $S_0 = 0$  and  $S_n = Y_1 + \dots + Y_n$ . Thus the steps of the walk  $(S_n)$  are  $\pm 1$ , the first step is  $\pm 1$  with probabilities  $1/2$  and, for  $n \geq 2$ , the  $n$ -th step equals the  $(n-1)$ -th step with probability  $p$  and the opposite of the  $(n-1)$ -th step with probability  $1 - p$ . Finally, define  $S_t$  for  $t \geq 0$  by linear interpolation.

(6) Prove that, for all  $i, n \geq 1$ , we have  $\mathbb{E}[Y_i Y_{i+n}] = (2p - 1)^n$ .

(7) When  $p = 1/2$ , prove that  $(S_{nt}/\sqrt{n} : 0 \leq t \leq 1)$  converges in distribution in  $\mathcal{C}$  to the law of  $W$  defined in part one for  $H = 1/2$ .

**Part four.** Let  $H \in ]1/2, 1[$ . Let  $P$  be a random variable taking values in  $[1/2, 1]$  with distribution

$$(1 - H)2^{3-2H}(1 - x)^{1-2H} \mathbb{1}_{[1/2, 1]}(x) dx.$$

In this part, let  $\mathbf{X} = (X_n)_{n \geq 1}$  denote a sequence of random variables which, conditionally on  $P$ , are independent and such that  $X_1$  has a Bernoulli distribution with parameter  $1/2$  and, for  $n \geq 2$ , the random variable  $X_n$  has a Bernoulli distribution with parameter  $1 - P$ . Formally, if, for  $p \in [0, 1]$ ,  $\mu_p$  is the law of a sequence of independent Bernoulli random variables whose first term has parameter  $1/2$  and whose other terms have parameter  $1 - p$ , the law  $\mathbb{P}_{\mathbf{X}}$  of  $\mathbf{X}$  is characterized by the fact that, for every event  $A$  in  $\{0, 1\}^{\mathbb{N}^*}$ , we have

$$\mathbb{P}_{\mathbf{X}}(A) = \int_0^1 \mu_x(A) \cdot (1 - H)2^{3-2H}(1 - x)^{1-2H} \mathbb{1}_{[1/2, 1]}(x) dx.$$

Now let  $(\mathbf{X}^k)_{k \geq 1}$  be a sequence of i.i.d. random variables with the same distribution as  $\mathbf{X}$ . For  $k \geq 1$ , set  $\mathbf{X}^k = (X_n^k)_{n \geq 1}$ . For  $k, n \geq 1$ , set  $Y_n^k = (-1)^{X_1^k + \dots + X_n^k}$ .

(8) Prove that, as  $k \rightarrow \infty$ ,  $\left( \frac{Y_{n+1}^1 + \dots + Y_{n+1}^k}{\sqrt{k}} : n \geq 1 \right)$  converges in finite-dimensional distributions to a centered Gaussian process  $(G_n)_{n \geq 1}$  satisfying:

- a) for every  $n \geq 1$ ,  $\mathbb{E}[G_n^2] = 1$ .

- b) for every  $n \geq 1$  and  $i \geq 1$ , the quantity  $r(n) = \mathbb{E}[G_i G_{i+n}]$  does not depend on  $i$ , and we have

$$r(n) \underset{n \rightarrow \infty}{\sim} \frac{1}{c_H^2} \frac{H(2H-1)}{n^{2-2H}} \quad \text{with } c_H = \sqrt{\frac{H(2H-1)}{\Gamma(3-2H)}}.$$

**Remark.** You may assume that  $\int_{1/2}^1 (2u-1)^n (1-u)^{1-2H} du \sim \frac{1}{c_H^2} \frac{H(2H-1)}{(1-H)2^{3-2H}} \cdot \frac{1}{n^{2-2H}}$  as  $n \rightarrow \infty$ .

- (9) Set  $V_0 = 0$  and  $V_n = G_1 + \dots + G_n$  for  $n \geq 1$ , and define  $V_t$  for  $t \geq 0$  by linear interpolation. Prove that

$$\left( c_H \frac{V_{nt}}{n^H} : 0 \leq t \leq 1 \right) \xrightarrow[n \rightarrow \infty]{(d)} (W_t : 0 \leq t \leq 1),$$

where the law of  $W$  was defined in part one.

## 5.2 Solutions

**Solution to Exercise 5.1.** [statement p. 179]

- (1) For every  $c > 0$ , the process  $B^{(c)}$  is continuous, centered Gaussian, and

$$\mathbb{E} [B_s^{(c)} B_t^{(c)}] = \frac{1}{c} \min(cs, ct) = \min(s, t).$$

It therefore has the law of Brownian motion. The family of its laws consists of a single probability measure on the Polish space  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ : it is tight.

- (2) The process  $(B_{c+t} - B_c)_{t \geq 0}$  is Brownian motion independent of  $(B_s)_{0 \leq s \leq c}$ , and hence of  $(B_s^{(c)})_{0 \leq s \leq 1}$ . Since  $B_c = \sqrt{c} B_1^{(c)}$ , the claimed equality in distribution follows.

- (3) By (2),

$$\begin{aligned} (B_{c+t_i} : 1 \leq i \leq k; B_{t'_j}^{(c)} : 1 \leq j \leq k') \\ \stackrel{(d)}{=} (\sqrt{c} B'_1 + B_{t_i} : 1 \leq i \leq k; B'_{t'_j} : 1 \leq j \leq k'). \end{aligned}$$

The right-hand side converges almost surely to the desired vector as  $c \rightarrow 0$ . By continuity of  $B$ ,  $\max_{1 \leq i \leq k} |B_{c+t_i} - B_{t_i}| \rightarrow 0$  almost surely. Slutsky's lemma therefore allows us to replace  $B_{c+t_i}$  by  $B_{t_i}$ .

- (4) Tightness of the two marginals implies tightness of the pairs: if a compact subset  $K$  of  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$  satisfies  $\mathbb{P}(B \notin K) \leq \varepsilon$ , then

$$\mathbb{P}((B, B^{(c)}) \notin K \times K) \leq 2\varepsilon.$$

The argument in (2)–(3) applies on every  $[0, T]$ : simply replace  $c$  by  $cT$  in the shift of the first process and  $\sqrt{c} B_1'$  by  $\sqrt{c} B_T'$ . The finite-dimensional distributions of the pairs therefore converge to those of two independent Brownian motions. Together with tightness, this gives convergence in distribution in  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})^2$ .

This result may seem surprising because, for each  $c$ , the process  $B^{(c)}$  is entirely determined by  $B$ , whereas the two limiting processes are independent.

- (5) The limit is again a pair of independent Brownian motions. Indeed, if  $W = B^{(c)}$ , then  $W$  is Brownian motion and  $W^{(1/c)} = B$ . Thus

$$(B, B^{(c)}) \stackrel{(d)}{=} (B^{(1/c)}, B).$$

Apply (4) with  $1/c \rightarrow 0$ , then swap the two coordinates.

\* \* \*

**Solution to Exercise 5.2.** [statement p. 179] The limit is the pair

$$\left( (B_t)_{0 \leq t \leq 1}, \left( \int_0^t B_s ds \right)_{0 \leq t \leq 1} \right),$$

where  $B$  is standard Brownian motion. To see this, set  $J_t^{(n)} = \int_0^t S_s^{(n)} ds$ . The map

$$f \mapsto \left( f, \left( t \mapsto \int_0^t f(s) ds \right) \right)$$

is continuous from  $\mathcal{C}([0, 1], \mathbb{R})$  to its square, for the uniform norms. Donsker's theorem and the continuous mapping theorem therefore give the claimed convergence for  $(S^{(n)}, J^{(n)})$ .

It remains to compare  $A^{(n)}$  and  $J^{(n)}$ . Linear interpolation of  $A$  gives the identity

$$A_t^{(n)} = \int_0^t \frac{S_{[ns]}}{\sqrt{n}} ds.$$

On the interval  $[(i-1)/n, i/n]$ ,  $|S_{[ns]} - S_{ns}| \leq |X_i|$  for almost every  $s$ . Thus

$$\sup_{0 \leq t \leq 1} |A_t^{(n)} - J_t^{(n)}| \leq \frac{1}{n\sqrt{n}} \sum_{i=1}^n |X_i|,$$

whose expectation equals  $\mathbb{E}[|X_1|]/\sqrt{n}$  and tends to 0. Slutsky's lemma completes the proof.

\* \* \*

**Solution to Exercise 5.3.** [statement p. 179] Fix  $r \in ]0, 1[$ . By Proposition 6.2, the law of  $(b_t)_{0 \leq t \leq r}$  is absolutely continuous with respect to that of Brownian motion on  $[0, r]$ . The assumed property of Brownian motion therefore implies that, almost surely, the values of  $b$  at its local maximum times in  $]0, r[$  are pairwise distinct.

For  $m \geq 2$ , set  $r_m = 1 - 1/m$ . Intersecting the preceding events over all  $m$  gives an event of probability 1. On this event, if  $s, t \in ]0, 1[$  are two distinct local maximum times of  $b$ , choose  $m$  large enough that  $\max(s, t) < r_m$ . These two times are also local maximum times of the restriction of  $b$  to  $[0, r_m]$ , so  $b_s \neq b_t$ . This completes the proof.

**Solution to Exercise 5.4.** [statement p. 180]

- (1) Set  $m = \mathbb{E}[X_1]$  and  $R_k = S_k - km$ . The strong law of large numbers gives  $R_k/k \rightarrow 0$  almost surely. On this event, for every  $\varepsilon > 0$ , there exists  $K$  such that  $|R_k| \leq \varepsilon k$  for  $k \geq K$ . For every  $T > 0$ ,

$$\frac{1}{n} \max_{0 \leq k \leq \lceil nT \rceil} |R_k| \leq \frac{1}{n} \max_{0 \leq k < K} |R_k| + \varepsilon \frac{\lceil nT \rceil}{n}.$$

Letting  $n$  tend to infinity, then  $\varepsilon$  to 0, we see that the left-hand side tends to 0. Linear interpolation then gives

$$\sup_{0 \leq t \leq T} \left| \frac{S_{nt}}{n} - mt \right| \leq \frac{1}{n} \max_{0 \leq k \leq \lceil nT \rceil} |R_k| \rightarrow 0.$$

This holds on the same event of probability 1 for every  $T > 0$ .

- (2) The increments  $(N_k - N_{k-1})_{k \geq 1}$  are i.i.d. with Poisson distribution of parameter 1. By (1), almost surely, for every  $T > 0$ ,

$$\frac{1}{n} \max_{0 \leq k \leq \lceil nT \rceil} |N_k - k| \rightarrow 0.$$

Since  $N_{\lfloor nt \rfloor} \leq N_{nt} \leq N_{\lceil nt \rceil}$ , we obtain

$$\sup_{0 \leq t \leq T} \left| \frac{N_{nt}}{n} - t \right| \leq \frac{1}{n} \max_{0 \leq k \leq \lceil nT \rceil} |N_k - k| + \frac{1}{n} \rightarrow 0.$$

\* \* \*

**Solution to Exercise 5.5.** [statement p. 180] Set  $R_n = S_n - I_n$  and  $\tau_k = \theta_k^{-1}$ . A step of  $R$  is zero exactly when  $S$  reaches a new strict minimum. Consequently,

$$n - \theta_n = -I_n.$$

The strong law of large numbers, in its functional form proved in the previous exercise, implies  $I_n/n \rightarrow 0$  almost surely. Thus  $\theta_n/n \rightarrow 1$ , and all the times  $\tau_k$  are almost surely finite.

- (1) The chain  $R$  has transition probabilities

$$p(0,0) = p(0,1) = \frac{1}{2}, \quad p(j,j-1) = p(j,j+1) = \frac{1}{2} \quad (j \geq 1).$$

The chain  $(X_k)_{k \geq 0} = (R_{\tau_k})_{k \geq 0}$  is obtained by removing the zero steps at 0. By independence of the walk's increments at stopping times, it starts at 0, moves from 0 to 1 with probability 1, and, from every  $j \geq 1$ , moves to  $j-1$  or  $j+1$  with probabilities 1/2. These are exactly the transitions of  $(|S_k|)_{k \geq 0}$ , so the chains have the same distribution. Finally, for a walk with steps  $\pm 1$ ,  $|S_t|$  is the linear interpolation of the  $|S_k|$ : no segment crosses 0 in its interior. The equality in distribution therefore also holds for the continuous processes.

