

# Weak Convergence of Probability Measures

SERIK SAGITOV, Chalmers University of Technology and Gothenburg University

## Abstract

This text contains my lecture notes for the graduate course “Weak Convergence” given in September-October 2013 and then in March-May 2015. The course is based on the book *Convergence of Probability Measures* by Patrick Billingsley, partially covering Chapters 1-3, 5-9, 12-14, 16, as well as appendices. In this text the formula label (\*) operates locally.

I am grateful to Timo Hirscher whose numerous valuable suggestions helped me to improve earlier versions of these notes. Last updated: July 21, 2020.

## Contents

|   |           |
|---|-----------|
| <b>Abstract</b>   | <b>1</b>  |
| <b>1 The Portmanteau and mapping theorems</b>   | <b>3</b>  |
| 1.1 Metric spaces . . . . .   | 3         |
| 1.2 Convergence in distribution and weak convergence . . . . .                        | 4         |
| 1.3 Convergence in probability and in total variation. Local limit theorems . . . . . | 6         |
| <b>2 Convergence of finite-dimensional distributions</b>                              | <b>7</b>  |
| 2.1 Separating and convergence-determining classes . . . . .                          | 7         |
| 2.2 Weak convergence in product spaces . . . . .                                      | 8         |
| 2.3 Weak convergence in $\mathbf{R}^k$ and $\mathbf{R}^\infty$ . . . . .              | 9         |
| 2.4 Kolmogorov’s extension theorem . . . . .  | 10        |
| <b>3 Tightness and Prokhorov’s theorem</b>  | <b>10</b> |
| 3.1 Tightness of probability measures . . . . .                                       | 10        |
| 3.2 Proof of Prokhorov’s theorem . . . . .  | 12        |
| 3.3 Skorokhod’s representation theorem . . . . .                                      | 13        |
| <b>4 Functional Central Limit Theorem on <math>C = C[0, 1]</math></b>                 | <b>14</b> |
| 4.1 Weak convergence in $C$ . . . . .   | 14        |
| 4.2 Wiener measure and Donsker’s theorem . . . . .                                    | 16        |
| 4.3 Tightness in $C$ . . . . .  | 18        |
| <b>5 Applications of the functional CLT</b>   | <b>19</b> |
| 5.1 The minimum and maximum of the Brownian path . . . . .                            | 19        |
| 5.2 The arcsine law . . . . .   | 21        |
| 5.3 The Brownian bridge . . . . .   | 23        |
| <b>6 The space <math>D = D[0, 1]</math></b>   | <b>25</b> |
| 6.1 Cadlag functions . . . . .  | 25        |
| 6.2 Two metrics in $D$ and the Skorokhod topology . . . . .                           | 27        |
| 6.3 Separability and completeness of $D$ . . . . .                                    | 29        |
| 6.4 Relative compactness in the Skorokhod topology . . . . .                          | 30        |
| <b>7 Probability measures on <math>D</math> and random elements</b>                   | <b>31</b> |
| 7.1 Finite-dimensional distributions on $D$ . . . . .                                 | 31        |
| 7.2 Tightness criteria in $D$ . . . . .   | 33        |
| 7.3 A key condition on 3-dimensional distributions . . . . .                          | 33        |
| 7.4 A criterion for existence . . . . .   | 35        |

|          |   |           |
|----------|---|-----------|
| <b>8</b> | <b>Weak convergence on <math>D</math></b>                         | <b>36</b> |
| 8.1      | Criteria for weak convergence in $D$ . . . . .                    | 37        |
| 8.2      | Functional CLT on $D$ . . . . .                                   | 38        |
| 8.3      | Empirical distribution functions . . . . .                        | 39        |
| <b>9</b> | <b>The space <math>D_\infty = D[0, \infty)</math></b>             | <b>42</b> |
| 9.1      | Two metrics on $D_\infty$ . . . . .                               | 42        |
| 9.2      | Characterization of Skorokhod convergence on $D_\infty$ . . . . . | 43        |
| 9.3      | Separability and completeness of $D_\infty$ . . . . .             | 45        |
| 9.4      | Weak convergence on $D_\infty$ . . . . .                          | 45        |

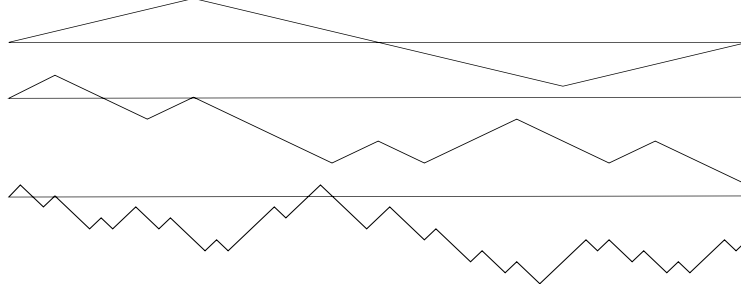


Figure 1: Scaled symmetric simple random walk  $X_t^n(\omega)$  for a fixed  $\omega \in \Omega$  and  $n = 4, 16, 64$ .

## Introduction

Throughout these lecture notes we use the following notation

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du.$$

Consider a symmetric simple random walk  $S_n = \xi_1 + \dots + \xi_n$  with  $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = 1/2$ . The random sequence  $S_n$  has no limit in the usual sense. However, by de Moivre's theorem (1733),

$$\mathbb{P}(S_n \leq z\sqrt{n}) \rightarrow \Phi(z) \text{ as } n \rightarrow \infty \text{ for any } z \in \mathbf{R}.$$

This is an example of convergence in distribution  $\frac{S_n}{\sqrt{n}} \Rightarrow Z$  to a normally distributed random variable. Define a sequence of stochastic processes  $X^n = (X_t^n)_{t \in [0,1]}$  by linear interpolation between its values  $X_{i/n}^n(\omega) = \frac{S_i(\omega)}{\sigma\sqrt{n}}$  at the points  $t = i/n$ , see Figure 1. The much more powerful functional CLT claims convergence in distribution towards the Wiener process  $X^n \Rightarrow W$ .

This course deals with weak convergence of probability measures on Polish spaces  $(\mathbf{S}, \mathcal{S})$ . For us, the principal examples of Polish spaces (complete separable metric spaces) are

- the space  $\mathbf{C} = \mathbf{C}[0, 1]$  of continuous trajectories  $x : [0, 1] \rightarrow \mathbf{R}$  (Section 4),
- the space  $\mathbf{D} = \mathbf{D}[0, 1]$  of cadlag trajectories  $x : [0, 1] \rightarrow \mathbf{R}$  (Section 6),
- the space  $\mathbf{D}[0, \infty)$  of cadlag trajectories  $x : [0, \infty) \rightarrow \mathbf{R}$  (Section 9).

To prove the functional CLT  $X^n \Rightarrow W$ , we have to check that  $\mathbb{E}f(X^n) \rightarrow \mathbb{E}f(W)$  for all bounded continuous functions  $f : \mathbf{C}[0, 1] \rightarrow \mathbf{R}$ , which is not practical to do straightforwardly. Instead, one starts with the finite-dimensional distributions

$$(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow (W_{t_1}, \dots, W_{t_k}).$$

To prove the weak convergence of the finite-dimensional distributions, it is enough to check the convergence of moment generating functions, thus allowing us to focus on a special class of continuous functions  $f_{\lambda_1, \dots, \lambda_k} : \mathbf{R}^k \rightarrow \mathbf{R}$ , where  $\lambda_i \geq 0$  and

$$f_{\lambda_1, \dots, \lambda_k}(x_1, \dots, x_k) = \exp(\lambda_1 x_1 + \dots + \lambda_k x_k).$$

For the weak convergence in the infinite-dimensional space  $\mathbf{C}[0, 1]$ , the usual additional step is to verify tightness of the distributions of the family of processes  $(X^n)$ . Loosely speaking, tightness means that no probability mass escapes to infinity. By Prokhorov theorem (Section 3), tightness implies relative compactness, which means that each subsequence of  $X^n$  contains a further subsequence converging weakly. Since all possible limits have the finite-dimensional distributions of  $W$ , we conclude that all subsequences converge to the same limit  $W$ , and by this we establish the convergence  $X^n \Rightarrow W$ .

This approach makes it crucial to find tightness criteria in  $\mathbf{C}[0, 1]$ ,  $\mathbf{D}[0, 1]$ , and then in  $\mathbf{D}[0, \infty)$ .

## 1 The Portmanteau and mapping theorems

### 1.1 Metric spaces

Consider a metric space  $\mathbf{S}$  with metric  $\rho(x, y)$ . For subsets  $A \subset \mathbf{S}$ , denote the closure by  $A^-$ , the interior by  $A^\circ$ , and the boundary by  $\partial A = A^- - A^\circ$ . We write

$$\rho(x, A) = \inf\{\rho(x, y) : y \in A\}, \quad A^\epsilon = \{x : \rho(x, A) < \epsilon\}.$$

**Definition 1.1** Open balls  $B(x, r) = \{y \in \mathbf{S} : \rho(x, y) < r\}$  form a *base* for  $\mathbf{S}$ : each *open set* in  $\mathbf{S}$  is a union of open balls. Complements to the open sets are called *closed sets*. The *Borel  $\sigma$ -algebra*  $\mathcal{S}$  is formed from the open and closed sets in  $\mathbf{S}$  using the operations of countable intersection, countable union, and set difference.

**Definition 1.2** A collection  $\mathcal{A}$  of  $\mathbf{S}$ -subsets is called a  $\pi$ -system if it is closed under intersection, that is if  $A, B \in \mathcal{A}$ , then  $A \cap B \in \mathcal{A}$ . We say that  $\mathcal{L}$  is a  $\lambda$ -system if: (i)  $\mathbf{S} \in \mathcal{L}$ , (ii)  $A \in \mathcal{L}$  implies  $A^c \in \mathcal{L}$ , (iii) for any sequence of disjoint sets  $A_n \in \mathcal{L}$ ,  $\cup_n A_n \in \mathcal{L}$ .

**Theorem 1.3** *Dynkin's  $\pi$ - $\lambda$  lemma.* If  $\mathcal{A}$  is a  $\pi$ -system such that  $\mathcal{A} \subset \mathcal{L}$ , where  $\mathcal{L}$  is a  $\lambda$ -system, then  $\sigma(\mathcal{A}) \subset \mathcal{L}$ , where  $\sigma(\mathcal{A})$  is the  $\sigma$ -algebra generated by  $\mathcal{A}$ .

**Definition 1.4** A metric space  $\mathbf{S}$  is called *separable* if it contains a countable dense subset. It is called *complete* if every Cauchy (fundamental) sequence has a limit lying in  $\mathbf{S}$ . A complete separable metric space is called a *Polish space*.

Separability is a topological property, while completeness is a property of the metric and not of the topology.

**Definition 1.5** An *open cover* of  $A \subset \mathbf{S}$  is a class of open sets whose union contains  $A$ .

**Theorem 1.6** *These three conditions are equivalent:*

- (i)  $\mathbf{S}$  is separable,
- (ii)  $\mathbf{S}$  has a countable base (a class of open sets such that each open set is a union of sets in the class),
- (iii) Each open cover of each subset of  $\mathbf{S}$  has a countable subcover.

**Theorem 1.7** *Suppose that the subset  $M$  of  $\mathbf{S}$  is separable.*

- (i) *There is a countable class  $\mathcal{A}$  of open sets with the property that, if  $x \in G \cap M$  and  $G$  is open, then  $x \in A \subset A^- \subset G$  for some  $A \in \mathcal{A}$ .*
- (ii) *Lindelöf property. Each open cover of  $M$  has a countable subcover.*

**Definition 1.8** A set  $K$  is called *compact* if each open cover of  $K$  has a finite subcover. A set  $A \subset \mathbf{S}$  is called *relatively compact* if each sequence in  $A$  has a convergent subsequence the limit of which may not lie in  $A$ .

**Theorem 1.9** *Let  $A$  be a subset of a metric space  $\mathbf{S}$ . The following three conditions are equivalent:*

- (i)  $A^-$  is compact,
- (ii)  $A$  is relatively compact,
- (iii)  $A^-$  is complete and  $A$  is totally bounded (that is for any  $\epsilon > 0$ ,  $A$  has a finite  $\epsilon$ -net the points of which are not required to lie in  $A$ ).

**Theorem 1.10** *Consider two metric spaces  $(\mathbf{S}, \rho)$  and  $(\mathbf{S}', \rho')$  and maps  $h, h_n : \mathbf{S} \rightarrow \mathbf{S}'$ . If  $h$  is continuous, then it is measurable  $\mathbf{S}/\mathbf{S}'$ . If each  $h_n$  is measurable  $\mathbf{S}/\mathbf{S}'$ , and if  $h_n x \rightarrow h x$  for every  $x \in \mathbf{S}$ , then  $h$  is also measurable  $\mathbf{S}/\mathbf{S}'$ .*

## 1.2 Convergence in distribution and weak convergence

**Definition 1.11** Let  $P_n, P$  be probability measures on  $(\mathbf{S}, \mathcal{S})$ . We say  $P_n \Rightarrow P$  *weakly converges* as  $n \rightarrow \infty$  if for any bounded continuous function  $f : \mathbf{S} \rightarrow \mathbf{R}$

$$\int_{\mathbf{S}} f(x) P_n(dx) \rightarrow \int_{\mathbf{S}} f(x) P(dx), \quad n \rightarrow \infty.$$

**Definition 1.12** Let  $X$  be a  $(\mathbf{S}, \mathcal{S})$ -valued random element defined on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . We say that a probability measure  $P$  on  $\mathbf{S}$  is the probability distribution of  $X$  if  $P(A) = \mathbb{P}(X \in A)$  for all  $A \in \mathcal{S}$ .

**Definition 1.13** Let  $X_n, X$  be  $(\mathbf{S}, \mathcal{S})$ -valued random elements defined on the probability spaces  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ ,  $(\Omega, \mathcal{F}, \mathbb{P})$ . We say  $X_n$  converge in distribution to  $X$  as  $n \rightarrow \infty$  and write  $X_n \Rightarrow X$ , if for any bounded continuous function  $f : \mathbf{S} \rightarrow \mathbf{R}$ ,

$$\mathbb{E}_n(f(X_n)) \rightarrow \mathbb{E}(f(X)), \quad n \rightarrow \infty.$$

This is equivalent to the weak convergence  $P_n \Rightarrow P$  of the respective probability distributions.

**Example 1.14** The function  $f(x) = 1_{\{x \in A\}}$  is bounded but not continuous, therefore if  $P_n \Rightarrow P$ , then  $P_n(A) \rightarrow P(A)$  does not always hold. For  $\mathbf{S} = \mathbf{R}$ , the function  $f(x) = x$  is continuous but not bounded, therefore if  $X_n \Rightarrow X$ , then  $\mathbb{E}_n(X_n) \rightarrow \mathbb{E}(X)$  does not always hold.

**Definition 1.15** Call  $A \in \mathcal{S}$  a *P-continuity set* if  $P(\partial A) = 0$ .

**Theorem 1.16** *Portmanteau's theorem. The following five statements are equivalent.*

- (i)  $P_n \Rightarrow P$ .
- (ii)  $\int f(x)P_n(dx) \rightarrow \int f(x)P(dx)$  for all bounded uniformly continuous  $f : \mathbf{S} \rightarrow \mathbf{R}$ .
- (iii)  $\limsup_{n \rightarrow \infty} P_n(F) \leq P(F)$  for all closed  $F \in \mathcal{S}$ .
- (iv)  $\liminf_{n \rightarrow \infty} P_n(G) \geq P(G)$  for all open  $G \in \mathcal{S}$ .
- (v)  $P_n(A) \rightarrow P(A)$  for all  $P$ -continuity sets  $A$ .

Proof. (i)  $\rightarrow$  (ii) is trivial.

(ii)  $\rightarrow$  (iii). For a closed  $F \in \mathcal{S}$  put

$$g(x) = (1 - \epsilon^{-1}\rho(x, F)) \vee 0.$$

This function is bounded and uniformly continuous since  $|g(x) - g(y)| \leq \epsilon^{-1}\rho(x, y)$ . Using

$$1_{\{x \in F\}} \leq g(x) \leq 1_{\{x \in F^\epsilon\}},$$

we derive (iii) from (ii):

$$\limsup_{n \rightarrow \infty} P_n(F) \leq \limsup_{n \rightarrow \infty} \int g(x)P_n(dx) = \int g(x)P(dx) \leq P(F^\epsilon) \rightarrow P(F), \quad \epsilon \rightarrow 0.$$

(iii)  $\rightarrow$  (iv) follows by complementation.

(iii) + (iv)  $\rightarrow$  (v). If  $P(\partial A) = 0$ , then the leftmost and rightmost probabilities coincide:

$$\begin{aligned} P(A^-) &\geq \limsup_{n \rightarrow \infty} P_n(A^-) \geq \limsup_{n \rightarrow \infty} P_n(A) \\ &\geq \liminf_{n \rightarrow \infty} P_n(A) \geq \liminf_{n \rightarrow \infty} P_n(A^\circ) \geq P(A^\circ). \end{aligned}$$

(v)  $\rightarrow$  (i). By linearity we may assume that the bounded continuous function  $f$  satisfies  $0 \leq f \leq 1$ . Then putting  $A_t = \{x : f(x) > t\}$  we get

$$\int_{\mathbf{S}} f(x)P_n(dx) = \int_0^1 P_n(A_t)dt \rightarrow \int_0^1 P(A_t)dt = \int_{\mathbf{S}} f(x)P(dx).$$

Here the convergence follows from (v) since  $f$  is continuous, implying that  $\partial A_t = \{x : f(x) = t\}$ , and since  $\{x : f(x) = t\}$  are  $P$ -continuity sets except for countably many  $t$ . We also used the bounded convergence theorem.

**Example 1.17** Let  $F(x) = \mathbb{P}(X \leq x)$ . Then  $X_n = X + n^{-1}$  has distribution  $F_n(x) = F(x - n^{-1})$ . As  $n \rightarrow \infty$ ,  $F_n(x) \rightarrow F(x-)$ , so convergence only occurs at continuity points.

**Corollary 1.18** *A single sequence of probability measures can not weakly converge to each of two different limits.*

Proof. It suffices to prove that if  $\int_{\mathbf{S}} f(x)P(dx) = \int_{\mathbf{S}} f(x)Q(dx)$  for all bounded, uniformly continuous functions  $f : \mathbf{S} \rightarrow \mathbf{R}$ , then  $P = Q$ . Using the bounded, uniformly continuous functions  $g(x) = (1 - \epsilon^{-1}\rho(x, F)) \vee 0$  we get

$$P(F) \leq \int_{\mathbf{S}} g(x)P(dx) = \int_{\mathbf{S}} g(x)Q(dx) \leq Q(F^\epsilon).$$

Letting  $\epsilon \rightarrow 0$  it gives for any closed set  $F$ , that  $P(F) \leq Q(F)$  and by symmetry we conclude that  $P(F) = Q(F)$ . It follows that  $P(G) = Q(G)$  for all open sets  $G$ .

It remains to use regularity of any probability measure  $P$ : if  $A \in \mathcal{S}$  and  $\epsilon > 0$ , then there exist a closed set  $F_\epsilon$  and an open set  $G_\epsilon$  such that  $F_\epsilon \subset A \subset G_\epsilon$  and  $P(G_\epsilon - F_\epsilon) < \epsilon$ . To this end we denote by  $\mathcal{G}_P$  the class of  $\mathcal{S}$ -sets with the just stated property. If  $A$  is closed, we can take  $F = A$  and  $G = F^\delta$ , where  $\delta$  is small enough. Thus all closed sets belong to  $\mathcal{G}_P$ , and we need to show that  $\mathcal{G}_P$  forms a  $\sigma$ -algebra. Given  $A_n \in \mathcal{G}_P$ , choose closed sets  $F_n$  and open sets  $G_n$  such that  $F_n \subset A_n \subset G_n$  and  $P(G_n - F_n) < 2^{-n-1}\epsilon$ . If  $G = \cup_n G_n$  and  $F = \cup_{n \leq n_0} F_n$  with  $n_0$  chosen so that  $P(\cup_n F_n - F) < \epsilon/2$ , then  $F \subset \cup_n A_n \subset G$  and  $P(G - F) < \epsilon$ . Thus  $\mathcal{G}_P$  is closed under the formation of countable unions. Since it is closed under complementation,  $\mathcal{G}_P$  is a  $\sigma$ -algebra.

**Theorem 1.19** *Mapping theorem. Let  $X_n$  and  $X$  be random elements of a metric space  $\mathbf{S}$ . Let  $h : \mathbf{S} \rightarrow \mathbf{S}'$  be a  $\mathcal{S}/\mathcal{S}'$ -measurable mapping and  $D_h$  be the set of its discontinuity points. If  $X_n \Rightarrow X$  and  $\mathbb{P}(X \in D_h) = 0$ , then  $h(X_n) \Rightarrow h(X)$ .*

*In other terms, if  $P_n \Rightarrow P$  and  $P(D_h) = 0$ , then  $P_n h^{-1} \Rightarrow P h^{-1}$ .*

Proof. We show first that  $D_h$  is a Borel subset of  $\mathcal{S}$ . For any pair  $(\epsilon, \delta)$  of positive rationals, the set

$$A_{\epsilon\delta} = \{x \in \mathcal{S} : \text{there exist } y, z \in \mathcal{S} \text{ such that } \rho(x, y) < \delta, \rho(x, z) < \delta, \rho'(hy, hz) \geq \epsilon\}$$

is open. Therefore,  $D_h = \bigcup_{\epsilon} \bigcap_{\delta} A_{\epsilon\delta} \in \mathcal{S}$ . Now, for each  $F \in \mathcal{S}'$ ,

$$\begin{aligned} \limsup_{n \rightarrow \infty} P_n(h^{-1}F) &\leq \limsup_{n \rightarrow \infty} P_n((h^{-1}F)^-) \leq P((h^{-1}F)^-) \\ &\leq P(h^{-1}(F^-) \cup D_h) = P(h^{-1}(F^-)). \end{aligned}$$

To see that  $(h^{-1}F)^- \subset h^{-1}(F^-) \cup D_h$  take an element  $x \in (h^{-1}F)^-$ . There is a sequence  $x_n \rightarrow x$  such that  $h(x_n) \in F$ , and therefore, either  $h(x_n) \rightarrow h(x)$  or  $x \in D_h$ . By the Portmanteau theorem, the last chain of inequalities implies  $P_n h^{-1} \Rightarrow P h^{-1}$ .

**Example 1.20** Let  $P_n \Rightarrow P$ . If  $A$  is a  $P$ -continuity set and  $h(x) = 1_{\{x \in A\}}$ , then by the mapping theorem,  $P_n h^{-1} \Rightarrow P h^{-1}$ .

### 1.3 Convergence in probability and in total variation. Local limit theorems

**Definition 1.21** Suppose  $X_n$  and  $X$  are random elements of  $\mathcal{S}$  defined on the same probability space. If  $\mathbb{P}(\rho(X_n, X) < \epsilon) \rightarrow 1$  for each positive  $\epsilon$ , we say  $X_n$  converge to  $X$  in probability and write  $X_n \xrightarrow{P} X$ .

**Exercise 1.22** Convergence in probability  $X_n \xrightarrow{P} X$  is equivalent to the weak convergence  $\rho(X^n, X) \Rightarrow 0$ . Moreover,  $(X_1^n, \dots, X_k^n) \xrightarrow{P} (X_1, \dots, X_k)$  if and only if  $X_i^n \xrightarrow{P} X_i$  for all  $i = 1, \dots, k$ .