- (2) The chain  $R$  is irreducible on  $\mathbb{N}$ . Until its first visit to 0, it evolves like a simple symmetric random walk. For this walk started at  $j \in \{1, \dots, M-1\}$ , let  $H_0$  and  $H_M$  denote the hitting times of 0 and  $M$ . The gambler's ruin formulas give

$$\mathbb{P}_j(H_0 < H_M) = 1 - \frac{j}{M}, \quad \mathbb{E}_1[H_0 \wedge H_M] = M - 1.$$

They can be recovered by solving, with the boundary conditions, the relations  $q_j = (q_{j-1} + q_{j+1})/2$  and  $u_j = 1 + (u_{j-1} + u_{j+1})/2$  :  $q_j = 1 - j/M$  with  $q_0 = 1, q_M = 0$ , and  $u_j = j(M - j)$ . Letting  $M \rightarrow \infty$  gives  $\mathbb{P}_j(H_0 < \infty) = 1$  and  $\mathbb{E}_1[H_0] = \infty$ . Thus  $R$  is recurrent. Started at 0, it takes its first step to 1 with probability  $1/2$ ; its first return time to 0 therefore has infinite expectation. It is null recurrent.

To deduce the required statement, let  $0 = T_0 < T_1 < T_2 < \dots$  denote its successive returns to 0. The random variables  $T_i - T_{i-1}$  are i.i.d., positive, and have infinite expectation. For each integer  $M \geq 1$ , the strong law gives

$$\liminf_{j \rightarrow \infty} \frac{T_j}{j} \geq \mathbb{E}[\min(T_1, M)] \quad \text{almost surely.}$$

Letting  $M$  tend to infinity shows that  $T_j/j \rightarrow \infty$  almost surely. If  $L_n = \max\{j : T_j \leq n\}$ , then

$$\frac{1}{n} \sum_{k=1}^n \mathbb{1}_{R_k=0} = \frac{L_n}{n} \leq \frac{L_n}{T_{L_n}} \rightarrow 0$$

almost surely, the inequality being used for  $L_n \geq 1$ .

- (3) The convergence  $\theta_n/n \rightarrow 1$  implies  $\tau_k/k \rightarrow 1$  almost surely: indeed,  $\theta_{\tau_k} = k$  and  $\tau_k \rightarrow \infty$ . Moreover,  $\tau_{k+1} \geq \tau_k + 1$ , so  $\tau_k - k$  is nondecreasing and nonnegative. For every  $T > 0$ ,

$$\max_{0 \leq k \leq \lfloor nT \rfloor} \left| \frac{\tau_k}{n} - \frac{k}{n} \right| = \frac{\tau_{\lfloor nT \rfloor} - \lfloor nT \rfloor}{n} \rightarrow 0 \quad \text{almost surely.}$$

We therefore even have almost sure convergence in the required statement.

Extend  $R$  by linear interpolation and set  $R_t^{(n)} = R_{nt}/\sqrt{n}$ . For the interpolated simple random walk,  $R_t = S_t - \inf_{0 \leq s \leq t} S_s$ . The map that subtracts the running infimum is continuous for uniform convergence on compact sets, since

$$\sup_{t \leq T} \left| \left( f(t) - \inf_{s \leq t} f(s) \right) - \left( g(t) - \inf_{s \leq t} g(s) \right) \right| \leq 2 \sup_{t \leq T} |f(t) - g(t)|.$$

Donsker's theorem therefore gives  $R^{(n)} \Rightarrow (B_t - I_t)_{t \geq 0}$ .

Let  $\lambda_n$  be the piecewise affine function taking the value  $\tau_k/n$  at time  $k/n$ . By the preceding argument,  $\lambda_n$  converges almost surely to the identity, uniformly on compact sets. Slutsky's lemma and continuity of composition give

$$R^{(n)} \circ \lambda_n \Rightarrow (B_t - I_t)_{t \geq 0}.$$

Interpolation must be taken into account: if  $\tau_{k+1} = \tau_k + 1$ , the paths  $X_{nt}/\sqrt{n}$  and  $R_{\lambda_n(t)}^{(n)}$  coincide on  $[k/n, (k+1)/n]$ . Otherwise,  $R$  stays at 0 between  $\tau_k$  and  $\tau_{k+1} - 1$ , then rises to 1, whereas  $X$  interpolates directly from 0 to 1. In all cases,

$$\sup_{t \geq 0} \left| \frac{X_{nt}}{\sqrt{n}} - R_{\lambda_n(t)}^{(n)} \right| \leq \frac{1}{\sqrt{n}}.$$

This gives the required convergence in distribution.

- (4) By (1) and Donsker's theorem,  $(X_{nt}/\sqrt{n})_{t \geq 0}$  also converges in distribution to  $(|B_t|)_{t \geq 0}$ , since the map  $f \mapsto |f|$  is continuous. Uniqueness of the limit in distribution gives

$$(B_t - I_t)_{t \geq 0} \stackrel{(d)}{=} (|B_t|)_{t \geq 0}.$$

\* \* \*

**Solution to Exercise 5.6.** [statement p. 181]

- (1) If  $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$ , then  $|\phi(2\pi/h)| = 1$ . Conversely, if  $|\phi(t_0)| = 1$ , the random variable  $e^{it_0 X}$ , of modulus 1, is almost surely constant. There therefore exists  $b \in \mathbb{R}$  such that

$$\mathbb{P}\left(X \in b + \frac{2\pi}{|t_0|}\mathbb{Z}\right) = 1.$$

- (2) Question (1) shows that every  $t \neq 0$  such that  $|\phi(t)| = 1$  yields a lattice spacing  $2\pi/|t|$ . Thus  $h$  is the span if and only if  $h$  is a lattice spacing and no spacing strictly greater than  $h$  is possible, which is exactly equivalent to the two stated conditions.
- (3) Yes (for nonconstant random variables): if  $\mathbb{P}(X \in b + h\mathbb{Z}) = 1$ , taking  $k_1 \neq k_2$  in the support of  $X$  gives  $k_1 = b + ha_1$  and  $k_2 = b + ha_2$  for integers  $a_1$  and  $a_2$ , so  $h$  is rational. Now suppose, for a contradiction, that the span of  $X$  is rational but not an integer. Writing it as  $h = p/q$  with  $p, q \geq 1$  coprime, and considering  $k_1 - k_2$ , we see that  $p$  divides  $k_1 - k_2$ . Since the span of a random variable is unchanged by adding the same quantity to every element of its support, we may assume that all elements of the support of  $X$  are divisible by  $p$ . The random variable  $Y = X/p$  taking values in  $\mathbb{Z}$  then has span  $1/q$ , which is a contradiction, since the span of an integer-valued random variable is clearly at least 1.
- (4) No; take, for example,  $X$  with support  $\{1, 3\}$ ; the span is 2.
- (5) Let  $k \in \text{supp}(X)$  be such that  $\gcd(\{i - k : i \in \text{supp}(X)\}) = 1$ . By (3), the span  $h$  is a strictly positive integer. Since the support is contained in a translate of  $h\mathbb{Z}$ ,  $h$  divides every difference  $i - k$ , and hence their greatest common divisor. Thus  $h = 1$ .

**Solution to Exercise 5.7.** [statement p. 181] For  $n \geq 2$ , take i.i.d. random variables in each row with distribution

$$\mathbb{P}(X_1^{(n)} = 0) = 1 - e^{-n} - n^{-1/2}, \quad \mathbb{P}(X_1^{(n)} = 1) = e^{-n}, \quad \mathbb{P}(X_1^{(n)} = 2) = n^{-1/2}.$$

These three probabilities are positive: the distribution is aperiodic. For  $n = 1$ , take a Bernoulli distribution with parameter  $1/2$ . Set  $m_n = \mathbb{E}[S_n^{(n)}]$  and  $s_n^2 = \text{Var}(S_n^{(n)})$ . For  $n \geq 2$ ,

$$m_n = n(e^{-n} + 2n^{-1/2}), \quad s_n^2 = n(e^{-n} + 4n^{-1/2} - (e^{-n} + 2n^{-1/2})^2) \sim 4\sqrt{n}.$$

The centered random variables in each row are bounded in absolute value by 2, while  $s_n \rightarrow \infty$ . Lindeberg's condition is therefore satisfied, giving

$$\frac{S_n^{(n)} - m_n}{s_n} \Rightarrow \mathcal{N}(0, 1), \quad \frac{S_n^{(n)} - m_n}{2n^{1/4}} \Rightarrow \mathcal{N}(0, 1).$$

However, the local convergence

$$\sup_{k \in \mathbb{Z}} \left| s_n \mathbb{P}(S_n^{(n)} = k) - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(k - m_n)^2}{2s_n^2}\right) \right| \rightarrow 0$$

fails. Indeed, choose an odd integer  $k_n$  with  $|k_n - m_n| \leq 1$ . An odd sum contains at least one term equal to 1, so

$$s_n \mathbb{P}(S_n^{(n)} = k_n) \leq s_n \mathbb{P}(S_n^{(n)} \text{ is odd}) \leq s_n n e^{-n} \rightarrow 0,$$

whereas the Gaussian term at  $k_n$  tends to  $1/\sqrt{2\pi}$ .

\* \* \*

**Solution to Exercise 5.8.** [statement p. 181] Let  $(X_i)_{i \geq 1}$  be a sequence of i.i.d. random variables with distribution  $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ , and write  $S'_n = X_1 + \dots + X_n$  for  $n \geq 0$ . We have  $\mathbb{E}[X_1] = 0$  and  $\mathbb{E}[X_1^2] = 1$ .

- (1) We have convergence in distribution to  $\sqrt{2}$  times Brownian motion on  $[0, 1]$ . Observe that  $(S_i)_{0 \leq i \leq 2n}$  has the same distribution as  $(S'_i)_{0 \leq i \leq 2n}$ , and the result follows from Donsker's theorem.
- (2) We show that the sequence converges in distribution to  $\sqrt{2}$  times the Brownian bridge. The paths  $(S_i)_{0 \leq i \leq 2n}$  have the same distribution as  $(S'_i)_{0 \leq i \leq 2n}$  conditioned on  $S'_{2n} = 0$ . To use exactly the aperiodic version of the theorem, set

$$Y_i = \frac{X_{2i-1} + X_{2i}}{2}, \quad T_k = Y_1 + \dots + Y_k = \frac{S'_{2k}}{2}.$$

The  $Y_i$  are i.i.d., centered, with variance  $1/2$ , and their distribution on  $\{-1, 0, 1\}$ , with probabilities  $1/4, 1/2, 1/4$ , is aperiodic. Interpolating  $T$  linearly, the conditioned Donsker theorem gives

$$\left( \frac{T_{nt}}{\sqrt{n/2}} : 0 \leq t \leq 1 \right) \text{ under } \mathbb{P}(\cdot \mid T_n = 0) \Rightarrow b.$$

On each two-step interval, the two interpolations satisfy  $|S'_{2nt} - 2T_{nt}| \leq 1$ . After division by  $\sqrt{n}$ , this difference tends uniformly to zero. Since  $T_n = 0$  is equivalent to  $S'_{2n} = 0$ , Slutsky's lemma gives the claimed limit  $\sqrt{2} b$ .

\* \* \*

**Solution to Exercise 5.9.** [statement p. 181]

(1) For  $k \in \mathbb{Z}$ , write:

$$\mu_n(k) = \mathbb{P}(X_1 = k | S_n = 0) = \mathbb{P}(X_1 = k) \cdot \frac{\mathbb{P}(S_{n-1} = -k)}{\mathbb{P}(S_n = 0)}.$$

By the local limit theorem, the numerator and denominator are both asymptotic to  $\frac{1}{\sqrt{2\pi\sigma^2 n}}$ , so that

$$\mu_n(k) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X_1 = k).$$

Thus  $\mu_n$  converges weakly to the law of  $X_1$ .