**Theorem 1.23** Suppose  $(X_n, X_{u,n})$  are random elements of  $\mathcal{S} \times \mathcal{S}$ . If  $X_{u,n} \Rightarrow Z_u$  as  $n \rightarrow \infty$  for any fixed  $u$ , and  $Z_u \Rightarrow X$  as  $u \rightarrow \infty$ , and

$$\lim_{u \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\rho(X_{u,n}, X_n) \geq \epsilon) = 0, \text{ for each positive } \epsilon,$$

then  $X_n \Rightarrow X$ .

Proof. Let  $F \in \mathcal{S}$  be closed and define  $F_\epsilon$  as the set  $\{x : \rho(x, F) \leq \epsilon\}$ . Then

$$\begin{aligned} \mathbb{P}(X_n \in F) &= \mathbb{P}(X_n \in F, X_{u,n} \notin F_\epsilon) + \mathbb{P}(X_n \in F, X_{u,n} \in F_\epsilon) \\ &\leq \mathbb{P}(\rho(X_{u,n}, X_n) \geq \epsilon) + \mathbb{P}(X_{u,n} \in F_\epsilon). \end{aligned}$$

Since  $F_\epsilon$  is also closed and  $F_\epsilon \downarrow F$  as  $\epsilon \downarrow 0$ , we get

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{P}(X_n \in F) &\leq \limsup_{\epsilon \rightarrow 0} \limsup_{u \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(X_{u,n} \in F_\epsilon) \\ &\leq \limsup_{\epsilon \rightarrow 0} \mathbb{P}(X \in F_\epsilon) = \mathbb{P}(X \in F). \end{aligned}$$

**Corollary 1.24** Suppose  $(X_n, Y_n)$  are random elements of  $\mathcal{S} \times \mathcal{S}$ . If  $Y_n \Rightarrow X$  as  $n \rightarrow \infty$  and  $\rho(X_n, Y_n) \Rightarrow 0$ , then  $X_n \Rightarrow X$ . Taking  $Y_n \equiv X$ , we conclude that convergence in probability implies convergence in distribution.

**Definition 1.25** Convergence in total variation  $P_n \xrightarrow{\text{TV}} P$  means

$$\sup_{A \in \mathcal{S}} |P_n(A) - P(A)| \rightarrow 0.$$

**Theorem 1.26** Scheffe's theorem. Suppose  $P_n$  and  $P$  have densities  $f_n$  and  $f$  with respect to a measure  $\mu$  on  $(\mathcal{S}, \mathcal{S})$ . If  $f_n \rightarrow f$  almost everywhere with respect to  $\mu$ , then  $P_n \xrightarrow{\text{TV}} P$  and therefore  $P_n \Rightarrow P$ .

Proof. For any  $A \in \mathcal{S}$

$$\begin{aligned} |P_n(A) - P(A)| &= \left| \int_A (f_n(x) - f(x)) \mu(dx) \right| \leq \int_{\mathcal{S}} |f(x) - f_n(x)| \mu(dx) \\ &= 2 \int_{\mathcal{S}} (f(x) - f_n(x))^+ \mu(dx), \end{aligned}$$

where the last equality follows from

$$0 = \int_{\mathcal{S}} (f(x) - f_n(x)) \mu(dx) = \int_{\mathcal{S}} (f(x) - f_n(x))^+ \mu(dx) - \int_{\mathcal{S}} (f(x) - f_n(x))^- \mu(dx).$$

On the other hand, by the dominated convergence theorem,  $\int (f(x) - f_n(x))^+ \mu(dx) \rightarrow 0$ .

**Example 1.27** According to Theorem 1.26 the local limit theorem implies the integral limit theorem  $P_n \Rightarrow P$ . The reverse implication is false. Indeed, let  $P = \mu$  be Lebesgue measure on  $\mathbf{S} = [0, 1]$  so that  $f \equiv 1$ . Let  $P_n$  be the uniform distribution on the set

$$B_n = \bigcup_{k=0}^{n-1} (kn^{-1}, kn^{-1} + n^{-3})$$

with density  $f_n(x) = n^2 1_{\{x \in B_n\}}$ . Since  $\mu(B_n) = n^{-2}$ , the Borel-Cantelli lemma implies that  $\mu(B_n \text{ i.o.}) = 0$ . Thus  $f_n(x) \rightarrow 0$  for almost all  $x$  and there is no local theorem. On the other hand,  $|P_n[0, x] - x| \leq n^{-1}$  implying  $P_n \Rightarrow P$ .

**Theorem 1.28** Let  $\mathbf{S} = \mathbf{R}^k$ . Denote by  $L_n \subset \mathbf{R}^k$  a lattice with cells having dimensions  $(\delta_1(n), \dots, \delta_k(n))$  so that the cells of the lattice  $L_n$  all having the form

$$B_n(x) = \{y : x_1 - \delta_1(n) < y_1 \leq x_1, \dots, x_k - \delta_k(n) < y_k \leq x_k\}, \quad x \in L_n$$

have size  $v_n = \delta_1(n) \cdots \delta_k(n)$ . Suppose that  $(P_n)$  is a sequence of probability measures on  $\mathbf{R}^k$ , where  $P_n$  is supported by  $L_n$  with probability mass function  $p_n(x)$ .

Suppose that  $P$  is a probability measure on  $\mathbf{R}^k$  having density  $f$  with respect to Lebesgue measure. Assume that all  $\delta_i(n) \rightarrow 0$  as  $n \rightarrow \infty$ . If  $\frac{p_n(x_n)}{v_n} \rightarrow f(x)$  whenever  $x_n \in L_n$  and  $x_n \rightarrow x$ , then  $P_n \Rightarrow P$ .

Proof. Define a probability density  $f_n$  on  $\mathbf{R}^k$  by setting  $f_n(y) = \frac{p_n(x)}{v_n}$  for  $y \in B_n(x)$ . It follows that  $f_n(y) \rightarrow f(y)$  for all  $y \in \mathbf{R}^k$ . Let a random vector  $Y_n$  have the density  $f_n$  and  $X$  have the density  $f$ . By Theorem 1.26,  $Y_n \Rightarrow X$ . Define  $X_n$  on the same probability space as  $Y_n$  by setting  $X_n = x$  if  $Y_n$  lies in the cell  $B_n(x)$ . Since  $\|X_n - Y_n\| \leq \|\delta(n)\|$ , we conclude using Corollary 1.24 that  $X_n \Rightarrow X$ .

**Example 1.29** If  $S_n$  is the number of successes in  $n$  Bernoulli trials, then according to the local form of the de Moivre-Laplace theorem,

$$\mathbb{P}(S_n = i) \sqrt{npq} = \binom{n}{i} p^i q^{n-i} \sqrt{npq} \rightarrow \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

provided  $i$  varies with  $n$  in such a way that  $\frac{i-np}{\sqrt{npq}} \rightarrow z$ . Therefore, Theorem 1.28 applies to the lattice

$$L_n = \left\{ \frac{i-np}{\sqrt{npq}}, i \in \mathbb{Z} \right\}$$

with  $v_n = \frac{1}{\sqrt{npq}}$  and the probability mass function  $p_n(\frac{i-np}{\sqrt{npq}}) = \mathbb{P}(S_n = i)$  for  $i = 0, \dots, n$ . As a result we get the integral form of the de Moivre-Laplace theorem:

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} \leq z\right) \rightarrow \Phi(z) \text{ as } n \rightarrow \infty \text{ for any } z \in \mathbf{R}.$$

## 2 Convergence of finite-dimensional distributions

### 2.1 Separating and convergence-determining classes

**Definition 2.1** Call a subclass  $\mathcal{A} \subset \mathcal{S}$  a *separating class* if any two probability measures with  $P(A) = Q(A)$  for all  $A \in \mathcal{A}$ , must be identical:  $P(A) = Q(A)$  for all  $A \in \mathcal{S}$ .

Call a subclass  $\mathcal{A} \subset \mathcal{S}$  a *convergence-determining class* if, for every  $P$  and every sequence  $(P_n)$ , convergence  $P_n(A) \rightarrow P(A)$  for all  $P$ -continuity sets  $A \in \mathcal{A}$  implies  $P_n \Rightarrow P$ .

**Lemma 2.2** If  $\mathcal{A} \subset \mathcal{S}$  is a  $\pi$ -system and  $\sigma(\mathcal{A}) = \mathcal{S}$ , then  $\mathcal{A}$  is a separating class.

Proof. Consider a pair of probability measures such that  $P(A) = Q(A)$  for all  $A \in \mathcal{A}$ . Let  $\mathcal{L} = \mathcal{L}_{P,Q}$  be the class of all sets  $A \in \mathcal{S}$  such that  $P(A) = Q(A)$ . Clearly,  $\mathbf{S} \in \mathcal{L}$ . If  $A \in \mathcal{L}$ , then  $A^c \in \mathcal{L}$  since  $P(A^c) = 1 - P(A) = 1 - Q(A) = Q(A^c)$ . If  $A_n$  are disjoint sets in  $\mathcal{L}$ , then  $\cup_n A_n \in \mathcal{L}$  since

$$P(\cup_n A_n) = \sum_n P(A_n) = \sum_n Q(A_n) = Q(\cup_n A_n).$$

Therefore  $\mathcal{L}$  is a  $\lambda$ -system, and since  $\mathcal{A} \subset \mathcal{L}$ , Theorem 1.3 gives  $\sigma(\mathcal{A}) \subset \mathcal{L}$ , and  $\mathcal{L} = \mathcal{S}$ .

**Theorem 2.3** Suppose that  $P$  is a probability measure on a separable  $\mathbf{S}$ , and a subclass  $\mathcal{A}_P \subset \mathcal{S}$  satisfies

(i)  $\mathcal{A}_P$  is a  $\pi$ -system,

(ii) for every  $x \in \mathbf{S}$  and  $\epsilon > 0$ , there is an  $A \in \mathcal{A}_P$  for which  $x \in A^\circ \subset A \subset B(x, \epsilon)$ .

If  $P_n(A) \rightarrow P(A)$  for every  $A \in \mathcal{A}_P$ , then  $P_n \Rightarrow P$ .

Proof. If  $A_1, \dots, A_r$  lie in  $\mathcal{A}_P$ , so do their intersections. Hence, by the inclusion-exclusion formula and a theorem assumption,

$$\begin{aligned} P_n\left(\bigcup_{i=1}^r A_i\right) &= \sum_i P_n(A_i) - \sum_{ij} P_n(A_i \cap A_j) + \sum_{ijk} P_n(A_i \cap A_j \cap A_k) - \dots \\ &\rightarrow \sum_i P(A_i) - \sum_{ij} P(A_i \cap A_j) + \sum_{ijk} P(A_i \cap A_j \cap A_k) - \dots = P\left(\bigcup_{i=1}^r A_i\right). \end{aligned}$$

If  $G \subset \mathbf{S}$  is open, then for each  $x \in G$ ,  $x \in A_x^\circ \subset A_x \subset G$  holds for some  $A_x \in \mathcal{A}_P$ . Since  $\mathbf{S}$  is separable, by Theorem 1.6 (iii), there is a countable sub-collection  $(A_{x_i})$  that covers  $G$ . Thus  $G = \cup_i A_{x_i}$ , where all  $A_{x_i}$  are  $\mathcal{A}_P$ -sets.

With  $A_i = A_{x_i}$  we have  $G = \cup_i A_i$ . Given  $\epsilon$ , choose  $r$  so that  $P(\cup_{i=1}^r A_i) > P(G) - \epsilon$ . Then,

$$P(G) - \epsilon < P\left(\bigcup_{i=1}^r A_i\right) = \lim_n P_n\left(\bigcup_{i=1}^r A_i\right) \leq \liminf_n P_n(G).$$

Now, letting  $\epsilon \rightarrow 0$  we find that for any open set  $\liminf_n P_n(G) \geq P(G)$ .

**Theorem 2.4** Suppose that  $\mathbf{S}$  is separable and consider a subclass  $\mathcal{A} \subset \mathcal{S}$ . Let  $\mathcal{A}_{x,\epsilon}$  be the class of  $A \in \mathcal{A}$  satisfying  $x \in A^\circ \subset A \subset B(x, \epsilon)$ , and let  $\partial\mathcal{A}_{x,\epsilon}$  be the class of their boundaries. If

(i)  $\mathcal{A}$  is a  $\pi$ -system,

(ii) for every  $x \in \mathbf{S}$  and  $\epsilon > 0$ ,  $\partial\mathcal{A}_{x,\epsilon}$  contains uncountably many disjoint sets,

then  $\mathcal{A}$  is a convergence-determining class.

Proof. For an arbitrary  $P$  let  $\mathcal{A}_P$  be the class of  $P$ -continuity sets in  $\mathcal{A}$ . We have to show that if  $P_n(A) \rightarrow P(A)$  holds for every  $A \in \mathcal{A}_P$ , then  $P_n \Rightarrow P$ . Indeed, by (i), since  $\partial(A \cap B) \subset \partial(A) \cup \partial(B)$ ,  $\mathcal{A}_P$  is a  $\pi$ -system. By (ii), there is an  $A_x \in \mathcal{A}_{x,\epsilon}$  such that  $P(\partial A_x) = 0$  so that  $A_x \in \mathcal{A}_P$ . It remains to apply Theorem 2.3.

## 2.2 Weak convergence in product spaces

**Definition 2.5** Let  $P$  be a probability measure on  $\mathbf{S} = \mathbf{S}' \times \mathbf{S}''$  with the product metric

$$\rho((x', x''), (y', y'')) = \rho'(x', y') \vee \rho''(x'', y'').$$

Define the marginal distributions by  $P'(A') = P(A' \times \mathbf{S}'')$  and  $P''(A'') = P(\mathbf{S}' \times A'')$ . If the marginals are independent, we write  $P = P' \times P''$ . We denote by  $\mathbf{S}' \times \mathbf{S}''$  the product  $\sigma$ -algebra generated by the measurable rectangles  $A' \times A''$  for  $A' \in \mathcal{S}'$  and  $A'' \in \mathcal{S}''$ .

**Lemma 2.6** If  $\mathbf{S} = \mathbf{S}' \times \mathbf{S}''$  is separable, then the three Borel  $\sigma$ -algebras are related by  $\mathcal{S} = \mathcal{S}' \times \mathcal{S}''$ .

Proof. Consider the projections  $\pi' : \mathbf{S} \rightarrow \mathbf{S}'$  and  $\pi'' : \mathbf{S} \rightarrow \mathbf{S}''$  defined by  $\pi'(x', x'') = x'$  and  $\pi''(x', x'') = x''$ , each is continuous. For  $A' \in \mathcal{S}'$  and  $A'' \in \mathcal{S}''$ , we have

$$A' \times A'' = (\pi')^{-1}A' \cap (\pi'')^{-1}A'' \in \mathcal{S},$$

since the two projections are continuous and therefore measurable. Thus  $\mathbf{S}' \times \mathbf{S}'' \subset \mathcal{S}$ . On the other hand, if  $\mathbf{S}$  is separable, then each open set in  $\mathbf{S}$  is a countable union of the balls

$$B((x', x''), r) = B'(x', r) \times B''(x'', r)$$

and hence lies in  $\mathcal{S}' \times \mathcal{S}''$ . Thus  $\mathcal{S} \subset \mathcal{S}' \times \mathcal{S}''$ .

**Theorem 2.7** Consider probability measures  $P_n$  and  $P$  on a separable metric space  $\mathbf{S} = \mathbf{S}' \times \mathbf{S}''$ .

(a)  $P_n \Rightarrow P$  implies  $P'_n \Rightarrow P'$  and  $P''_n \Rightarrow P''$ .

(b)  $P_n \Rightarrow P$  if and only if  $P_n(A' \times A'') \rightarrow P(A' \times A'')$  for each  $P'$ -continuity set  $A'$  and each  $P''$ -continuity set  $A''$ .

(c)  $P'_n \times P''_n \Rightarrow P$  if and only if  $P'_n \Rightarrow P'$ ,  $P''_n \Rightarrow P''$ , and  $P = P' \times P''$ .

Proof. (a) Since  $P' = P(\pi')^{-1}$ ,  $P'' = P(\pi'')^{-1}$  and the projections  $\pi'$ ,  $\pi''$  are continuous, it follows by the mapping theorem that  $P_n \Rightarrow P$  implies  $P'_n \Rightarrow P'$  and  $P''_n \Rightarrow P''$ .

(b) Consider the  $\pi$ -system  $\mathcal{A}$  of measurable rectangles  $A' \times A''$ :  $A' \in \mathcal{S}'$  and  $A'' \in \mathcal{S}''$ . Let  $\mathcal{A}_P$  be the class of  $A' \times A'' \in \mathcal{A}$  such that  $P'(\partial A') = P''(\partial A'') = 0$ . Since

$$\partial(A' \cap B') \subset (\partial A') \cup (\partial B'), \quad \partial(A'' \cap B'') \subset (\partial A'') \cup (\partial B''),$$



it follows that  $\mathcal{A}_P$  is a  $\pi$ -system:

$$A' \times A'', B' \times B'' \in \mathcal{A}_P \Rightarrow (A' \times A'') \cap (B' \times B'') \in \mathcal{A}_P.$$

And since

$$\partial(A' \times A'') \subset ((\partial A') \times S'') \cup (S' \times (\partial A'')),$$

each set in  $\mathcal{A}_P$  is a  $P$ -continuity set. Since  $B'(x', r)$  have disjoint boundaries for different values of  $r$ , and since the same is true of the  $B''(x'', r)$ , there are arbitrarily small  $r$  for which  $B(x, r) = B'(x', r) \times B''(x'', r)$  lies in  $\mathcal{A}_P$ . It follows that Theorem 2.3 applies to  $\mathcal{A}_P$ :  $P_n \Rightarrow P$  if and only if  $P_n(A) \rightarrow P(A)$  for each  $A \in \mathcal{A}_P$ .

The statement (c) is a consequence of (b).

**Exercise 2.8** The uniform distribution on the unit square and the uniform distribution on its diagonal have identical marginal distributions. Use this fact to demonstrate that the reverse to (a) in Theorem 2.7 is false.

**Exercise 2.9** Let  $(X_n, Y_n)$  be a sequence of two-dimensional random vectors. Show that if  $(X_n, Y_n) \Rightarrow (X, Y)$ , then besides  $X_n \Rightarrow X$  and  $Y_n \Rightarrow Y$ , we have  $X_n + Y_n \Rightarrow X + Y$ .

Give an example of  $(X_n, Y_n)$  such that  $X_n \Rightarrow X$  and  $Y_n \Rightarrow Y$  but the sum  $X_n + Y_n$  has no limit distribution.

## 2.3 Weak convergence in $\mathbf{R}^k$ and $\mathbf{R}^\infty$

Let  $\mathbf{R}^k$  denote the  $k$ -dimensional Euclidean space with elements  $x = (x_1, \dots, x_k)$  and the ordinary metric

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + \dots + (x_k - y_k)^2}.$$

Denote by  $\mathcal{R}^k$  the corresponding class of  $k$ -dimensional Borel sets. Put  $A_x = \{y : y_1 \leq x_1, \dots, y_k \leq x_k\}$ ,  $x \in \mathbf{R}^k$ . The probability measures on  $(\mathbf{R}^k, \mathcal{R}^k)$  are completely determined by their distribution functions  $F(x) = P(A_x)$  at the points of continuity  $x \in \mathbf{R}^k$ .

**Lemma 2.10** *The Weierstrass M-test. Suppose that sequences of real numbers  $x_i^n \rightarrow x_i$  converge for each  $i$ , and for all  $(n, i)$ ,  $|x_i^n| \leq M_i$ , where  $\sum_i M_i < \infty$ . Then  $\sum_i x_i < \infty$ ,  $\sum_i x_i^n < \infty$ , and  $\sum_i x_i^n \rightarrow \sum_i x_i$ .*

Proof. The series of course converge absolutely, since  $\sum_i M_i < \infty$ . Now for any  $(n, i_0)$ ,

$$\left| \sum_i x_i^n - \sum_i x_i \right| \leq \sum_{i \leq i_0} |x_i^n - x_i| + 2 \sum_{i > i_0} M_i.$$

Given  $\epsilon > 0$ , choose  $i_0$  so that  $\sum_{i > i_0} M_i < \epsilon/3$ , and then choose  $n_0$  so that  $n > n_0$  implies  $|x_i^n - x_i| < \frac{\epsilon}{3i_0}$  for  $i \leq i_0$ . Then  $n > n_0$  implies  $|\sum_i x_i^n - \sum_i x_i| < \epsilon$ .

**Lemma 2.11** *Let  $\mathbf{R}^\infty$  denote the space of the sequences  $x = (x_1, x_2, \dots)$  of real numbers with metric*

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

*Then  $\rho(x^n, x) \rightarrow 0$  if and only if  $|x_i^n - x_i| \rightarrow 0$  for each  $i$ .*

Proof. If  $\rho(x^n, x) \rightarrow 0$ , then for each  $i$  we have  $1 \wedge |x_i^n - x_i| \rightarrow 0$  and therefore  $|x_i^n - x_i| \rightarrow 0$ . The reverse implication holds by Lemma 2.10.

**Definition 2.12** Let  $\pi_k : \mathbf{R}^\infty \rightarrow \mathbf{R}^k$  be the natural projections  $\pi_k(x) = (x_1, \dots, x_k)$ ,  $k = 1, 2, \dots$ , and let  $P$  be a probability measure on  $(\mathbf{R}^\infty, \mathcal{R}^\infty)$ . The probability measures  $P\pi_k^{-1}$  defined on  $(\mathbf{R}^k, \mathcal{R}^k)$  are called the *finite-dimensional distributions* of  $P$ .

**Theorem 2.13** *The space  $\mathbf{R}^\infty$  is separable and complete. Let  $P$  and  $Q$  be two probability measures on  $(\mathbf{R}^\infty, \mathcal{R}^\infty)$ . If  $P\pi_k^{-1} = Q\pi_k^{-1}$  for each  $k$ , then  $P = Q$ .*

Proof. Convergence in  $\mathbf{R}^\infty$  implies coordinatewise convergence, therefore  $\pi_k$  is continuous so that the sets

$$B_k(x, \epsilon) = \{y \in \mathbf{R}^\infty : |y_i - x_i| < \epsilon, i = 1, \dots, k\} = \pi_k^{-1}\{y \in \mathbf{R}^k : |y_i - x_i| < \epsilon, i = 1, \dots, k\}$$

are open. Moreover,  $y \in B_k(x, \epsilon)$  implies  $\rho(x, y) < \epsilon + 2^{-k}$ . Thus  $B_k(x, \epsilon) \subset B(x, r)$  for  $r > \epsilon + 2^{-k}$ . This means that the sets  $B_k(x, \epsilon)$  form a base for the topology of  $\mathbf{R}^\infty$ . It follows that the space is separable: one countable, dense subset consists of those points having only finitely many nonzero coordinates, each of them rational.

If  $(x^n)$  is a fundamental sequence, then each coordinate sequence  $(x_i^n)$  is fundamental and hence converges to some  $x_i$ , implying  $x^n \rightarrow x$ . Therefore,  $\mathbf{R}^\infty$  is also complete.

Let  $\mathcal{A}$  be the class of finite-dimensional sets  $\{x : \pi_k(x) \in H\}$  for some  $k$  and some  $H \in \mathcal{R}^k$ . This class of cylinders is closed under finite intersections. To be able to apply Lemma 2.2 it remains to observe that  $\mathcal{A}$  generates  $\mathcal{R}^\infty$ : by separability each open set  $G \subset \mathbf{R}^\infty$  is a countable union of sets in  $\mathcal{A}$ , since the sets  $B_k(x, \epsilon) \in \mathcal{A}$  form a base.