(2) Yes, she is right. Let us check that  $(Y_n)_{n \geq 1}$  is bounded in  $L^2$ , which implies the desired result. We have

$$\mathbb{E}[Y_n^2] = \sum_{k \in \mathbb{Z}} k^2 \mathbb{P}(X_1 = k) \cdot \frac{\mathbb{P}(S_{n-1} = -k)}{\mathbb{P}(S_n = 0)}.$$

By the uniform local limit theorem, there exists  $C > 0$  such that, for every  $n \geq 2$  and  $k \in \mathbb{Z}$ ,

$$\frac{\mathbb{P}(S_{n-1} = -k)}{\mathbb{P}(S_n = 0)} \leq C.$$

Indeed, the numerator is uniformly  $O(n^{-1/2})$  and the denominator is asymptotic to  $(2\pi\sigma^2 n)^{-1/2}$ . We absorb the initial values of  $n$  into  $C$  since  $\mathbb{P}(S_n = 0) > 0$ . Thus  $\mathbb{E}[Y_n^2] \leq C \mathbb{E}[X_1^2]$  for  $n \geq 2$ . Finally,  $Y_1 = 0$  almost surely. The sequence is therefore bounded in  $L^2$ , and consequently uniformly integrable.

\* \* \*

**Solution to Problem 5.10.** [statement p. 182]

(1) This is a variant of the converging-together principle. We show that, for every closed set  $F$ ,

$$\limsup_{n \rightarrow \infty} \mathbb{P}(I_n \in F) \leq \mathbb{P}(I \in F). \quad (10)$$

By Markov's inequality, for  $\eta > 0$ , we have  $\mathbb{P}(|I_n(\delta) - I_n| \geq \eta) \leq \frac{\delta}{\eta}$ . Thus, writing  $F^\eta$  for the closed  $\eta$ -neighborhood of  $F$ , we have

$$\limsup_{n \rightarrow \infty} \mathbb{P}(I_n \in F) \leq \limsup_{n \rightarrow \infty} \mathbb{P}(I_n(\delta) \in F^\eta) + \frac{\delta}{\eta} \leq \mathbb{P}(I(\delta) \in F^\eta) + \frac{\delta}{\eta}$$

by convergence in distribution of  $I_n(\delta)$  to  $I(\delta)$ . Letting  $\delta \rightarrow 0$  and then  $\eta \rightarrow 0$  yields the desired result (10).

*Alternative solution.* Alternatively, use the characterization of convergence in distribution by bounded Lipschitz functions (the portmanteau theorem).

(2) Apply Fubini's theorem to  $(\omega, t) \mapsto \mathbb{1}_{B_t=\delta}$ :

$$\mathbb{E} \left[ \int_0^1 \mathbb{1}_{B_t=\delta} dt \right] = \int_0^1 \mathbb{P}(B_t = \delta) dt = 0,$$

so  $\int_0^1 \mathbb{1}_{B_t=\delta} dt = 0$  a.s., which gives the result.

(3) For  $\delta > 0$ , set

$$g_\delta(x) = \begin{cases} |x|^{-\gamma}, & |x| > \delta, \\ 0, & |x| \leq \delta. \end{cases}$$

We have the exact identity

$$\frac{1}{n^{1-\gamma/2}} \sum_{i=1}^n |S_i|^{-\gamma} \mathbb{1}_{|S_i| > \delta\sqrt{n}} = \int_0^1 g_\delta \left( \frac{S_{[nt]}}{\sqrt{n}} \right) dt. \quad (11)$$

By Donsker's theorem and the Skorokhod representation theorem, we can realize the interpolated walks  $S_{nt}/\sqrt{n}$  so that they converge uniformly almost surely on  $[0, 1]$  to  $B$ . Their step-function versions  $S_{[nt]}/\sqrt{n}$  then also converge uniformly to  $B$ , by uniform continuity of the limiting path. The only discontinuities of  $g_\delta$  are  $\delta$  and  $-\delta$ . By (2), also applied to  $-B$ , these levels are visited on a set of times of Lebesgue measure zero, almost surely. Finally,  $|g_\delta| \leq \delta^{-\gamma}$ . Dominated convergence therefore gives the claimed convergence.

(4) We show that

$$\mathbb{E} \left[ \int_0^1 \frac{1}{|B_t|^\gamma} dt \right] < \infty \quad (12)$$

which will imply that  $\int_0^1 \frac{1}{|B_t|^\gamma} dt < \infty$  a.s. By Fubini's theorem and the scaling property of Brownian motion,

$$\mathbb{E} \left[ \int_0^1 \frac{1}{|B_t|^\gamma} dt \right] = \int_0^1 \mathbb{E} \left[ \frac{1}{|B_t|^\gamma} \right] dt = \int_0^1 \frac{1}{t^{\gamma/2}} dt \cdot \mathbb{E} \left[ \frac{1}{|B_1|^\gamma} \right].$$

Since  $\gamma/2 < 1$ , it suffices to check that  $\mathbb{E}[1/|B_1|^\gamma] < \infty$ . For this, simply write

$$\mathbb{E} \left[ \frac{1}{|B_1|^\gamma} \right] = \int_{\mathbb{R}} \frac{1}{|x|^\gamma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx < \infty$$

because  $\gamma < 1$ , and (12) follows.

(5) Set

$$I_n = \frac{1}{n^{1-\gamma/2}} \sum_{i=1, S_i \neq 0}^n \frac{1}{|S_i|^\gamma}, \quad I_n(\delta) = \frac{1}{n^{1-\gamma/2}} \sum_{i=1}^n \frac{1}{|S_i|^\gamma} \mathbb{1}_{|S_i| > \delta\sqrt{n}}$$

and

$$I = \int_0^1 \frac{1}{|B_t|^\gamma} dt, \quad I(\delta) = \int_0^1 \frac{1}{|B_t|^\gamma} \mathbb{1}_{|B_t| > \delta} dt$$

and check that (1) applies.

Point (i) follows by monotone convergence, and point (ii) follows from question (3). The delicate point is (iii). To prove it, write:

$$\mathbb{E} [|I_n(\delta) - I_n|] = \frac{1}{n^{1-\gamma/2}} \sum_{i=1}^n \mathbb{E} \left[ \frac{1}{|S_i|^\gamma} \mathbb{1}_{0 < |S_i| \leq \delta\sqrt{n}} \right].$$

As a corollary of the local limit theorem, there exists a constant  $C > 0$  such that  $\mathbb{P}(S_i = k) \leq \frac{C}{\sqrt{i}}$  for every  $k \in \mathbb{Z}$  and  $i \geq 1$  (below,  $C$  is a constant depending on the distribution of the increments and on  $\gamma$ , which may change from line to line). Thus

$$\begin{aligned} \mathbb{E} \left[ \frac{1}{|S_i|^\gamma} \mathbb{1}_{0 < |S_i| \leq \delta\sqrt{n}} \right] &= \sum_{k=1}^{\lfloor \delta\sqrt{n} \rfloor} \frac{1}{k^\gamma} \mathbb{P}(|S_i| = k) \\ &\leq C \sum_{k=1}^{\lfloor \delta\sqrt{n} \rfloor} \frac{1}{k^\gamma} \cdot \frac{1}{\sqrt{i}} \\ &\leq C \cdot \delta^{1-\gamma} \cdot n^{1/2-\gamma/2} \cdot \frac{1}{\sqrt{i}}. \end{aligned}$$

Hence

$$\begin{aligned} \frac{1}{n^{1-\gamma/2}} \sum_{i=1}^n \mathbb{E} \left[ \frac{1}{|S_i|^\gamma} \mathbb{1}_{0 < |S_i| \leq \delta\sqrt{n}} \right] &\leq C \frac{\delta^{1-\gamma} \cdot n^{1/2-\gamma/2}}{n^{1-\gamma/2}} \sum_{i=1}^n \frac{1}{\sqrt{i}} \\ &\leq C \frac{\delta^{1-\gamma} \cdot n^{1/2-\gamma/2}}{n^{1-\gamma/2}} \cdot n^{1/2} \leq C \delta^{1-\gamma}. \end{aligned}$$

To apply (iii) exactly, set

$$r(\delta) = (\delta/C)^{1/(1-\gamma)}.$$

Replace  $I_n(\delta), I(\delta)$  by  $I_n(r(\delta)), I(r(\delta))$ . The resulting bound is

$$\mathbb{E} [|I_n(r(\delta)) - I_n|] \leq \delta,$$

and properties (i) and (ii) are preserved.

(\*) For  $\gamma = 1$ , we have

$$\int_0^1 \frac{dt}{|B_t|} = +\infty \quad \text{almost surely.}$$

This can be seen using Brownian local time, normalized by  $\int_0^1 h(B_t) dt = \int_{\mathbb{R}} h(x) L_1^x dx$ . Here we assume the usual properties:  $x \mapsto L_1^x$  is continuous and  $L_1^0 > 0$  almost surely. The occupation time formula and divergence of  $\int_0^\varepsilon dx/x$  then give the assertion.

The discrete quantity also diverges:

$$J_n := \frac{1}{\sqrt{n}} \sum_{\substack{1 \leq i \leq n \\ S_i \neq 0}} \frac{1}{|S_i|} \longrightarrow +\infty \quad \text{in probability.}$$

Indeed, for fixed  $\delta > 0$ , the proof of (3) remains valid with  $\gamma = 1$  and gives

$$J_n(\delta) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\mathbb{1}_{\{|S_i| > \delta\sqrt{n}\}}}{|S_i|} \Rightarrow J(\delta) := \int_0^1 \frac{\mathbb{1}_{\{|B_t| > \delta\}}}{|B_t|} dt,$$

with the convention in the statement for zero terms. Since  $J_n \geq J_n(\delta)$ , for every  $M > 0$ ,

$$\liminf_{n \rightarrow \infty} \mathbb{P}(J_n > M) \geq \mathbb{P}(J(\delta) > M).$$

Now  $J(\delta) \uparrow +\infty$  almost surely as  $\delta \downarrow 0$ . The right-hand side therefore tends to 1.

\* \* \*

### Solution to Problem 5.11. [statement p. 182]

- (1) If  $\sigma = 0$ , then  $Z = 0$  almost surely and the identity is immediate. We therefore assume that  $\sigma > 0$ . We have

$$\mathbb{E}[Z^4] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^4 e^{-x^2/(2\sigma^2)} dx.$$

Integration by parts gives

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^3 \cdot x e^{-\frac{x^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} 3x^2 \cdot \sigma^2 e^{-\frac{x^2}{2\sigma^2}} dx = 3\sigma^4.$$

- (2) Let  $H_n \rightarrow H$  with  $H \in [1/2, 1]$ . We show that  $\mathbb{P}_{H_n}$  converges weakly to  $\mathbb{P}_H$ . Suppose that  $W^n$  has distribution  $\mathbb{P}_{H_n}$ .

*Convergence of finite-dimensional distributions.* Let  $0 \leq t_1 < t_2 < \dots < t_k \leq 1$ . To prove convergence in distribution of  $((W_{t_1}^n, \dots, W_{t_k}^n))_{n \geq 1}$ , given the form of the characteristic functions of Gaussian vectors, Lévy's theorem shows that it suffices to prove convergence of the covariance matrices. We have

$$\mathbb{E}[W_{t_i}^n W_{t_j}^n] = \frac{1}{2} (t_i^{2H_n} + t_j^{2H_n} - |t_i - t_j|^{2H_n}) \xrightarrow{n \rightarrow \infty} \frac{1}{2} (t_i^{2H} + t_j^{2H} - |t_i - t_j|^{2H}),$$

so the finite-dimensional distributions of  $\mathbb{P}_{H_n}$  converge to those of  $\mathbb{P}_H$ .

*Tightness.* The increments are centered Gaussian and

$$\mathbb{E}[(W_s^n - W_t^n)^2] = |s - t|^{2H_n}.$$

By (1), for all  $n$  and  $s, t \in [0, 1]$ ,

$$\mathbb{E}[(W_s^n - W_t^n)^4] = 3|s - t|^{4H_n} \leq 3|s - t|^2,$$

because  $H_n \geq 1/2$ . Since  $W_0^n = 0$  almost surely, Kolmogorov's criterion gives tightness, with the same constants for all  $H_n \in [1/2, 1]$ . Convergence of the finite-dimensional distributions completes the proof.