**Theorem 2.14** Let  $P_n, P$  be probability measures on  $(\mathbf{R}^\infty, \mathcal{R}^\infty)$ . Then  $P_n \Rightarrow P$  if and only if  $P_n \pi_k^{-1} \Rightarrow P \pi_k^{-1}$  for each  $k$ .

Proof. Necessity follows from the mapping theorem. Turning to sufficiency, let  $\mathcal{A}$ , again, be the class of finite-dimensional sets  $\{x : \pi_k(x) \in H\}$  for some  $k$  and some  $H \in \mathcal{R}^k$ . We proceed in three steps.

Step 1. Show that  $\mathcal{A}$  is a convergence-determining class. This is proven using Theorem 2.4. Given  $x$  and  $\epsilon$ , choose  $k$  so that  $2^{-k} < \epsilon/2$  and consider the collection of uncountably many finite-dimensional sets

$$A_\eta = \{y : |y_i - x_i| < \eta, i = 1, \dots, k\} \text{ for } 0 < \eta < \epsilon/2.$$

We have  $A_\eta \in \mathcal{A}_{x, \epsilon}$ . On the other hand,  $\partial A_\eta$  consists of the points  $y$  such that  $|y_i - x_i| \leq \eta$  with equality for some  $i$ , hence these boundaries are disjoint. And since  $\mathbf{R}^\infty$  is separable, Theorem 2.4 applies.

Step 2. Show that  $\partial(\pi_k^{-1}H) = \pi_k^{-1}\partial H$ .

From the continuity of  $\pi_k$  it follows that  $\partial(\pi_k^{-1}H) \subset \pi_k^{-1}\partial H$ . Using special properties of the projections we can prove inclusion in the other direction. If  $x \in \pi_k^{-1}\partial H$ , so that  $\pi_k x \in \partial H$ , then there are points  $\alpha^{(u)} \in H$ ,  $\beta^{(u)} \in H^c$  such that  $\alpha^{(u)} \rightarrow \pi_k x$  and  $\beta^{(u)} \rightarrow \pi_k x$  as  $u \rightarrow \infty$ . Since the points  $(\alpha_1^{(u)}, \dots, \alpha_k^{(u)}, x_{k+1}, \dots)$  lie in  $\pi_k^{-1}H$  and converge to  $x$ , and since the points  $(\beta_1^{(u)}, \dots, \beta_k^{(u)}, x_{k+1}, \dots)$  lie in  $(\pi_k^{-1}H)^c$  and converge to  $x$ , we conclude that  $x \in \partial(\pi_k^{-1}H)$ .

Step 3. Suppose that  $P\pi_k^{-1}(\partial H) = 0$  implies  $P_n\pi_k^{-1}(H) \rightarrow P\pi_k^{-1}(H)$  and show that  $P_n \Rightarrow P$ .

If  $A \in \mathcal{A}$  is a finite-dimensional  $P$ -continuity set, then we have  $A = \pi_k^{-1}H$  and

$$P\pi_k^{-1}(\partial H) = P(\pi_k^{-1}\partial H) = P(\partial\pi_k^{-1}H) = P(\partial A) = 0.$$

Thus by assumption,  $P_n(A) \rightarrow P(A)$  and according to step 1,  $P_n \Rightarrow P$ .

## 2.4 Kolmogorov's extension theorem

**Definition 2.15** We say that the system of finite-dimensional distributions  $\mu_{t_1, \dots, t_k}$  is consistent if the joint distribution functions

$$F_{t_1, \dots, t_k}(z_1, \dots, z_k) = \mu_{t_1, \dots, t_k}((-\infty, z_1] \times \dots \times (-\infty, z_k])$$

satisfy two consistency conditions

- (i)  $F_{t_1, \dots, t_k, t_{k+1}}(z_1, \dots, z_k, \infty) = F_{t_1, \dots, t_k}(z_1, \dots, z_k)$ ,
- (ii) if  $\pi$  is a permutation of  $(1, \dots, k)$ , then

$$F_{t_{\pi(1)}, \dots, t_{\pi(k)}}(z_{\pi(1)}, \dots, z_{\pi(k)}) = F_{t_1, \dots, t_k}(z_1, \dots, z_k).$$

**Theorem 2.16** Let  $\mu_{t_1, \dots, t_k}$  be a consistent system of finite-dimensional distributions. Put  $\Omega = \{\text{functions } \omega : [0, 1] \rightarrow \mathbb{R}\}$  and  $\mathcal{F}$  is the  $\sigma$ -algebra generated by the finite-dimensional sets  $\{\omega : \omega(t_i) \in B_i, i = 1, \dots, n\}$ , where  $B_i$  are Borel subsets of  $\mathbb{R}$ . Then there is a unique probability measure  $\mathbb{P}$  on  $(\Omega, \mathcal{F})$  such that a stochastic process defined by  $X_t(\omega) = \omega(t)$  has the finite-dimensional distributions  $\mu_{t_1, \dots, t_k}$ .

Without proof. Kolmogorov's extension theorem does not directly imply the existence of the Wiener process because the  $\sigma$ -algebra  $\mathcal{F}$  is not rich enough to ensure the continuity property for trajectories. However, it is used in the proof of Theorem 7.17 establishing the existence of processes with cadlag trajectories.

## 3 Tightness and Prokhorov's theorem

### 3.1 Tightness of probability measures

Convergence of finite-dimensional distributions does not always imply weak convergence. This makes important the following concept of tightness.

**Definition 3.1** A family of probability measures  $\Pi$  on  $(\mathcal{S}, \mathcal{S})$  is called *tight* if for every  $\epsilon$  there exists a compact set  $K \subset \mathcal{S}$  such that  $P(K) > 1 - \epsilon$  for all  $P \in \Pi$ .

**Lemma 3.2** If  $\mathcal{S}$  is separable and complete, then each probability measure  $P$  on  $(\mathcal{S}, \mathcal{S})$  is tight.

Proof. Separability: for each  $k$  there is a sequence  $A_{k,i}$  of open  $1/k$ -balls covering  $\mathcal{S}$ . Choose  $n_k$  large enough that  $P(B_k) > 1 - \epsilon 2^{-k}$  where  $B_k = A_{k,1} \cup \dots \cup A_{k,n_k}$ . Completeness: the totally bounded set  $B_1 \cap B_2 \cap \dots$  has compact closure  $K$ . But clearly  $P(K^c) \leq \sum_k P(B_k^c) < \epsilon$ .

**Exercise 3.3** Check whether the following sequence of distributions on  $\mathbf{R}$

$$P_n(A) = (1 - n^{-1})1_{\{0 \in A\}} + n^{-1}1_{\{n^2 \in A\}}, \quad n \geq 1,$$

is tight or it “leaks” towards infinity. Notice that the corresponding mean value is  $n$ .

**Definition 3.4** A family of probability measures  $\Pi$  on  $(\mathbf{S}, \mathcal{S})$  is called *relatively compact* if any sequence of its elements contains a weakly convergent subsequence. The limiting probability measures might be different for different subsequences and lie outside  $\Pi$ .

**Definition 3.5** Let  $\mathbf{P}$  be the space of probability measures on  $(\mathbf{S}, \mathcal{S})$ . The *Prokhorov distance*  $\pi(P, Q)$  between  $P, Q \in \mathbf{P}$  is defined as the infimum of those positive  $\epsilon$  for which

$$P(A) \leq Q(A^\epsilon) + \epsilon, \quad Q(A) \leq P(A^\epsilon) + \epsilon, \quad \text{for all } A \in \mathcal{S}.$$

**Lemma 3.6** The Prokhorov distance  $\pi$  is a metric on  $\mathbf{P}$ .

Proof. Obviously  $\pi(P, Q) = \pi(Q, P)$  and  $\pi(P, P) = 0$ . If  $\pi(P, Q) = 0$ , then for any  $F \in \mathcal{S}$  and  $\epsilon > 0$ ,  $P(F) \leq Q(F^\epsilon) + \epsilon$ . For closed  $F$  letting  $\epsilon \rightarrow 0$  gives  $P(F) \leq Q(F)$ . By symmetry, we have  $P(F) = Q(F)$  implying  $P = Q$ .

To verify the triangle inequality notice that if  $\pi(P, Q) < \epsilon_1$  and  $\pi(Q, R) < \epsilon_2$ , then

$$P(A) \leq Q(A^{\epsilon_1}) + \epsilon_1 \leq R((A^{\epsilon_1})^{\epsilon_2}) + \epsilon_1 + \epsilon_2 \leq R(A^{\epsilon_1 + \epsilon_2}) + \epsilon_1 + \epsilon_2.$$

Thus, using the symmetric relation we obtain  $\pi(P, R) < \epsilon_1 + \epsilon_2$ . Therefore,  $\pi(P, R) \leq \pi(P, Q) + \pi(Q, R)$ .

**Theorem 3.7** Suppose  $\mathbf{S}$  is a complete separable metric space. Then weak convergence is equivalent to  $\pi$ -convergence,  $(\mathbf{P}, \pi)$  is separable and complete, and  $\Pi \subset \mathbf{P}$  is relatively compact iff its  $\pi$ -closure is  $\pi$ -compact.

Without proof.

**Theorem 3.8** A necessary and sufficient condition for  $P_n \Rightarrow P$  is that each subsequence  $P_{n'}$  contains a further subsequence  $P_{n''}$  converging weakly to  $P$ .

Proof. The necessity is easy but useless. As for sufficiency, if  $P_n \not\Rightarrow P$ , then  $\int_{\mathbf{S}} f(x) P_n(dx) \not\rightarrow \int_{\mathbf{S}} f(x) P(dx)$  for some bounded, continuous  $f$ . But then, for some  $\epsilon > 0$  and some subsequence  $P_{n'}$ ,

$$\left| \int_{\mathbf{S}} f(x) P_{n'}(dx) - \int_{\mathbf{S}} f(x) P(dx) \right| \geq \epsilon \quad \text{for all } n',$$

and no further subsequence can converge weakly to  $P$ .

**Theorem 3.9** Prokhorov's theorem, the direct part. If a family of probability measures  $\Pi$  on  $(\mathbf{S}, \mathcal{S})$  is tight, then it is relatively compact.

Proof. See the next subsection.

**Theorem 3.10** Prokhorov's theorem, the reverse part. Suppose  $\mathbf{S}$  is a complete separable metric space. If  $\Pi$  is relatively compact, then it is tight.

Proof. Consider open sets  $G_n \uparrow \mathbf{S}$ . For each  $\epsilon$  there is an  $n$  such that  $P(G_n) > 1 - \epsilon$  for all  $P \in \Pi$ . To show this we assume the opposite:  $P_n(G_n) \leq 1 - \epsilon$  for some  $P_n \in \Pi$ . By the assumed relative compactness,  $P_{n'} \Rightarrow Q$  for some subsequence and some probability measure  $Q$ . Then

$$Q(G_n) \leq \liminf_{n'} P_{n'}(G_n) \leq \liminf_{n'} P_{n'}(G_{n'}) \leq 1 - \epsilon$$

which is impossible since  $G_n \uparrow \mathbf{S}$ .

If  $A_{ki}$  is a sequence of open balls of radius  $1/k$  covering  $\mathbf{S}$  (separability), so that  $\mathbf{S} = \cup_i A_{k,i}$  for each  $k$ . From the previous step, it follows that there is an  $n_k$  such that  $P(\cup_{i \leq n_k} A_{k,i}) > 1 - \epsilon 2^{-k}$  for all  $P \in \Pi$ . Let  $K$  be the closure of the totally bounded set  $\cap_{k \geq 1} \cup_{i \leq n_k} A_{k,i}$ , then  $K$  is compact (completeness) and  $P(K) > 1 - \epsilon$  for all  $P \in \Pi$ .