- (3) Let  $N$  be a standard normal random variable. Set  $W_s = sN$  for  $0 \leq s \leq 1$ . Thus  $(W_s)_{0 \leq s \leq 1}$  is a Gaussian vector with covariance  $\mathbb{E}[W_s W_t] = st$ , whose law is therefore  $\mathbb{P}_1$ .
- (4) It suffices to check Kolmogorov's tightness criterion. We show that there exist  $C > 0$  and  $\eta > 0$  such that, for sufficiently large  $n$  and every  $s, t \in [0, 1]$ :

$$\mathbb{E} \left[ |Z_n(s) - Z_n(t)|^2 \right] \leq C|s - t|^{1+\eta}.$$

Choose  $C' > 0$  such that  $\mathbb{E}[S_n^2]^{1/2} \leq C'n^H$  for every  $n \geq 1$ . Let  $s \leq t \in [0, 1]$ .

*First case:*  $\lfloor ns \rfloor = \lfloor nt \rfloor = k$ . Then  $S_{ns} - S_{nt} = n(s - t)(S_{k+1} - S_k)$ . Since  $0 \leq n(t - s) \leq 1$ , we obtain

$$\mathbb{E} \left[ |Z_n(s) - Z_n(t)|^2 \right] = n^{2-2H}(t - s)^2 \mathbb{E} \left[ S_1^2 \right] \leq \mathbb{E} \left[ S_1^2 \right] (t - s)^{2H}.$$

The constant is independent of  $n, s, t$ .

*Second case:*  $\lfloor ns \rfloor < \lfloor nt \rfloor$ . Set  $k = \lfloor ns \rfloor$  and  $\ell = \lfloor nt \rfloor$ . Then, using the fact that

$$|S_{nt} - S_{ns}| \leq |S_{nt} - S_\ell| + |S_\ell - S_{k+1}| + |S_{k+1} - S_{ns}|,$$

the first case and the inequality  $x \leq x^H$  for  $x \in [0, 1]$ , we have

$$\begin{aligned} & \mathbb{E} \left[ (S_{ns} - S_{nt})^2 \right]^{1/2} \\ & \leq \mathbb{E} \left[ (S_{nt} - S_\ell)^2 \right]^{1/2} + \mathbb{E} \left[ (S_\ell - S_{k+1})^2 \right]^{1/2} + \mathbb{E} \left[ (S_{k+1} - S_{ns})^2 \right]^{1/2} \\ & \leq C(nt - \ell) + C'(\ell - (k + 1))^H + C(k + 1 - ns) \\ & \leq C(nt - \ell)^H + C'(\ell - (k + 1))^H + C(k + 1 - ns)^H \\ & \leq C(nt - ns)^H + C'(nt - ns)^H + C(nt - ns)^H \\ & \leq (2C + C')n^H(t - s)^H, \end{aligned}$$

which implies

$$\mathbb{E} \left[ |Z_n(s) - Z_n(t)|^2 \right] \leq C''(t - s)^{2H}.$$

Since  $2H > 1$ , Kolmogorov's tightness criterion is satisfied.

- (5) Assumptions (a) and (b) of question (4) are satisfied; its proof therefore gives tightness of  $(Z_n)$ . We will establish convergence of the finite-dimensional distributions without assuming in advance the existence of  $W$ . To do so, it suffices to show convergence of the covariance matrices of the corresponding Gaussian vectors. Fix  $0 \leq s < t \leq 1$  and use  $xy = (x^2 + y^2 - (x - y)^2)/2$ :

$$\begin{aligned} & \frac{1}{n^{2H}} \mathbb{E} [S_{\lfloor ns \rfloor} S_{\lfloor nt \rfloor}] \\ & = \frac{1}{2n^{2H}} \mathbb{E} [S_{\lfloor ns \rfloor}^2] + \frac{1}{2n^{2H}} \mathbb{E} [S_{\lfloor nt \rfloor}^2] - \frac{1}{2n^{2H}} \mathbb{E} [(S_{\lfloor nt \rfloor} - S_{\lfloor ns \rfloor})^2] \\ & = \frac{1}{2n^{2H}} \mathbb{E} [S_{\lfloor ns \rfloor}^2] + \frac{1}{2n^{2H}} \mathbb{E} [S_{\lfloor nt \rfloor}^2] - \frac{1}{2n^{2H}} \mathbb{E} [S_{\lfloor nt \rfloor - \lfloor ns \rfloor}^2] \\ & \rightarrow \frac{c}{2} (s^{2H} + t^{2H} - |s - t|^{2H}). \end{aligned}$$

To conclude, for every  $t \in [0, 1]$ ,

$$\frac{|S_{nt} - S_{[nt]}|}{n^H} \leq \frac{|S_{[nt]+1} - S_{[nt]}|}{n^H} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

Indeed, the increment in the numerator has the same distribution as  $S_1$  (also when  $[nt] = 0$ ). Slutsky's lemma therefore transfers these Gaussian limits to the interpolated processes. The limiting covariance matrices are positive semidefinite, being limits of covariance matrices. Tightness and convergence of all finite-dimensional distributions then give a limit in distribution  $Z$  taking values in  $\mathcal{C}$ . Setting  $W = Z/\sqrt{c}$  constructs a continuous, centered Gaussian process with covariance

$$\mathbb{E}[W_s W_t] = \frac{1}{2}(s^{2H} + t^{2H} - |s - t|^{2H}).$$

Its law is therefore  $\mathbb{P}_H$ . Applying this result in (9) will establish its existence for every  $H \in ]1/2, 1[$ .

(6) We have

$$\mathbb{E}[Y_i Y_{i+n}] = \mathbb{E}[(-1)^{X_{i+1} + \dots + X_{i+n}}] = \mathbb{E}[(-1)^{X_2}]^n = (p - (1-p))^n = (2p-1)^n.$$

(7) This follows from Donsker's theorem: when  $p = 1/2$ , the  $(Y_n)_{n \geq 1}$  are i.i.d. with  $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = 1/2$ .

(8) For fixed  $m \geq 1$ , the vectors

$$(Y_2^j, \dots, Y_{m+1}^j), \quad j \geq 1,$$

are i.i.d., centered and bounded. The multivariate central limit theorem therefore gives

$$\left( \frac{1}{\sqrt{k}} \sum_{j=1}^k Y_{i+1}^j : 1 \leq i \leq m \right) \Rightarrow \mathcal{N}(0, K_m),$$

where

$$(K_m)_{i\ell} = \mathbb{E}[Y_{i+1}^1 Y_{\ell+1}^1] = \mathbb{E}[(2p-1)^{|i-\ell|}].$$

These Gaussian laws are consistent as  $m$  varies; Kolmogorov's extension theorem provides a centered Gaussian process  $(G_i)_{i \geq 1}$  with these marginals. We have  $\mathbb{E}[G_i^2] = 1$  and, for  $n \geq 1$ ,

$$r(n) = \mathbb{E}[G_i G_{i+n}] = \mathbb{E}[(2p-1)^n] = (1-H)2^{3-2H} \int_{1/2}^1 (2u-1)^n (1-u)^{1-2H} du.$$

The asymptotic equivalent assumed in the statement gives the result.

- (9) Set  $r(0) = 1$ . The process  $(G_i)$  is centered Gaussian and its covariance  $r(|i - j|)$  depends only on the lag. Thus  $V_{n+i} - V_n$  has the same distribution as  $V_i$ . Moreover,

$$\mathbb{E} [V_n^2] = \sum_{i,j=1}^n r(|i - j|) = n + 2 \sum_{k=1}^{n-1} (n - k)r(k).$$

By (8),  $r(k) \sim c_H^{-2H}(2H - 1)k^{2H-2}$ . Since  $2H - 2 > -1$ , an integral comparison gives

$$\sum_{k=1}^{n-1} (n - k)k^{2H-2} \sim \frac{n^{2H}}{(2H - 1)2H}.$$

Consequently,

$$\mathbb{E} [V_n^2] \sim c_H^{-2} n^{2H},$$

the term  $n$  being negligible since  $2H > 1$ . The assumptions of (5) are therefore satisfied with  $c = c_H^{-2}$ , which gives convergence of  $c_H V_{nt}/n^H$  to  $W_t$ .

**Remark.** The process  $W$  is called *fractional Brownian motion*; this problem is inspired by Nathanaël Enriquez's paper "A simple construction of the fractional Brownian motion." *Stochastic Processes and their Applications* 109.2 (2004): 203-223.

## 6 The space of càdlàg functions

Related lecture notes: [Chapter 7](#), p. 68.

### 6.1 Exercises

**Exercise 6.1.** [solution p. 200] Write  $\mathbb{D} = \mathbb{D}([0, 1], \mathbb{R})$  for the set of real-valued càdlàg functions on  $[0, 1]$ , equipped with the Skorokhod  $J_1$  topology defined in the lectures. For a random variable  $Z$ , write  $F_Z$  for its distribution function defined by  $F_Z(t) = \mathbb{P}(Z \leq t)$  for  $t \in \mathbb{R}$ , viewed as an element of  $\mathbb{D}$  by restriction to  $[0, 1]$ .

Let  $(X_n)_{n \geq 1}$ ,  $X$  be real-valued random variables taking values in  $[\frac{1}{2}, 1]$ .

(1) Alix says: “if  $X_n$  converges in distribution, then  $F_{X_n}$  converges in  $\mathbb{D}$ .” Is he right? Justify your answer.

(2) Billie says: “if  $F_{X_n}$  converges in  $\mathbb{D}$ , then  $X_n$  converges in distribution.” Is she right? Justify your answer.

\* \* \*

**Problem 6.2.** [solution p. 201]

The three parts of this problem are independent, except for question (6), which uses results from earlier parts.

**Notation and review.** For a vector space  $E$  and  $A \subset E$ , write  $\text{co}(A)$  for the convex hull of  $A$ , defined by

$$\text{co}(A) = \left\{ \sum_{i=1}^n \lambda_i a_i : n \geq 1, a_i \in A \text{ and } \lambda_i \in [0, 1] \text{ for } 1 \leq i \leq n, \sum_{i=1}^n \lambda_i = 1 \right\}.$$

Write  $\mathbb{D} = \mathbb{D}([0, 1], \mathbb{R})$  for the space of càdlàg functions from  $[0, 1]$  to  $\mathbb{R}$ , equipped with the Skorokhod  $J_1$  topology. For  $x \in \mathbb{D}$  and  $0 < t \leq 1$ , write  $\Delta x(t) = x(t) - x(t-)$  and  $\Delta x(0) = 0$ . For  $0 < \delta < 1$  and  $x \in \mathbb{D}$ , write

$$\omega'(x, \delta) = \inf_{1 \leq i \leq n} \max_{s, t \in [t_{i-1}, t_i]} |x(s) - x(t)|,$$

where the infimum is taken over all partitions  $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = 1$  of  $[0, 1]$  such that  $\min_{1 \leq i \leq n} (t_i - t_{i-1}) > \delta$ .

For a metric space  $E$  and  $A \subset E$ , recall that  $A$  is called totally bounded if, for every  $\varepsilon > 0$ ,  $A$  can be covered by finitely many balls of radius  $\varepsilon$ , and  $A$  is called relatively compact if  $\overline{A}$  is compact. If  $E$  is complete, recall that a subset is totally bounded if and only if it is relatively compact.

Recall that a family  $\mathcal{F} \subset \mathbb{D}$  is relatively compact if and only if the following two conditions hold:

- $\sup_{x \in \mathcal{F}} \|x\|_\infty < \infty$
- $\forall \varepsilon > 0, \exists \delta > 0, \sup_{x \in \mathcal{F}} \omega'(x, \delta) \leq \varepsilon$ .

**Part One.** Let  $(E, \|\cdot\|)$  be a complete separable normed vector space.

- (1) If  $K \subset E$  is compact, show that the convex hull of  $K$ , denoted by  $\text{co}(K)$ , is totally bounded.
- (2) Show that every probability measure on  $E$  is convex-tight, meaning that, for every  $\varepsilon > 0$ , there exists a compact convex set  $K$  such that  $\mu(K) \geq 1 - \varepsilon$ .

**Part Two.**

- (3) Consider

$$\mathcal{F}_0 = \left\{ \mathbb{1}_{[s,1]} : \frac{1}{4} \leq s \leq \frac{3}{4} \right\}.$$

Justify that  $\mathcal{F}_0$  is relatively compact in  $\mathbb{D}$  and  $\text{co}(\mathcal{F}_0)$  is not relatively compact in  $\mathbb{D}$ .

For  $\mathcal{F} \subset \mathbb{D}$  and  $\varepsilon > 0$ , write

$$S_\varepsilon(\mathcal{F}) = \left\{ t \in [0, 1] : \sup_{x \in \mathcal{F}} |\Delta x(t)| > \varepsilon \right\}.$$

- (4) Assume that  $K \subset \mathbb{D}$  is relatively compact. Show that if  $\text{co}(K)$  is relatively compact, then for every  $\varepsilon > 0$ ,  $\text{Card}(S_\varepsilon(K)) < \infty$ .

**Part Three.** Let  $X$  be a random variable taking values in  $\mathbb{D}$ . We say that  $X$  is continuous in probability if, for every sequence  $t_n \rightarrow t$ , the sequence  $(X(t_n))_{n \geq 1}$  converges in probability to  $X(t)$  as  $n \rightarrow \infty$ .

- (5) Give an example where  $\forall \omega \in \Omega$ ,  $X_\omega$  is not continuous but  $X$  is continuous in probability.
- (6) Assume that  $X$  is continuous in probability and its law is convex-tight. Prove that  $X$  is a.s. continuous.
- (7) Is every probability measure on  $\mathbb{D}$  convex-tight? Does this contradict (2)?

*Bonus question:* Assume that  $K \subset \mathbb{D}$  is relatively compact. Show that if  $\text{Card}(S_\varepsilon(K)) < \infty$  for every  $\varepsilon > 0$ , then  $\text{co}(K)$  is relatively compact.

\* \* \*

**Problem 6.3.** [solution p. 203] Write  $\mathcal{C}$  for the set of continuous functions from  $[0, 1]$  to  $\mathbb{R}$ , equipped with the topology of uniform convergence, and  $\mathbb{D}$  for the set of càdlàg functions from  $[0, 1]$  to  $\mathbb{R}$ , equipped with the Skorokhod topology introduced in the lectures. Let  $(U_i)_{i \geq 1}$  be i.i.d. random variables uniform on  $[0, 1]$ . For  $n \geq 1$  and  $t \in [0, 1]$ , set:

$$F_n(t) = \frac{1}{n} \text{Card}(\{i \in \{1, 2, \dots, n\} : U_i \leq t\}), \quad B_n(t) = \sqrt{n}(F_n(t) - t).$$

The aim of this problem is to show that  $B_n$  converges in distribution in  $\mathbb{D}$  to the Brownian bridge.

We say that  $(M_1, \dots, M_n)$  has the multinomial distribution  $M(k, p_1, \dots, p_n)$  if  $k \geq 1$ ,  $p_i \geq 0$ ,  $\sum_{i=1}^n p_i = 1$ , and, for all integers  $m_i \geq 0$  such that  $\sum_{i=1}^n m_i = k$ ,

$$\mathbb{P}((M_1, \dots, M_n) = (m_1, \dots, m_n)) = \frac{k!}{\prod_{i=1}^n m_i!} \prod_{i=1}^n p_i^{m_i}.$$

This probability is 0 elsewhere; we use the convention  $0^0 = 1$ .

For  $1 \leq j \leq n$ , set:

$$N_j = \text{Card}(\{i \in \{1, 2, \dots, n\} : U_i \in [(j-1)/n, j/n]\}).$$

(1) Prove that  $(N_1, \dots, N_n)$  has a multinomial distribution and identify its parameters.

Let  $(P_i)_{i \geq 1}$  be i.i.d. Poisson random variables with parameter 1. For  $0 \leq k \leq n$ , set:

$$S_k = \sum_{i=1}^k (P_i - 1)$$

and define  $S_t$  for  $t \geq 0$  by linear interpolation. Finally, let  $\widehat{B}_n \in \mathcal{C}$  be the function obtained by linearly interpolating  $B_n$  at the points  $\{k/n, 0 \leq k \leq n\}$ , and let  $\widehat{S}_n \in \mathcal{C}$  be a random variable whose law is the conditional law of  $(\frac{1}{\sqrt{n}} S_{nt})_{0 \leq t \leq 1}$  given  $S_n = 0$ .

(2) Prove that  $\widehat{B}_n$  and  $\widehat{S}_n$  have the same law.

(3) Justify that  $\widehat{B}_n$  converges in distribution in  $\mathbb{D}$  to the Brownian bridge.

(4) Prove that  $\sup_{0 \leq t \leq 1} |\widehat{B}_n(t) - B_n(t)| \rightarrow 0$  in probability.

(5) Conclude.

## 6.2 Solutions

### Solution to Exercise 6.1. [statement p. 198]

(1) No. For  $n \geq 4$ , take  $\mathcal{L}(X_n) = \frac{1}{2}\delta_{1/2+1/n} + \frac{1}{2}\delta_{1/2+2/n}$ , and set  $X_n = 1/2$  for  $n < 4$ . These variables do take values in  $[1/2, 1]$  and converge in distribution to  $1/2$ . If  $F_{X_n}$  converged in  $\mathbb{D}$  to  $F$ , convergence at continuity points, right-continuity and convergence at the terminal point 1 would force  $F = \mathbb{1}_{[1/2, 1]}$ . Yet, for every time change  $\lambda_n$ , there exists  $t_n$  such that  $\lambda_n(t_n) = 1/2 + 1/n$ , and

$$|F_{X_n}(\lambda_n(t_n)) - F(t_n)| = \frac{1}{2}.$$

Thus  $J_1$  convergence is impossible.

- (2) Yes. Let  $F$  be the limit in  $\mathbb{D}$  and choose time changes  $\lambda_n$  such that  $F_{X_n} \circ \lambda_n \rightarrow F$  uniformly. Since the functions  $F_{X_n} \circ \lambda_n$  are nondecreasing,  $F$  is nondecreasing. It is càdlàg, and convergence at the endpoints gives  $F(0) = 0$  and  $F(1) = 1$ . Extending  $F$  by 0 on  $] - \infty, 0[$  and by 1 on  $]1, \infty[$  therefore yields a distribution function, still denoted by  $F$ , of a probability law  $\nu$  on  $\mathbb{R}$ . The  $J_1$  convergence gives  $F_{X_n}(t) \rightarrow F(t)$  at every continuity point of  $F$  in  $[0, 1]$ ; outside this interval, convergence is immediate. The convergence criterion for distribution functions therefore implies  $\mathcal{L}(X_n) \Rightarrow \nu$ .

\* \* \*

**Solution to Problem 6.2.** [statement p. 198]

- (1) Let  $\varepsilon > 0$ . Choose  $x_1, \dots, x_N \in K$  such that  $K \subset \bigcup_{i=1}^N B(x_i, \varepsilon)$ . Every convex combination of elements of  $K$  lies at distance strictly less than  $\varepsilon$  from  $P = \text{co}(\{x_1, \dots, x_N\})$ . Since the simplex of coefficients is compact,  $P$  is compact. Thus cover  $P$  by finitely many balls of radius  $\varepsilon$ : the same centers with radius  $2\varepsilon$  cover  $\text{co}(K)$ . Since  $\varepsilon$  is arbitrary,  $\text{co}(K)$  is totally bounded.
- (2) Let  $\varepsilon > 0$ . By Prokhorov's theorem, there exists a compact set  $C$  such that  $\mu(C) \geq 1 - \varepsilon$ . The desired result follows by taking  $K = \overline{\text{co}(C)}$ .
- (3) We indeed have  $\sup_{x \in \mathcal{F}_0} \|x\|_\infty < \infty$ . Moreover, for  $\varepsilon > 0$ , taking  $\delta < 1/4$  gives  $\sup_{x \in \mathcal{F}_0} \omega'(x, \delta) = 0$ . Thus  $\mathcal{F}_0$  is relatively compact. On the other hand, taking

$$x_n = \frac{1}{2} \mathbb{1}_{[1/2-1/n, 1]} + \frac{1}{2} \mathbb{1}_{[1/2+1/n, 1]} \in \text{co}(\mathcal{F}_0), \quad n \geq 4,$$

the sequence  $(x_n)$  has no convergent subsequence, so  $\text{co}(\mathcal{F}_0)$  is not relatively compact in  $\mathbb{D}$ .

- (4) Argue by contradiction. If  $S_\varepsilon(K)$  is infinite for some  $\varepsilon > 0$ , there exist distinct times  $u_n \rightarrow u_0$  and functions  $x_n \in K$  such that  $|\Delta x_n(u_n)| > \varepsilon$ . We can discard the possible term  $u_n = 1$ ; by convention,  $u_n \neq 0$ . By relative compactness of  $K$ , choose  $\delta_0 > 0$  such that

$$\sup_{x \in K} \omega'(x, \delta_0) < \frac{\varepsilon}{2}.$$

For each  $n$ , choose a partition with intervals of length strictly greater than  $\delta_0$  whose maximal oscillation for  $x_n$  is strictly less than  $\varepsilon/2$ . The jump at  $u_n$  forces this point to be an interior partition point. The oscillations of  $x_n$  on  $[u_n - \delta_0, u_n[$  and  $[u_n, u_n + \delta_0]$  are therefore less than  $\varepsilon/2$ .

Fix  $0 < \delta < \delta_0$  and choose  $m \neq n$  such that  $0 < |u_m - u_n| < \delta$ . Each point then lies in the interior of an interval of the partition chosen for the other function, so that

$$|\Delta x_m(u_n)| \leq \frac{\varepsilon}{2}, \quad |\Delta x_n(u_m)| \leq \frac{\varepsilon}{2}.$$

For  $z = (x_m + x_n)/2 \in \text{co}(K)$ , the reverse triangle inequality therefore gives

$$|\Delta z(u_m)| > \frac{\varepsilon}{4}, \quad |\Delta z(u_n)| > \frac{\varepsilon}{4}.$$

A partition with intervals of length strictly greater than  $\delta$  cannot contain both points among its partition points. At least one of these jumps lies in the interior of an interval, whose oscillation is greater than  $\varepsilon/4$ . Thus  $\omega'(z, \delta) \geq \varepsilon/4$ . For every  $0 < \delta < \delta_0$ , we have constructed such a  $z \in \text{co}(K)$ , contradicting relative compactness of  $\text{co}(K)$ .

- (5) Take  $\Omega = ]0, 1[$  equipped with Lebesgue measure and  $U(\omega) = \omega$ . Set  $X(t) = \mathbb{1}_{\{U \leq t\}}$ . Every path has a jump, so is discontinuous. For  $0 < \varepsilon \leq 1$ ,

$$\mathbb{P}(|X(t_n) - X(t)| \geq \varepsilon) \leq \mathbb{P}(U \in [\min(t_n, t), \max(t_n, t)]) = |t_n - t| \rightarrow 0.$$

For  $\varepsilon > 1$ , the probability is zero.

- (6) Let  $\eta > 0$  and let  $K$  be a compact convex set such that  $\mathbb{P}(X \in K) \geq 1 - \eta$ . For  $\varepsilon > 0$ , the set  $D_\varepsilon = S_\varepsilon(K)$  is finite by (4). On  $\{X \in K\}$ , we have  $|\Delta X(t)| \leq \varepsilon$  for every  $t \notin D_\varepsilon$ . For each fixed  $t > 0$ , as  $s$  tends to  $t$  from below,  $X(s) \rightarrow X(t-)$  almost surely and  $X(s) \rightarrow X(t)$  in probability. Thus  $\Delta X(t) = 0$  almost surely. This property holds simultaneously on the finite set  $D_\varepsilon$ . Consequently,

$$\mathbb{P}\left(\sup_{0 \leq t \leq 1} |\Delta X(t)| \leq \varepsilon\right) \geq 1 - \eta.$$

Let  $\varepsilon \downarrow 0$ , then  $\eta \downarrow 0$ . A càdlàg function without jumps is continuous, completing the proof.

- (7) Let  $X$  be the example in question (5). By questions (5) and (6), its law is not convex-tight. This does not contradict (2), since  $\mathbb{D}$  equipped with the Skorokhod  $J_1$  topology does not carry a normed vector space topology.

**Remark.** This problem is inspired by Remark 2.5 in

Basse-O'Connor, Andreas, and Jan Rosiński. "On the uniform convergence of random series in Skorokhod space and representations of cadlag infinitely divisible processes." The Annals of Probability 41.6 (2013): 4317-4341.