### 3.2 Proof of Prokhorov's theorem

This subsection contains a proof of the direct half of Prokhorov's theorem. Let  $(P_n)$  be a sequence in the tight family  $\Pi$ . We are to find a subsequence  $(P_{n'})$  and a probability measure  $P$  such that  $P_{n'} \Rightarrow P$ . The proof, like that of Helly's selection theorem will depend on a diagonal argument.

Choose compact sets  $K_1 \subset K_2 \subset \dots$  such that  $P_n(K_i) > 1 - i^{-1}$  for all  $n$  and  $i$ . The set  $K_\infty = \cup_i K_i$  is separable: compactness = each open cover has a finite subcover, separability = each open cover has a countable subcover. Hence, by Theorem 1.7, there exists a countable class  $\mathcal{A}$  of open sets with the following property: if  $G$  is open and  $x \in K_\infty \cap G$ , then  $x \in A \subset A^- \subset G$  for some  $A \in \mathcal{A}$ . Let  $\mathcal{H}$  consist of  $\emptyset$  and the finite unions of sets of the form  $A^- \cap K_i$  for  $A \in \mathcal{A}$  and  $i \geq 1$ .

Consider the countable class  $\mathcal{H} = (H_j)$ . For  $(P_n)$  there is a subsequence  $(P_{n_1})$  such that  $P_{n_1}(H_1)$  converges as  $n_1 \rightarrow \infty$ . For  $(P_{n_1})$  there is a further subsequence  $(P_{n_2})$  such that  $P_{n_2}(H_2)$  converges as  $n_2 \rightarrow \infty$ . Continuing in this way we get a collection of indices  $(n_{1k}) \supset (n_{2k}) \supset \dots$  such that  $P_{n_{jk}}(H_j)$  converges as  $k \rightarrow \infty$  for each  $j \geq 1$ . Putting  $n'_j = n_{jj}$  we find a subsequence  $(P_{n'})$  for which the limit

$$\alpha(H) = \lim_{n'} P_{n'}(H) \text{ exists for each } H \in \mathcal{H}.$$

Furthermore, for open sets  $G \subset \mathcal{S}$  and arbitrary sets  $M \subset \mathcal{S}$  define

$$\beta(G) = \sup_{H \subset G} \alpha(H), \quad \gamma(M) = \inf_{G \supset M} \beta(G).$$

Our objective is to construct on  $(\mathcal{S}, \mathcal{S})$  a probability measure  $P$  such that  $P(G) = \beta(G)$  for all open sets  $G$ . If there does exist such a  $P$ , then the proof will be complete: if  $H \subset G$ , then

$$\alpha(H) = \lim_{n'} P_{n'}(H) \leq \liminf_{n'} P_{n'}(G),$$

whence  $P(G) \leq \liminf_{n'} P_{n'}(G)$ , and therefore  $P_{n'} \Rightarrow P$ . The construction of the probability measure  $P$  is divided in seven steps.

Step 1: if  $F \subset G$ , where  $F$  is closed and  $G$  is open, and if  $F \subset H$ , for some  $H \in \mathcal{H}$ , then  $F \subset H_0 \subset G$ , for some  $H_0 \in \mathcal{H}$ .

Since  $F \subset K_{i_0}$  for some  $i_0$ , the closed set  $F$  is compact. For each  $x \in F$ , choose an  $A_x \in \mathcal{A}$  such that  $x \in A_x \subset A_x^- \subset G$ . The sets  $A_x$  cover the compact  $F$ , and there is a finite subcover  $A_{x_1}, \dots, A_{x_k}$ . We can take  $H_0 = \cup_{j=1}^k (A_{x_j}^- \cap K_{i_0})$ .

Step 2:  $\beta$  is finitely subadditive on the open sets.

Suppose that  $H \subset G_1 \cup G_2$ , where  $H \in \mathcal{H}$  and  $G_1, G_2$  are open. Define

$$\begin{aligned} F_1 &= \{x \in H : \rho(x, G_1^c) \geq \rho(x, G_2^c)\}, \\ F_2 &= \{x \in H : \rho(x, G_2^c) \geq \rho(x, G_1^c)\}, \end{aligned}$$

so that  $H = F_1 \cup F_2$  with  $F_1 \subset G_1$  and  $F_2 \subset G_2$ . According to Step 1, since  $F_i \subset H$ , we have  $F_i \subset H_i \subset G_i$  for some  $H_i \in \mathcal{H}$ .

The function  $\alpha(H)$  has these three properties

$$\begin{aligned} \alpha(H_1) &\leq \alpha(H_2) && \text{if } H_1 \subset H_2, \\ \alpha(H_1 \cup H_2) &= \alpha(H_1) + \alpha(H_2) && \text{if } H_1 \cap H_2 = \emptyset, \\ \alpha(H_1 \cup H_2) &\leq \alpha(H_1) + \alpha(H_2). \end{aligned}$$

It follows first,

$$\alpha(H) \leq \alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2) \leq \beta(G_1) + \beta(G_2),$$

and then

$$\beta(G_1 \cup G_2) = \sup_{H \subset G_1 \cup G_2} \alpha(H) \leq \beta(G_1) + \beta(G_2).$$

Step 3:  $\beta$  is countably subadditive on the open sets.

If  $H \subset \cup_n G_n$ , then, since  $H$  is compact,  $H \subset \cup_{n \leq n_0} G_n$  for some  $n_0$ , and finite subadditivity implies

$$\alpha(H) \leq \sum_{n \leq n_0} \beta(G_n) \leq \sum_n \beta(G_n).$$

Taking the supremum over  $H$  contained in  $\cup_n G_n$  gives  $\beta(\cup_n G_n) \leq \sum_n \beta(G_n)$ .

Step 4:  $\gamma$  is an outer measure.

Since  $\gamma$  is clearly monotone and satisfies  $\gamma(\emptyset) = 0$ , we need only prove that it is countably subadditive. Given a positive  $\epsilon$  and arbitrary  $M_n \subset \mathbf{S}$ , choose open sets  $G_n$  such that  $M_n \subset G_n$  and  $\beta(G_n) < \gamma(M_n) + \epsilon/2^n$ . Apply Step 3

$$\gamma\left(\bigcup_n M_n\right) \leq \beta\left(\bigcup_n G_n\right) \leq \sum_n \beta(G_n) \leq \sum_n \gamma(M_n) + \epsilon,$$

and let  $\epsilon \rightarrow 0$  to get  $\gamma\left(\bigcup_n M_n\right) \leq \sum_n \gamma(M_n)$ .

Step 5:  $\beta(G) \geq \gamma(F \cap G) + \gamma(F^c \cap G)$  for  $F$  closed and  $G$  open.

Choose  $H_3, H_4 \in \mathcal{H}$  for which

$$\begin{aligned} H_3 &\subset F^c \cap G & \text{and} & & \alpha(H_3) &> \beta(F^c \cap G) - \epsilon, \\ H_4 &\subset H_3^c \cap G & \text{and} & & \alpha(H_4) &> \beta(H_3^c \cap G) - \epsilon. \end{aligned}$$

Since  $H_3$  and  $H_4$  are disjoint and are contained in  $G$ , it follows from the properties of the functions  $\alpha, \beta$ , and  $\gamma$  that

$$\begin{aligned} \beta(G) &\geq \alpha(H_3 \cup H_4) = \alpha(H_3) + \alpha(H_4) > \beta(F^c \cap G) + \beta(H_3^c \cap G) - 2\epsilon \\ &\geq \gamma(F^c \cap G) + \gamma(F \cap G) - 2\epsilon. \end{aligned}$$

Now it remains to let  $\epsilon \rightarrow 0$ .

Step 6: if  $F \subset \mathbf{S}$  is closed, then  $F$  is in the class  $\mathcal{M}$  of  $\gamma$ -measurable sets.

By Step 5,  $\beta(G) \geq \gamma(F \cap L) + \gamma(F^c \cap L)$  if  $F$  is closed,  $G$  is open, and  $G \supset L$ . Taking the infimum over these  $G$  gives  $\gamma(L) \geq \gamma(F \cap L) + \gamma(F^c \cap L)$  confirming that  $F$  is  $\gamma$ -measurable.

Step 7:  $\mathcal{S} \subset \mathcal{M}$ , and the restriction  $P$  of  $\gamma$  to  $\mathcal{S}$  is a probability measure satisfying  $P(G) = \gamma(G) = \beta(G)$  for all open sets  $G \subset \mathbf{S}$ .

Since each closed set lies in  $\mathcal{M}$  and  $\mathcal{M}$  is a  $\sigma$ -algebra, we have  $\mathcal{S} \subset \mathcal{M}$ . To see that the  $P$  is a probability measure, observe that each  $K_i$  has a finite covering by  $\mathcal{A}$ -sets and therefore  $K_i \in \mathcal{H}$ . Thus

$$1 \geq P(\mathbf{S}) = \beta(\mathbf{S}) \geq \sup_i \alpha(K_i) \geq \sup_i (1 - i^{-1}) = 1.$$

### 3.3 Skorokhod's representation theorem

**Theorem 3.11** *Suppose that  $P_n \Rightarrow P$  and  $P$  has a separable support. Then there exist random elements  $X_n$  and  $X$ , defined on a common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , such that  $P_n$  is the probability distribution of  $X_n$ ,  $P$  is the probability distribution of  $X$ , and  $X_n(\omega) \rightarrow X(\omega)$  for every  $\omega$ .*

*Proof.* We split the proof in four steps.

Step 1: show that for each  $\epsilon$ , there is a finite  $\mathcal{S}$ -partition  $B_0, B_1, \dots, B_k$  of  $\mathbf{S}$  such that

$$0 < P(B_0) < \epsilon, \quad P(\partial B_i) = 0, \quad \text{diam}(B_i) < \epsilon, \quad i = 1, \dots, k.$$

Let  $M$  be a separable  $\mathcal{S}$ -set for which  $P(M) = 1$ . For each  $x \in M$ , choose  $r_x$  so that  $0 < r_x < \epsilon/2$  and  $P(\partial B(x, r_x)) = 0$ . Since  $M$  is a separable, it can be covered by a countable subcollection  $A_1, A_2, \dots$  of the balls  $B(x, r_x)$ . Choose  $k$  so that  $P(\cup_{i=1}^k A_i) > 1 - \epsilon$ . Take

$$B_0 = \left(\bigcup_{i=1}^k A_i\right)^c, \quad B_1 = A_1, \quad B_i = A_1^c \cap \dots \cap A_{i-1}^c \cap A_i,$$

and notice that  $\partial B_i \subset \partial A_1 \cup \dots \cup \partial A_k$ .

Step 2: definition of  $n_j$ .

Take  $\epsilon_j = 2^{-j}$ . By step 1, there are  $\mathcal{S}$ -partitions  $B_0^j, B_1^j, \dots, B_{k_j}^j$  such that

$$0 < P(B_0^j) < \epsilon_j, \quad P(\partial B_i^j) = 0, \quad \text{diam}(B_i^j) < \epsilon_j, \quad i = 1, \dots, k_j.$$

If some  $P(B_i^j) = 0$ , we redefine these partitions by amalgamating such  $B_i^j$  with  $B_0^j$ , so that  $P(\cdot | B_i^j)$  is well defined for  $i \geq 1$ . By the assumption  $P_n \Rightarrow P$ , there is for each  $j$  an  $n_j$  such that

$$P_n(B_i^j) \geq (1 - \epsilon_j)P(B_i^j), \quad i = 0, 1, \dots, k_j, \quad n \geq n_j.$$

Putting  $n_0 = 1$ , we can assume  $n_0 < n_1 < \dots$ .

Step 3: construction of  $X, Y_n, Y_{ni}, Z_n, \xi$ .