<https://arxiv.org/pdf/1111.1682.pdf>

**Bonus question.** Fix  $\varepsilon > 0$ . Relative compactness of  $K$  yields  $\delta > 0$  such that  $\sup_{x \in K} \omega'(x, \delta) < \varepsilon$ . Choose a common partition  $0 = s_0 < \dots < s_q = 1$  containing all points of  $S_\varepsilon(K)$  and whose intervals all have length strictly less than  $\delta$ . This is possible since  $S_\varepsilon(K)$  is finite.

For each  $x \in K$ , also choose a partition with intervals of length strictly greater than  $\delta$  on which the oscillations of  $x$  are less than  $\varepsilon$ . The open interval  $]s_j, s_{j+1}[$  contains at most

one partition point of this second partition. If it contains one, the corresponding jump has size at most  $\varepsilon$ , since this partition point does not belong to  $S_\varepsilon(K)$ . The oscillations on either side are less than  $\varepsilon$ , so

$$\sup_{x \in K} \omega(x, [s_j, s_{j+1}[) \leq 3\varepsilon.$$

The same bound holds for every convex combination of elements of  $K$ . For  $z \in \text{co}(K)$ , set

$$Tz = \sum_{j=0}^{q-1} z(s_j) \mathbb{1}_{[s_j, s_{j+1}[} + z(1) \mathbb{1}_{\{1\}}.$$

Then  $\|z - Tz\|_\infty \leq 3\varepsilon$ . Since  $\sup_{x \in K} \|x\|_\infty < \infty$ , the functions  $Tz$ ,  $z \in \text{co}(K)$ , form a bounded subset of a finite-dimensional space. They are therefore totally bounded for the uniform norm. Since  $\varepsilon$  is arbitrary,  $\text{co}(K)$  itself is totally bounded for this norm. Since the space  $(\mathbb{D}, \|\cdot\|_\infty)$  is complete, its uniform closure is compact. Since uniform convergence implies  $J_1$  convergence, the desired relative compactness follows.

See also Peter Z. Daffer and Robert L. Taylor, “Laws of Large Numbers for  $D[0, 1]$ ”, *Annals of Probability* 7(1), 85–95 (1979), Theorem 6.

\* \* \*

### Solution to Problem 6.3. [statement p. 199]

- (1) For nonnegative integers  $m_1, \dots, m_n$  such that  $m_1 + \dots + m_n = n$ , we have

$$\mathbb{P}((N_1, \dots, N_n) = (m_1, \dots, m_n)) = \binom{n}{m_1, \dots, m_n} \frac{1}{n^n} = \frac{n!}{\prod_{i=1}^n m_i!} \frac{1}{n^n}$$

so  $(N_1, \dots, N_n)$  has the multinomial distribution  $M(n, 1/n, \dots, 1/n)$ .

- (2) It suffices to check that the following two vectors have the same law:

$$(\widehat{B}_n(0), \widehat{B}_n(1/n), \dots, \widehat{B}_n(n/n))$$

and  $(\widehat{S}_n(0), \widehat{S}_n(1/n), \dots, \widehat{S}_n(n/n))$

Note that, for  $0 \leq k \leq n$ , we have

$$\widehat{B}_n(k/n) = \frac{1}{\sqrt{n}} \sum_{i=1}^k (N_i - 1), \quad \widehat{S}_n(k/n) \stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^k (P_i - 1) \quad \text{under } \mathbb{P}(\cdot | S_n = 0).$$

It therefore suffices to check that, under the conditional probability  $\mathbb{P}(\cdot | S_n = 0)$ , the vector  $(P_1, \dots, P_n)$  has the multinomial distribution

$$M(n, 1/n, \dots, 1/n).$$

By the conditioning, the support of its law is indeed the set of nonnegative integers  $m_1, \dots, m_n$  such that  $m_1 + \dots + m_n = n$ . For such integers, we have

$$\mathbb{P}((P_1, \dots, P_n) = (m_1, \dots, m_n) | S_n = 0) = \frac{1}{\mathbb{P}(S_n = 0)} \prod_{i=1}^n e^{-1} \frac{1}{m_i!}.$$

Now  $S_n + n$  has a Poisson distribution with parameter  $n$ , so  $\mathbb{P}(S_n = 0) = e^{-n} \frac{n^n}{n!}$ . Thus

$$\mathbb{P}((P_1, \dots, P_n) = (m_1, \dots, m_n) | S_n = 0) = \frac{n!}{\prod_{i=1}^n m_i!} \frac{1}{n^n},$$

which gives the result.

- (3) Since  $\mathbb{E}[P_1 - 1] = 0$  and  $\text{Var}(P_1 - 1) = 1$ , by the conditional Donsker theorem proved in the lectures,  $\widehat{S}_n$  converges in distribution in  $\mathcal{C}$  to the Brownian bridge. Since  $\widehat{B}_n$  and  $\widehat{S}_n$  have the same law, the same holds for  $\widehat{B}_n$ . Next, if  $f_n, f \in \mathcal{C}$  satisfy  $f_n \rightarrow f$  in  $\mathcal{C}$ , then  $f_n \rightarrow f$  in  $\mathbb{D}$  (take the time change to be the identity). In other words, the inclusion of  $\mathcal{C}$  into  $\mathbb{D}$  is continuous. We conclude by the continuous mapping theorem.

- (4) *First solution.* Begin by writing

$$\sup_{0 \leq t \leq 1} |\widehat{B}_n(t) - B_n(t)| \leq \frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} N_k.$$

Noting that the  $N_i$  have distributions  $\text{Bin}(n, 1/n)$ , we obtain, for every  $\varepsilon > 0$ :

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq k \leq n} N_k \geq \varepsilon \sqrt{n}\right) &\leq n \mathbb{P}(N_1 \geq \varepsilon \sqrt{n}) \\ &\leq n \mathbb{E}\left[e^{N_1}\right] e^{-\varepsilon \sqrt{n}} \\ &= n \left(1 + \frac{e-1}{n}\right)^n e^{-\varepsilon \sqrt{n}} \sim n e^{e-1} e^{-\varepsilon \sqrt{n}} \end{aligned}$$

which converges to 0 as  $n \rightarrow \infty$ .

*Second solution.* If  $\frac{i}{n} \leq t < \frac{i+1}{n}$  for  $0 \leq i \leq n-1$ , note that

$$B_n\left(\frac{i}{n}\right) - \frac{1}{\sqrt{n}} \leq B_n(t) \leq B_n\left(\frac{i+1}{n}\right) + \frac{1}{\sqrt{n}},$$

$$\min\{B_n(i/n), B_n((i+1)/n)\} \leq \widehat{B}_n(t) \leq \max\{B_n(i/n), B_n((i+1)/n)\}.$$

Since  $\widehat{B}_n(i/n) = B_n(i/n)$  for  $0 \leq i \leq n$ , we deduce that

$$\sup_{0 \leq t \leq 1} |\widehat{B}_n(t) - B_n(t)| \leq \max_{0 \leq i \leq n-1} \left| \widehat{B}_n\left(\frac{i+1}{n}\right) - \widehat{B}_n\left(\frac{i}{n}\right) \right| + \frac{1}{\sqrt{n}}.$$

Now  $\widehat{B}_n$  converges in distribution in  $\mathcal{C}$  to the Brownian bridge. Thus, for every  $\varepsilon, \eta > 0$ , there exists  $\delta > 0$  such that, for all sufficiently large  $n$ ,

$$\mathbb{P}\left(\omega(\widehat{B}_n, \delta) > \eta\right) \leq \varepsilon.$$

We deduce that, for  $n > 1/\delta$  sufficiently large, we have

$$\max_{0 \leq i \leq n-1} \left| \widehat{B}_n\left(\frac{i+1}{n}\right) - \widehat{B}_n\left(\frac{i}{n}\right) \right| \leq \eta$$

with probability at least  $1 - \varepsilon$ . This completes the proof.

(5) First, write  $d_S$  for the Skorokhod metric

$$d_S(f, g) = \inf_{\lambda \in \Lambda} \max(\|\lambda - \text{Id}\|_\infty, \|f \circ \lambda - g\|_\infty)$$

where  $\Lambda$  is the set of time changes. Taking  $\lambda = \text{Id}$  directly gives

$$d_S(\widehat{B}_n, B_n) \leq \|\widehat{B}_n - B_n\|_\infty \xrightarrow{\mathbb{P}} 0$$

by (4).

Thus,  $\widehat{B}_n$  converges in distribution to the Brownian bridge in  $\mathbb{D}$  and  $d_S(\widehat{B}_n, B_n)$  converges in probability to 0. We deduce that  $B_n$  converges in distribution to the Brownian bridge.

**Remark.** This problem is inspired by the following article:

Jean-François Marckert. "One more approach to the convergence of the empirical process to the Brownian bridge." *Electronic Journal of Statistics*, 2 118-126 2008

<https://arxiv.org/pdf/0710.3296.pdf>

## 7 Poisson random measures

Related lecture notes: [Chapter 8](#), p. 81.

### 7.1 Exercises

**Exercise 7.1.** [solution p. 208] Let  $M$  be a Poisson random measure with intensity  $dx dy$  on  $\mathbb{R}^2$ . Write  $M_\theta$  and  $M_R$  for the Poisson random measures on  $\mathbb{R}/2\pi\mathbb{Z}$  and  $\mathbb{R}_+$  obtained from  $M$  by considering the angles and distances from the origin of the points of  $M$ . Give the intensities of these two random measures. Are they independent?

\* \* \*

**Exercise 7.2.** [solution p. 209] Let  $M$  be a Poisson random measure on  $\mathbb{R}_+$  with intensity  $\lambda(t)dt$ , where  $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a continuous function such that  $\int_0^\infty \lambda(t)dt = \infty$ .

- (1) Justify that  $M$  can be written as  $M = \sum_{k=1}^\infty \delta_{T_k}$ , where  $0 < T_1 < T_2 < \dots$  are random variables.
- (2) Set  $a(t) = \int_0^t \lambda(s)ds$ . Show that  $M' = \sum_{k \geq 1} \delta_{a(T_k)}$  is a Poisson random measure on  $\mathbb{R}_+$  with intensity  $dx$ .

From now on, assume that  $\lambda(t) = \frac{t}{1+t}$  for  $t \geq 0$ .

- (3) Compute the law of  $T_1$  and that of  $(T_1, T_2)$ .
- (4) Show that almost surely

$$\sum_{n \geq 1} \frac{1}{T_n^2} < \infty \quad \text{and} \quad \sum_{n \geq 1} \frac{1}{T_n} = \infty.$$

- (5) Show that  $T_{n+1} - T_n$  converges in distribution as  $n \rightarrow \infty$  and identify the limiting law.

\* \* \*

**Exercise 7.3.** [solution p. 210] Let  $\mu$  be a probability measure on  $\mathbb{R}$  and  $N$  a Poisson random measure with intensity  $dx \otimes \mu$  on  $\mathbb{R} \times \mathbb{R}$ . Write  $N = \sum_{n \in \mathbb{Z}} \delta_{(X_n, V_n)}$  with

$$\dots < X_{-k} < \dots < X_0 < 0 < X_1 < \dots < X_\ell < \dots,$$

and view  $N$  as a distribution of particles on the real line, each particle  $n$  initially at position  $X_n$  with velocity  $V_n$ .

- (a) Set  $N_t = \sum_{n \in \mathbb{Z}} \delta_{(X_n + tV_n, V_n)}$ . Show that  $N_t$  is still a Poisson random measure with intensity  $dx \otimes \mu$ .

(b) Assume  $\int_{\mathbb{R}} |v| \mu(dv) < \infty$ . For each particle with velocity  $V_n \neq 0$ , let

$$T_n = -\frac{X_n}{V_n}$$

be its hitting time of 0, which may be negative. Particles with zero velocity do not contribute to arrivals. Show that

$$\sum_{n \in \mathbb{Z}: V_n \neq 0} \delta_{(T_n, V_n)}$$

is a Poisson random measure on  $\mathbb{R} \times \mathbb{R}$  and determine its intensity.

(c) Compute  $\mathbb{P}(\forall s \in [0, t], \forall n \in \mathbb{Z}, X_n + sV_n \notin [a, b])$ , where  $[a, b]$  is a fixed interval, and determine 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 7.4.** [solution p. 211] Let  $M$  be a Poisson random measure on  $\mathbb{R}^2$  with intensity  $\lambda dx dy$  for  $\lambda > 0$ , written as  $M = \sum_{i \geq 1} \delta_{X_i}$ , where the enumeration  $(X_i)_{i \geq 1}$  is a measurable function of  $M$  fixed in advance.