Define  $m_n = j$  for  $n_j \leq n < n_{j+1}$  and write  $m$  instead of  $m_n$ . By Theorem 2.16 we can find an  $(\Omega, \mathcal{F}, \mathbb{P})$  supporting random elements  $X, Y_n, Y_{ni}, Z_n$  of  $\mathcal{S}$  and a random variable  $\xi$ , all independent of each other and having distributions satisfying:  $X$  has distribution  $P$ ,  $Y_n$  has distribution  $P_n$ ,

$$\begin{aligned}\mathbb{P}(Y_{ni} \in A) &= P_n(A|B_i^m), \quad \mathbb{P}(\xi \leq \epsilon) = \epsilon, \\ \epsilon_m \mathbb{P}(Z_n \in A) &= \sum_{i=0}^{k_m} P_n(A|B_i^m) \left( P_n(B_i^m) - (1 - \epsilon_m)P(B_i^m) \right).\end{aligned}$$

Note that  $\mathbb{P}(Y_{ni} \in B_i^m) = 1$ .

Step 4: construction of  $X_n$ .

Put  $X_n = Y_n$  for  $n < n_1$ . For  $n \geq n_1$ , put

$$X_n = 1_{\{\xi \leq 1 - \epsilon_m\}} \sum_{i=0}^{k_m} 1_{\{X \in B_i^m\}} Y_{ni} + 1_{\{\xi > 1 - \epsilon_m\}} Z_n.$$

By step 3, we  $X_n$  has distribution  $P_n$  because

$$\begin{aligned}\mathbb{P}(X_n \in A) &= (1 - \epsilon_m) \sum_{i=0}^{k_m} \mathbb{P}(X \in B_i^m, Y_{ni} \in A) + \epsilon_m \mathbb{P}(Z_n \in A) \\ &= (1 - \epsilon_m) \sum_{i=0}^{k_m} \mathbb{P}(X \in B_i^m) P_n(A|B_i^m) \\ &\quad + \sum_{i=0}^{k_m} P_n(A|B_i^m) \left( P_n(B_i^m) - (1 - \epsilon_m)P(B_i^m) \right) \\ &= P_n(A).\end{aligned}$$

Let

$$E_j = \{X \notin B_0^j; \xi \leq 1 - \epsilon_j\} \text{ and } E = \liminf_j E_j = \bigcup_{j=1}^{\infty} \bigcap_{i=j}^{\infty} E_i.$$

Since  $\mathbb{P}(E_j^c) < 2\epsilon_j$ , by the Borel-Cantelli lemma,  $\mathbb{P}(E^c) = \mathbb{P}(E_j^c \text{ i.o.}) = 0$  implying  $\mathbb{P}(E) = 1$ . If  $\omega \in E$ , then both  $X_n(\omega)$  and  $X(\omega)$  lie in the same  $B_i^m$  having diameter less than  $\epsilon_m$ . Thus,  $\rho(X_n(\omega), X(\omega)) < \epsilon_m$  and  $X_n(\omega) \rightarrow X(\omega)$  for  $\omega \in E$ . It remains to redefine  $X_n$  as  $X$  outside  $E$ .

**Corollary 3.12** *The mapping theorem. Let  $h : \mathcal{S} \rightarrow \mathcal{S}'$  be a continuous mapping between two metric spaces. If  $P_n \Rightarrow P$  on  $\mathcal{S}$  and  $P$  has a separable support, then  $P_n h^{-1} \Rightarrow P h^{-1}$  on  $\mathcal{S}'$ .*

Proof. Having  $X_n(\omega) \rightarrow X(\omega)$  we get  $h(X_n(\omega)) \rightarrow h(X(\omega))$  for every  $\omega$ . It follows, by Corollary 1.24 that  $h(X_n) \Rightarrow h(X)$  which is equivalent to  $P_n h^{-1} \Rightarrow P h^{-1}$ .

## 4 Functional Central Limit Theorem on $\mathcal{C} = \mathcal{C}[0, 1]$

### 4.1 Weak convergence in $\mathcal{C}$

**Definition 4.1** An element of the set  $\mathcal{C} = \mathcal{C}[0, 1]$  is a continuous function  $x = x(t)$ . The distance between points in  $\mathcal{C}$  is measured by the uniform metric

$$\rho(x, y) = \|x - y\| = \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

Denote by  $\mathcal{C}$  the Borel  $\sigma$ -algebra of subsets of  $\mathcal{C}$ .

**Exercise 4.2** Draw a picture for an open ball  $B(x, r)$  in  $\mathcal{C}$ .

For any real number  $a$  and  $t \in [0, 1]$  the set  $\{x : x(t) < a\}$  is an open subset of  $\mathcal{C}$ .

**Example 4.3** Convergence  $\rho(x_n, x) \rightarrow 0$  means uniform convergence of continuous functions, it is stronger than pointwise convergence. Consider the function  $z_n(t)$  that increases linearly from 0 to 1 over  $[0, n^{-1}]$ , decreases linearly from 1 to 0 over  $[n^{-1}, 2n^{-1}]$ , and equals 0 over  $[2n^{-1}, 1]$ . Despite  $z_n(t) \rightarrow 0$  for any  $t$  we have  $\|z_n\| = 1$  for all  $n$ .

**Theorem 4.4** *The space  $\mathcal{C}$  is separable and complete.*

Proof. Separability. Let  $L_k$  be the set of polygonal functions that are linear over each subinterval  $[\frac{i-1}{k}, \frac{i}{k}]$  and have rational values at the end points. We will show that the countable set  $\cup_{k \geq 1} L_k$  is dense in  $\mathbf{C}$ . For given  $x \in \mathbf{C}$  and  $\epsilon > 0$ , choose  $k$  so that

$$|x(t) - x(i/k)| < \epsilon \quad \text{for all } t \in [(i-1)/k, i/k], \quad 1 \leq i \leq k$$

which is possible by uniform continuity. Then choose  $y \in L_k$  so that  $|y(i/k) - x(i/k)| < \epsilon$  for each  $i$ . It remains to draw a picture with trajectories over an interval  $[\frac{i-1}{k}, \frac{i}{k}]$  and check that  $\rho(x, y) \leq 3\epsilon$ .

Completeness. Let  $(x_n)$  be a fundamental sequence so that

$$\epsilon_n = \sup_{m > n} \sup_{0 \leq t \leq 1} |x_n(t) - x_m(t)| \rightarrow 0, \quad n \rightarrow \infty.$$

Then for each  $t$ , the sequence  $(x_n(t))$  is fundamental on  $\mathbf{R}$  and hence has a limit  $x(t)$ . Letting  $m \rightarrow \infty$  in the inequality  $|x_n(t) - x_m(t)| \leq \epsilon_n$  gives  $|x_n(t) - x(t)| \leq \epsilon_n$ . Thus  $x_n$  converges uniformly to  $x \in \mathbf{C}$ .

**Definition 4.5** Convergence of finite-dimensional distributions  $X^n \xrightarrow{\text{fdd}} X$  means that for all  $t_1, \dots, t_k$

$$(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow (X_{t_1}, \dots, X_{t_k}).$$

**Exercise 4.6** The projection  $\pi_{t_1, \dots, t_k} : \mathbf{C} \rightarrow \mathbf{R}^k$  defined by  $\pi_{t_1, \dots, t_k}(x) = (x(t_1), \dots, x(t_k))$  is a continuous map.

**Example 4.7** By the mapping theorem, if  $X^n \Rightarrow X$ , then  $X^n \xrightarrow{\text{fdd}} X$ . The reverse is not true. Consider  $z_n(t)$  from Example 4.3 and put  $X^n = z_n$ ,  $X = 0$  so that  $X^n \xrightarrow{\text{fdd}} X$ . Take  $h(x) = \sup_t x(t)$ . It satisfies  $|h(x) - h(y)| \leq \rho(x, y)$  and therefore is a continuous function on  $\mathbf{C}$ . Since  $h(z_n) \equiv 1$ , we have  $h(X^n) \not\Rightarrow h(X)$ , and according to the mapping theorem  $X^n \not\Rightarrow X$ .

**Definition 4.8** Define a modulus of continuity of a function  $x : [0, 1] \rightarrow \mathbf{R}$  by

$$w_x(\delta) = w(x, \delta) = \sup_{|s-t| \leq \delta} |x(s) - x(t)|, \quad \delta \in (0, 1].$$

For any  $x : [0, 1] \rightarrow \mathbf{R}$  its modulus of continuity  $w_x(\delta)$  is non-decreasing over  $\delta$ . Clearly,  $x \in \mathbf{C}$  if and only if  $w_x(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ . The limit  $j_x = \lim_{\delta \rightarrow 0} w_x(\delta)$  is the absolute value of the largest jump of  $x$ .

**Exercise 4.9** Show that for any fixed  $\delta \in (0, 1]$  we have  $|w_x(\delta) - w_y(\delta)| \leq 2\rho(x, y)$  implying that  $w_x(\delta)$  is a continuous function on  $\mathbf{C}$ .

**Example 4.10** For  $z_n \in \mathbf{C}$  defined in Example 4.3 we have  $w(z_n, \delta) = 1$  for  $n \geq \delta^{-1}$ .

**Exercise 4.11** Given a probability measure  $P$  on the measurable space  $(\mathbf{C}, \mathcal{C})$  there exists a random process  $X$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  such that  $\mathbb{P}(X \in A) = P(A)$  for any  $A \in \mathcal{C}$ .

**Theorem 4.12** Let  $P_n, P$  be probability measures on  $(\mathbf{C}, \mathcal{C})$ . Suppose  $P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow P \pi_{t_1, \dots, t_k}^{-1}$  holds for all tuples  $(t_1, \dots, t_k) \subset [0, 1]$ . If for every positive  $\epsilon$

$$(i) \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P_n(x : w_x(\delta) \geq \epsilon) = 0,$$

then  $P_n \Rightarrow P$ .

Proof. The proof is given in terms of convergence in distribution using Theorem 1.23.

For  $u = 1, 2, \dots$ , define  $M_u : \mathbf{C} \rightarrow \mathbf{C}$  in the following way. Let  $(M_u x)(t)$  agree with  $x(t)$  at the points  $0, 1/u, 2/u, \dots, 1$  and be defined by linear interpolation between these points. Observe that  $\rho(M_u x, x) \leq 2w_x(1/u)$ .

Further, for a vector  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_u)$  define  $(L_u \alpha)(t)$  as an element of  $\mathbf{C}$  such that it has values  $\alpha_i$  at points  $t = i/u$  and is linear in between. Clearly,  $\rho(L_u \alpha, L_u \beta) = \max_i |\alpha_i - \beta_i|$ , so that  $L_u : \mathbf{R}^{u+1} \rightarrow \mathbf{C}$  is continuous.

Let  $t_i = i/u$ . Observe that  $M_u = L_u \pi_{t_0, \dots, t_u}$ . Since  $\pi_{t_0, \dots, t_u} X^n \Rightarrow \pi_{t_0, \dots, t_u} X$  and  $L_u$  is continuous, the mapping theorem gives  $M_u X^n \Rightarrow M_u X$  as  $n \rightarrow \infty$ . Since

$$\limsup_{u \rightarrow \infty} \rho(M_u X, X) \leq 2 \limsup_{u \rightarrow \infty} w(X, 1/u) = 0,$$

we have  $M_u X \rightarrow X$  in probability and therefore  $M_u X \Rightarrow X$ .

Finally, due to  $\rho(M_u X^n, X^n) \leq 2w(X^n, 1/u)$  and condition (i) we have

$$\limsup_{u \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(\rho(M_u X^n, X^n) \geq \epsilon) \leq \limsup_{u \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}(2w(X^n, 1/u) \geq \epsilon) = 0.$$

It remains to apply Theorem 1.23.