- (1) Compute the law of  $\inf_{i \geq 1} |X_i|$ .
- (2) Let  $(Y_n)_{n \geq 1}$  be i.i.d. random variables taking values in  $\mathbb{R}^2$ , independent of  $M$ . Show that  $M' = \sum_{i \geq 1} \delta_{X_i + Y_i}$  has the same law as  $M$ .
- (3) Now let  $(R_n)_{n \geq 1}$  be i.i.d. nonnegative real-valued random variables independent of  $M$ . Consider the random set

$$A = \bigcup_{i \geq 1} B(X_i, R_i),$$

where  $B(x, r)$  is the ball with center  $x$  and radius  $r$  in  $\mathbb{R}^2$ .

- (a) Compute the law of the number of balls in  $A$  covering a given point  $x \in \mathbb{R}^2$ .
  - (b) Show that if  $\mathbb{E}[R_1^2] = \infty$ , then for every  $x \in \mathbb{R}^2$ , almost surely  $x \in A$ .
  - (c) Deduce that if  $\mathbb{E}[R_1^2] = \infty$ , then  $A = \mathbb{R}^2$  almost surely.
  - (d) Show that if  $\mathbb{E}[R_1^2] < \infty$ , then  $A \neq \mathbb{R}^2$  almost surely.
- (4) Assume that  $R_i = r > 0$ . Write  $\mathcal{C}(0)$  for the cluster of the origin, that is, the points accessible from the origin by a path that stays in  $A$ .
- (a) Write  $\mathbb{P}_{\lambda, r}$  for the law of  $A$  with intensity  $\lambda > 0$  and radius  $r > 0$ . Show that

$$\begin{aligned} \mathbb{P}_{\lambda, r}(\mathcal{C}(0) \text{ unbounded}) &= \mathbb{P}_{1, \sqrt{\lambda}r}(\mathcal{C}(0) \text{ unbounded}) \\ &= \mathbb{P}_{\lambda r^2, 1}(\mathcal{C}(0) \text{ unbounded}). \end{aligned}$$

- (b) Show that  $\mathbb{P}_{\lambda, r}(\mathcal{C}(0) \text{ unbounded})$  is nondecreasing in  $\lambda$  and in  $r$ .

**Exercise 7.5.** [solution p. 213] **(Rényi's theorem)** Let  $M$  be a random variable taking values in  $M_{\text{atom}}^s(\mathbb{R}_+)$  (the subset of measures  $\mu \in M_{\text{atom}}(\mathbb{R}_+)$  whose atoms are pairwise distinct), almost surely finite on compact sets. Assume that, for every Borel subset  $A$  of  $\mathbb{R}_+$ ,

$$\mathbb{P}(M(A) = 0) = e^{-\lambda(A)},$$

where  $\lambda$  is Lebesgue measure on  $\mathbb{R}_+$ .

- (1) Let  $I_1, \dots, I_m$  be pairwise disjoint intervals. Show that the events  $\{M(I_j) = 0 : 1 \leq j \leq m\}$  are independent.
- (2) For all  $n, k \geq 0$ , set  $D_{n,k} = [k2^{-n}, (k+1)2^{-n}[$ . Let  $J$  be a bounded subinterval of  $\mathbb{R}_+$  with nonempty interior. For  $n \geq 0$ , write  $K_n = \text{Card}\{k \geq 0 : D_{n,k} \subseteq J\}$  and

$$M_n(J) = \sum_{k \geq 0 : D_{n,k} \subseteq J} \mathbb{1}_{\{M(D_{n,k}) \neq 0\}}.$$

Determine the law of  $M_n(J)$  in terms of  $n$  and  $K_n$ .

- (3) Show that  $M_n(J) \uparrow M(J)$  almost surely as  $n \rightarrow \infty$  (consider the atoms of  $M$  in  $J$  and first check that they all differ from  $\inf J$  and  $\sup J$ ), and deduce the law of  $M(J)$ .
- (4) Let  $J_1, \dots, J_k$  be pairwise disjoint bounded subintervals of  $\mathbb{R}_+$ . Show that the random variables  $M(J_1), \dots, M(J_k)$  are independent.
- (5) Write  $N_t = M([0, t])$ ,  $t \geq 0$ , for the counting process associated with  $M$ . Show that  $N$  is a Poisson process with intensity 1. Deduce that  $M$  is a Poisson random measure with intensity  $\lambda(dx)$ .

\* \* \*

**Exercise 7.6.** [solution p. 214] Find a sigma-algebra  $\mathcal{E}$  on  $[0, 1]$  and a probability measure  $\mu$  on  $([0, 1], \mathcal{E})$  such that  $\mu(B) \in \{0, 1\}$  for every  $B \in \mathcal{E}$  but  $\mu$  is not a Dirac measure.

## 7.2 Solutions

**Solution to Exercise 7.1.** [statement p. 206] The point 0 almost surely does not belong to the point cloud. By the mapping theorem and a change to polar coordinates, the desired intensities are

$$\mu_R(dr) = 2\pi r dr, \quad \mu_\theta(B) = \int_B \int_0^\infty r dr d\theta = \begin{cases} 0 & \text{if } \lambda(B) = 0, \\ \infty & \text{if } \lambda(B) > 0, \end{cases}$$

where  $\lambda$  denotes length measure on  $\mathbb{R}/2\pi\mathbb{Z}$ . The measure  $\mu_\theta$  is  $s$ -finite, since it equals  $\sum_{j \geq 1} \lambda$ , but it is not  $\sigma$ -finite.

The two random measures are independent. Indeed, for each fixed Borel set  $B$ ,  $M_\theta(B)$  is almost surely equal to 0 or  $\infty$ , according as  $\lambda(B) = 0$  or  $\lambda(B) > 0$ . All events generating  $\sigma(M_\theta)$  therefore have probability 0 or 1. Since events of probability 0 or 1 form a sigma-algebra, the same holds for every event in  $\sigma(M_\theta)$ . Such a sigma-algebra is independent of every other sigma-algebra, in particular of  $\sigma(M_R)$ . Here we use the sigma-algebra on random measures defined in the lectures, generated by the evaluation maps.

\* \* \*

### Solution to Exercise 7.2. [statement p. 206]

- (1) The intensity is atomless and finite on compact sets. The point cloud is therefore almost surely simple, finite on every compact set and has no point at 0. Moreover,  $M(\mathbb{R}_+) = \infty$  almost surely, since its total intensity is infinite. Its atoms can thus be arranged in increasing order. The variables

$$T_k = \inf\{t \geq 0 : M([0, t]) \geq k\}$$

are measurable, since  $\{T_k \leq t\} = \{M([0, t]) \geq k\}$ , and satisfy almost surely  $0 < T_1 < T_2 < \dots$  and  $T_k \rightarrow \infty$ .

- (2) By the mapping theorem, it suffices to show that the image of  $\lambda(t)dt$  under  $a$  is Lebesgue measure. The map  $a$  is continuous and nondecreasing,  $a(0) = 0$  and  $a(t) \rightarrow \infty$ . For  $x \geq 0$ , set

$$b(x) = \sup\{t \geq 0 : a(t) \leq x\}.$$

Then  $b(x) < \infty$ ,  $a(b(x)) = x$  and  $a^{-1}([0, x]) = [0, b(x)]$ . Consequently,

$$\int_{a^{-1}([0, x])} \lambda(t) dt = a(b(x)) = x.$$

These equalities characterize Lebesgue measure. This argument remains valid when  $a$  is not strictly increasing.

- (3) We have  $a(t) = t - \log(1 + t)$ . For  $t \geq 0$ ,

$$\mathbb{P}(T_1 > t) = \mathbb{P}(M([0, t]) = 0) = e^{-a(t)} = (1 + t)e^{-t}.$$

Thus  $T_1$  has density  $te^{-t}\mathbb{1}_{\{t>0\}}$ . By (2), the variables  $S_k = a(T_k)$  are the successive arrival times of a Poisson process with intensity 1. In particular,  $(S_1, S_2)$  has density  $e^{-v}\mathbb{1}_{\{0 < u < v\}}$  at  $(u, v)$ . The change of variables  $(u, v) = (a(s), a(t))$  gives the density of  $(T_1, T_2)$ :

$$\lambda(s)\lambda(t)e^{-a(t)}\mathbb{1}_{\{0 < s < t\}} = \frac{st}{1+s}e^{-t}\mathbb{1}_{\{0 < s < t\}}.$$

- (4) Set  $S_0 = 0$  and  $E_k = S_k - S_{k-1}$  for  $k \geq 1$ . The variables  $(E_k)$  are i.i.d. with exponential distribution of parameter 1. The strong law of large numbers gives  $S_n/n \rightarrow 1$  almost surely. Since  $T_n \rightarrow \infty$  and  $a(t)/t \rightarrow 1$ , we deduce

$$\frac{T_n}{n} = \frac{T_n}{a(T_n)} \frac{S_n}{n} \rightarrow 1 \quad \text{almost surely.}$$

The two assertions then follow by comparison with  $\sum_n n^{-2}$  and  $\sum_n n^{-1}$ .

- (5) Set  $G_n = T_{n+1} - T_n$  and  $E_{n+1} = a(T_{n+1}) - a(T_n)$ , which has distribution  $\text{Exp}(1)$ . Since  $a'(t) = t/(1+t)$  is nondecreasing and bounded above by 1,

$$\frac{T_n}{1+T_n} G_n \leq E_{n+1} \leq G_n.$$

Thus

$$0 \leq G_n - E_{n+1} \leq \frac{E_{n+1}}{T_n} \xrightarrow{\mathbb{P}} 0,$$

since  $(E_{n+1})$  is tight and  $T_n \rightarrow \infty$  almost surely. Slutsky's lemma therefore gives  $T_{n+1} - T_n \xrightarrow{d} \text{Exp}(1)$ .

\* \* \*

### Solution to Exercise 7.3. [statement p. 206]

- (a) The random measure  $N_t$  is the image of  $N$  under  $(x, v) \mapsto (x + tv, v)$ . For every nonnegative measurable function  $f$ , the change of variables  $y = x + tv$  gives

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x + tv, v) dx \mu(dv) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(y, v) dy \mu(dv).$$

The intensity is therefore unchanged, and the mapping theorem completes the proof.

- (b) First restrict  $N$  to  $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$ , then apply the transformation  $(x, v) \mapsto (-x/v, v)$ . The restriction and mapping properties show that the resulting measure is Poisson. The change of variables  $s = -x/v$  gives, for measurable  $f \geq 0$ ,

$$\int_{\mathbb{R} \setminus \{0\}} \int_{\mathbb{R}} f(-x/v, v) dx \mu(dv) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(s, v) |v| ds \mu(dv).$$

The desired intensity is therefore  $ds \otimes (|v| \mu(dv))$ . In particular, the arrival times form a Poisson random measure with intensity  $m_1 ds$ , where  $m_1 = \int_{\mathbb{R}} |v| \mu(dv)$ .

- (c) Set  $\ell = b - a$  and

$$A_t = \{(x, v) : \text{there exists } s \in [0, t] \text{ such that } x + sv \in [a, b]\}.$$

For a velocity  $v \geq 0$ , the corresponding initial positions form the interval  $[a - tv, b]$ ; for  $v < 0$ , they form  $[a, b - tv]$ . In both cases, its length is  $\ell + t|v|$ . Thus, for  $t > 0$ ,

$$\mathbb{P}(\forall s \in [0, t], \forall n \in \mathbb{Z}, X_n + sV_n \notin [a, b]) = \mathbb{P}(N(A_t) = 0) = e^{-\ell - tm_1}.$$

For  $t = 0$ , this probability equals  $e^{-\ell}$ .

Write  $H$  for the time defined in the statement, with the convention  $\inf \emptyset = \infty$ . If  $m_1 < \infty$ , only finitely many particles visit  $[a, b]$  during a bounded time interval. The first visit time is therefore attained whenever it is finite, and

$$\mathbb{P}(H > t) = e^{-\ell - tm_1}, \quad t \geq 0.$$

For  $0 < m_1 < \infty$ , we obtain

$$\mathcal{L}(H) = (1 - e^{-\ell})\delta_0 + e^{-\ell} \text{Exp}(m_1).$$

If  $m_1 = 0$ , all particles are stationary, and the law on  $[0, \infty]$  is  $(1 - e^{-\ell})\delta_0 + e^{-\ell}\delta_\infty$ .

If the integrability assumption in (b) is dropped and  $m_1 = \infty$ , the probability of no visit is 0 for every  $t > 0$ . Applying this to the times  $1/k$  gives  $H = 0$  almost surely, even though the interval may contain no particles at time 0.

\* \* \*

#### Solution to Exercise 7.4. [statement p. 207]

- (1) Set  $D = \inf_{i \geq 1} |X_i|$ . The point cloud is almost surely finite on compact sets. Thus, for  $t \geq 0$ ,

$$\mathbb{P}(D > t) = \mathbb{P}(M(\bar{B}(0, t)) = 0) = e^{-\lambda \pi t^2}.$$

In particular,  $D^2$  has an exponential distribution with parameter  $\lambda\pi$ , and  $D$  has density  $2\lambda\pi t e^{-\lambda\pi t^2} \mathbb{1}_{\{t>0\}}$ .

- (2) Write  $\nu$  for the law of  $Y_1$ . Since the enumeration  $(X_i)_{i \geq 1}$  is a measurable function of  $M$ , it is independent of  $(Y_i)_{i \geq 1}$ . By the marking theorem,  $\sum_{i \geq 1} \delta_{(X_i, Y_i)}$  is therefore a Poisson random measure with intensity  $\lambda dz \otimes \nu(dy)$ . Its image under  $(z, y) \mapsto z + y$  is  $M'$ . For every Borel set  $C \subset \mathbb{R}^2$ , the intensity of this image is, by Fubini and translation invariance of Lebesgue measure,

$$\lambda \int_{\mathbb{R}^2} \nu(dy) \int_{\mathbb{R}^2} \mathbb{1}_{\{z+y \in C\}} dz = \lambda \text{Area}(C).$$

The mapping theorem and uniqueness in law of Poisson random measures give the result.

- (3) Now write  $\rho$  for the law of  $R_1$ . Similarly,  $\widetilde{M} = \sum_{i \geq 1} \delta_{(X_i, R_i)}$  is a Poisson random measure with intensity  $\mu(dz, dr) = \lambda dz \rho(dr)$ .

- (a) For  $x \in \mathbb{R}^2$ , set  $H_x = \{(z, r) : x \in B(z, r)\}$ . The number of balls containing  $x$  is  $\widetilde{M}(H_x)$ , so has a Poisson distribution with parameter

$$\mu(H_x) = \lambda \int_0^\infty \pi r^2 \rho(dr) = \lambda \pi \mathbb{E}[R_1^2].$$

This formula also holds for an infinite parameter, in which case this number is almost surely infinite. It does not depend on the choice of open or closed balls, since their boundaries have Lebesgue measure zero.

(b) If  $\mathbb{E} [R_1^2] = \infty$ , for every fixed  $x$ ,  $\widetilde{M}(H_x) = \infty$  almost surely. In particular,  $\mathbb{P}(x \in A) = 1$ .

(c) Here we must obtain an event of probability 1 on which all points are covered simultaneously. Fix an integer  $m \geq 1$  and consider balls such that  $R_i > |X_i| + m$ . Each contains  $\overline{B}(0, m)$ . Their number has a Poisson distribution with parameter

$$\lambda \int_0^\infty \pi(r-m)_+^2 \rho(dr) = \infty.$$

Indeed,  $(r-m)_+^2 \geq r^2/4$  for  $r \geq 2m$ , and  $\mathbb{E} [R_1^2 \mathbb{1}_{\{R_1 \geq 2m\}}] = \infty$ . Thus  $\overline{B}(0, m) \subset A$  almost surely. Intersecting these events for  $m \geq 1$  gives  $A = \mathbb{R}^2$  almost surely.

(d) Assume  $a = \lambda \pi \mathbb{E} [R_1^2] < \infty$  and set  $p = e^{-a} > 0$ . By (a), a given point is not covered with probability  $p$ . To deduce the almost sure existence of an uncovered point, we control the dependence between two distant points.

For  $h \in \mathbb{R}^2$ , set

$$c(h) = \mu(H_0 \cap H_h) = \lambda \int_0^\infty \text{Area}(B(0, r) \cap B(h, r)) \rho(dr).$$

Two balls of radius  $r$  are disjoint when  $2r < |h|$ , so

$$0 \leq c(h) \leq \lambda \pi \mathbb{E} [R_1^2 \mathbb{1}_{\{2R_1 \geq |h|\}}] \longrightarrow 0 \quad (|h| \rightarrow \infty).$$

The probability that two points  $x, y$  are not covered is the probability that  $\widetilde{M}(H_x \cup H_y) = 0$ , namely

$$\mathbb{P}(x \notin A, y \notin A) = e^{-2a+c(y-x)}.$$

Thus, setting  $Z_j = \mathbb{1}_{\{(j,0) \notin A\}}$ , we have  $\mathbb{E} [Z_j] = p$  and

$$\text{Cov}(Z_i, Z_j) = p^2 (e^{c((j-i,0))} - 1).$$

These covariances are bounded and tend to 0 as  $|i-j| \rightarrow \infty$ . Consequently,

$$\text{Var} \left( \frac{1}{n} \sum_{j=1}^n Z_j \right) = \frac{p(1-p)}{n} + \frac{2p^2}{n^2} \sum_{h=1}^{n-1} (n-h) (e^{c((h,0))} - 1) \longrightarrow 0.$$

For the last convergence, it suffices to split the sum into  $h \leq H$  and  $h > H$ , then let first  $n$ , then  $H$ , tend to infinity. Chebyshev's inequality then gives

$$\mathbb{P}(Z_1 = \dots = Z_n = 0) \leq \frac{1}{p^2} \text{Var} \left( \frac{1}{n} \sum_{j=1}^n Z_j \right) \longrightarrow 0.$$

Almost surely, at least one of the points  $(j, 0)$  therefore does not belong to  $A$ , proving  $A \neq \mathbb{R}^2$ .

- (4) (a) For  $s > 0$ , the image of  $M$  under the dilation  $z \mapsto sz$  is a Poisson random measure with intensity  $(\lambda/s^2) dz$ . The image of  $A$  is the union of the balls  $B(sX_i, sr)$ , and the image of the cluster of the origin is the cluster of the origin in this set. A dilation preserves unboundedness. Consequently,

$$\mathbb{P}_{\lambda,r}(\mathcal{C}(0) \text{ unbounded}) = \mathbb{P}_{\lambda/s^2, sr}(\mathcal{C}(0) \text{ unbounded}).$$

Choosing  $s = \sqrt{\lambda}$  gives the first scaling relation; choosing  $s = 1/r$  gives the second.

- (b) For a fixed point cloud, if  $r \leq r'$ , then  $\bigcup_i B(X_i, r) \subset \bigcup_i B(X_i, r')$ . Every path in the first set is a path in the second, so the probability under consideration is nondecreasing in  $r$ .

If  $0 < \lambda \leq \lambda'$ , take two independent Poisson random measures  $M_1, M_2$  with respective intensities  $\lambda dz$  and  $(\lambda' - \lambda) dz$ . By superposition,  $M_1 + M_2$  has intensity  $\lambda' dz$ . The union of balls of radius  $r$  centered at points of  $M_1$  is contained in that associated with  $M_1 + M_2$ . The same argument using paths proves monotonicity in  $\lambda$ .

\* \* \*

#### Solution to Exercise 7.5. [statement p. 208]

- (1) For every subset  $A$  of  $\{1, \dots, m\}$ ,

$$\mathbb{P} \left( \bigcap_{j \in A} \{M(I_j) = 0\} \right) = \mathbb{P} \left( M \left( \bigcup_{j \in A} I_j \right) = 0 \right) = e^{-\sum_{j \in A} \lambda(I_j)} = \prod_{j \in A} \mathbb{P}(M(I_j) = 0).$$

This is exactly independence of the events under consideration.

- (2) By (1), the indicators being summed are independent, and each has a Bernoulli distribution with parameter  $1 - e^{-2^{-n}}$ . Consequently,

$$\mathcal{L}(M_n(J)) = \text{Bin}(K_n, 1 - e^{-2^{-n}}).$$

- (3) For each fixed  $x \geq 0$ ,  $\mathbb{P}(M(\{x\}) = 0) = 1$ . Thus there is almost surely no atom at  $\inf J$  or  $\sup J$ . Moreover, there are finitely many atoms in  $J$ , and they are distinct. For every realization satisfying these properties, as soon as  $n$  is sufficiently large, the dyadic cells containing these atoms are contained in  $J$  and each contains a single atom. Thus eventually  $M_n(J) = M(J)$ .

On the other hand, a dyadic cell contained in  $J$  splits into two cells of the next level, also contained in  $J$ . If it contains an atom, at least one of its two cells contains one. Thus  $M_n(J) \leq M_{n+1}(J)$  and  $M_n(J) \uparrow M(J)$  almost surely.

We have  $K_n 2^{-n} \rightarrow \lambda(J)$  and  $1 - e^{-2^{-n}} \sim 2^{-n}$ . For  $z \in [0, 1]$ , dominated convergence then gives

$$\mathbb{E} \left[ z^{M(J)} \right] = \lim_{n \rightarrow \infty} (1 - (1 - z)(1 - e^{-2^{-n}}))^{K_n} = e^{-\lambda(J)(1-z)}.$$

This is the probability generating function of a Poisson distribution with parameter  $\lambda(J)$ , so  $M(J) \sim \mathcal{P}(\lambda(J))$ .

- (4) For fixed  $n$ , the dyadic cells contained in the different  $J_i$  are pairwise disjoint. By (1), the variables  $M_n(J_1), \dots, M_n(J_k)$  are therefore independent. Letting  $n$  tend to infinity in their joint probability generating function and using (3), we obtain independence of  $M(J_1), \dots, M(J_k)$ . If one of the intervals reduces to a point or is empty, its mass is almost surely zero, and the conclusion remains valid.
- (5) Almost surely,  $N_0 = 0$  and the paths of  $N$  are nondecreasing, càdlàg, with jumps of size 1, by simplicity and local finiteness of  $M$ . For  $0 = t_0 < t_1 < \dots < t_k$ ,

$$N_{t_i} - N_{t_{i-1}} = M([t_{i-1}, t_i]), \quad 1 \leq i \leq k,$$

and, by (3) and (4), these variables are independent with distributions  $\mathcal{P}(t_i - t_{i-1})$ . Thus  $N$  is a Poisson process with intensity 1.

Finally, the law of a locally finite random measure on  $\mathbb{R}_+$  is determined by the joint laws of its masses on the intervals  $[0, t]$ ,  $t \in \mathbb{Q}_+$ . Indeed, on each  $[0, m]$ , evaluations on intervals with rational endpoints generate all Borel evaluations by a monotone class argument; we then let  $m$  tend to infinity. Here these joint laws are those of a Poisson random measure with intensity  $\lambda$ . Thus  $M$  is a Poisson random measure with this intensity.

\* \* \*

**Solution to Exercise 7.6.** [statement p. 208] Take the countable–cocountable sigma-algebra

$$\mathcal{E} = \{B \subseteq [0, 1] : B \text{ is countable or } [0, 1] \setminus B \text{ is countable}\},$$

and set

$$\mu(B) = \begin{cases} 0 & \text{if } B \text{ is countable,} \\ 1 & \text{if } [0, 1] \setminus B \text{ is countable.} \end{cases}$$

The two cases are disjoint since  $[0, 1]$  is uncountable. The family  $\mathcal{E}$  is closed under complements. A countable union of its elements is countable if they all are, and cocountable as soon as one is: it is therefore indeed a sigma-algebra.

Let us check countable additivity of  $\mu$ . In a countable family of pairwise disjoint elements of  $\mathcal{E}$ , there is at most one cocountable set. If there is none, the union is countable and both sides of the additivity identity are zero. If there is one, all the others are countable and both sides equal 1. Thus  $\mu$  is a probability measure. Finally, for every  $x \in [0, 1]$ ,  $\{x\} \in \mathcal{E}$  and  $\mu(\{x\}) = 0 \neq 1 = \delta_x(\{x\})$ : thus  $\mu$  is not a Dirac measure.