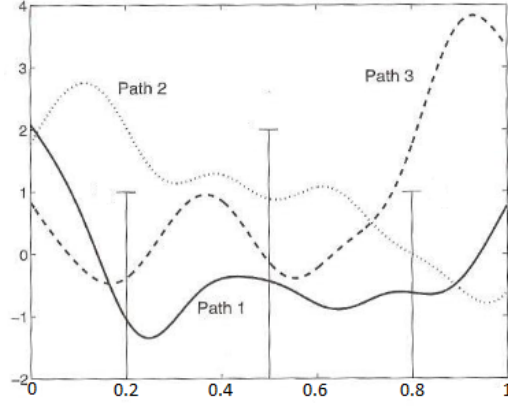


Figure 2: Cylinder sets.

**Lemma 4.13** Let  $P$  and  $Q$  be two probability measures on  $(\mathcal{C}, \mathcal{C})$ . If  $P\pi_{t_1, \dots, t_k}^{-1} = Q\pi_{t_1, \dots, t_k}^{-1}$  for all  $0 \leq t_1 < \dots < t_k \leq 1$ , then  $P = Q$ .

Proof. Denote by  $\mathcal{C}_f$  the collection of *cylinder sets* of the form

$$\pi_{t_1, \dots, t_k}^{-1}(H) = \{y \in \mathcal{C} : (y(t_1), \dots, y(t_k)) \in H\}, \quad (*)$$

where  $0 \leq t_1 < \dots < t_k \leq 1$  and a Borel subset  $H \subset \mathbf{R}^k$ . Due to the continuity of the projections we have  $\mathcal{C}_f \subset \mathcal{C}$ .

It suffices to check, using Lemma 2.2, that  $\mathcal{C}_f$  is a separating class. Clearly,  $\mathcal{C}_f$  is closed under formation of finite intersections. To show that  $\sigma(\mathcal{C}_f) = \mathcal{C}$ , observe that a closed ball centered at  $x$  of radius  $a$  can be represented as  $\cap_r (y : |y(r) - x(r)| \leq a)$ , where  $r$  ranges over rationals in  $[0, 1]$ . It follows that  $\sigma(\mathcal{C}_f)$  contains all closed balls, hence the open balls, and hence the  $\sigma$ -algebra generated by the open balls. By separability, the  $\sigma$ -algebra generated by the open balls, the so-called *ball  $\sigma$ -algebra*, coincides with the Borel  $\sigma$ -algebra generated by the open sets.

**Exercise 4.14** Which of the three paths on Figure 2 belong to the cylinder set  $(*)$  with  $k = 3$ ,  $t_1 = 0.2, t_2 = 0.5, t_3 = 0.8$ , and  $H = [-2, 1] \times [-2, 2] \times [-2, 1]$ .

**Theorem 4.15** Let  $P_n$  be probability measures on  $(\mathcal{C}, \mathcal{C})$ . If their finite-dimensional distributions converge weakly  $P_n\pi_{t_1, \dots, t_k}^{-1} \Rightarrow \mu_{t_1, \dots, t_k}$ , and if  $P_n$  is tight, then

- (a) there exists a probability measure  $P$  on  $(\mathcal{C}, \mathcal{C})$  with  $P\pi_{t_1, \dots, t_k}^{-1} = \mu_{t_1, \dots, t_k}$ , and
- (b)  $P_n \Rightarrow P$ .

Proof. Tightness implies relative compactness which in turn implies that each subsequence  $(P_{n'}) \subset (P_n)$  contains a further subsequence  $(P_{n''}) \subset (P_{n'})$  converging weakly to some probability measure  $P$ . By the mapping theorem  $P_{n''}\pi_{t_1, \dots, t_k}^{-1} \Rightarrow P\pi_{t_1, \dots, t_k}^{-1}$ . Thus by hypothesis,  $P\pi_{t_1, \dots, t_k}^{-1} = \mu_{t_1, \dots, t_k}$ . Moreover, by Lemma 4.13, the limit  $P$  must be the same for all converging subsequences, thus applying Theorem 3.8 we may conclude that  $P_n \Rightarrow P$ .

## 4.2 Wiener measure and Donsker's theorem

**Definition 4.16** Let  $\xi_i$  be a sequence of r.v. defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Put  $S_n = \xi_1 + \dots + \xi_n$  and let  $X_t^n(\omega)$  as a function of  $t$  be the element of  $\mathcal{C}$  defined by linear interpolation between its values  $X_{i/n}^n(\omega) = \frac{S_i(\omega)}{\sigma\sqrt{n}}$  at the points  $t = i/n$ .

**Theorem 4.17** Let  $X^n = (X_t^n : 0 \leq t \leq 1)$  be defined by Definition 4.16 and let  $P_n$  be the probability distribution of  $X^n$ . If  $\xi_i$  are iid with zero mean and finite variance  $\sigma^2$ , then

- (a)  $P_n\pi_{t_1, \dots, t_k}^{-1} \Rightarrow \mu_{t_1, \dots, t_k}$ , where  $\mu_{t_1, \dots, t_k}$  are Gaussian distributions on  $\mathbf{R}^k$  satisfying

$$\mu_{t_1, \dots, t_k} \left\{ (x_1, \dots, x_k) : x_i - x_{i-1} \leq \alpha_i, i = 1, \dots, k \right\} = \prod_{i=1}^k \Phi\left(\frac{\alpha_i}{\sqrt{t_i - t_{i-1}}}\right), \text{ where } x_0 = 0,$$

- (b) the sequence  $(P_n)$  of probability measures on  $(\mathcal{C}, \mathcal{C})$  is tight.



Proof. The claim (a) follows from the classical CLT and independence of increments of  $S_n$ . For example, if  $0 \leq s \leq t \leq 1$ , then

$$(X_s^n, X_t^n - X_s^n) = \frac{1}{\sigma\sqrt{n}}(S_{\lfloor ns \rfloor}, S_{\lfloor nt \rfloor} - S_{\lfloor ns \rfloor}) + \epsilon_{s,t}^n,$$

$$\epsilon_{s,t}^n = \frac{1}{\sigma\sqrt{n}}(\{ns\}\xi_{\lfloor ns \rfloor+1}, \{nt\}\xi_{\lfloor nt \rfloor+1} - \{ns\}\xi_{\lfloor ns \rfloor+1}),$$

where  $\{nt\}$  stands for the fractional part of  $nt$ . By the classical CLT and Theorem 2.7c,  $\frac{1}{\sigma\sqrt{n}}(S_{\lfloor ns \rfloor}, S_{\lfloor nt \rfloor} - S_{\lfloor ns \rfloor})$  has  $\mu_{s,t}$  as a limit distribution. Applying Corollary 1.24 to  $\epsilon_{s,t}^n$ , we derive  $P_n \pi_{s,t}^{-1} \Rightarrow \mu_{s,t}$ .

The proof of (b) is postponed until the next subsection.

**Definition 4.18** Wiener measure  $\mathbb{W}$  is a probability measure on  $\mathbf{C}$  with  $\mathbb{W}\pi_{t_1, \dots, t_k}^{-1} = \mu_{t_1, \dots, t_k}$  given by the formula in Theorem 4.17 part (a). The standard Wiener process  $W$  is the random element on  $(\mathbf{C}, \mathcal{C}, \mathbb{W})$  defined by  $W_t(x) = x(t)$ .

The existence of  $\mathbb{W}$  follows from Theorems 4.15 and 4.17.

**Theorem 4.19** Let  $X^n = (X_t^n : 0 \leq t \leq 1)$  be defined by Definition 4.16. If  $\xi_i$  are iid with zero mean and finite variance  $\sigma^2$ , then  $X^n$  converges in distribution to the standard Wiener process.

Proof 1. This is a corollary of Theorems 4.15 and 4.17.

Proof 2. An alternative proof is based on Theorem 4.12. We have to verify that condition (i) of Theorem 4.12 holds under the assumptions of Theorem 4.17. To this end take  $t_j = j\delta$ ,  $j = 0, \dots, \delta^{-1}$  assuming  $n\delta > 1$ . Then

$$\begin{aligned} \mathbb{P}(w(X^n, \delta) \geq 3\epsilon) &\leq \sum_{j=1}^{1/\delta} \mathbb{P}\left(\sup_{t_{j-1} \leq s \leq t_j} |X_s^n - X_{t_{j-1}}^n| \geq \epsilon\right) \\ &= \sum_{j=1}^{1/\delta} \mathbb{P}\left(\max_{(j-1)n\delta \leq k \leq jn\delta} \frac{|S_k - S_{(j-1)n\delta}|}{\sigma\sqrt{n}} \geq \epsilon\right) = \sum_{j=1}^{1/\delta} \mathbb{P}\left(\max_{k \leq n\delta} |S_k| \geq \epsilon\sigma\sqrt{n}\right) \\ &= \delta^{-1} \mathbb{P}\left(\max_{k \leq n\delta} |S_k| \geq \epsilon\sigma\sqrt{n}\right) \leq 3\delta^{-1} \max_{k \leq n\delta} \mathbb{P}\left(|S_k| \geq \epsilon\sigma\sqrt{n}/3\right), \end{aligned}$$

where the last is Etemadi's inequality:

$$\mathbb{P}\left(\max_{k \leq n} |S_k| \geq \alpha\right) \leq 3 \max_{k \leq n} \mathbb{P}\left(|S_k| \geq \alpha/3\right).$$

Remark: compare this with Kolmogorov's inequality  $\mathbb{P}(\max_{k \leq n} |S_k| \geq \alpha) \leq \frac{n\sigma^2}{\alpha^2}$ .

It suffices to check that assuming  $\sigma = 1$ ,

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} \lambda^2 \max_{k \leq n} \mathbb{P}\left(|S_k| \geq \epsilon\lambda\sqrt{n}\right) = 0.$$

Indeed, by the classical CLT,

$$\mathbb{P}(|S_k| \geq \epsilon\lambda\sqrt{k}) < 4(1 - \Phi(\epsilon\lambda)) \leq \frac{6}{\epsilon^4\lambda^4}$$

for sufficiently large  $k \geq k(\lambda\epsilon)$ . It follows,

$$\limsup_{n \rightarrow \infty} \lambda^2 \max_{k(\lambda\epsilon) \leq k \leq n} \mathbb{P}\left(|S_k| \geq \epsilon\lambda\sqrt{n}\right) \leq \limsup_{n \rightarrow \infty} \lambda^2 \max_{k \geq k(\lambda\epsilon)} \mathbb{P}\left(|S_k| \geq \epsilon\lambda\sqrt{k}\right) \leq \frac{6}{\epsilon^4\lambda^2}.$$

On the other hand, by Chebyshev's inequality,

$$\limsup_{n \rightarrow \infty} \lambda^2 \max_{k \leq k(\lambda\epsilon)} \mathbb{P}\left(|S_k| \geq \epsilon\lambda\sqrt{n}\right) \leq \limsup_{n \rightarrow \infty} \frac{\lambda^2 k(\lambda\epsilon)}{\epsilon^2 \lambda^2 n} = 0$$

finishing the proof of (i) of Theorem 4.12.

**Example 4.20** We show that  $h(x) = \sup_t x(t)$  is a continuous mapping from  $\mathbf{C}$  to  $\mathbf{R}$ . Indeed, if  $h(x) \geq h(y)$ , then there are  $t_i$  such that

$$0 \leq h(x) - h(y) = x(t_1) - y(t_2) \leq x(t_1) - y(t_1) \leq \|x - y\|.$$

Thus, we have  $|h(x) - h(y)| \leq \rho(x, y)$  and continuity follows.

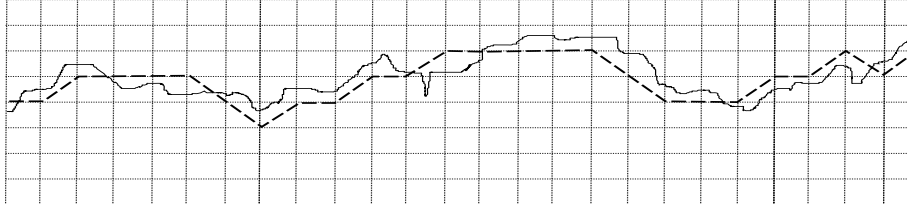


Figure 3: The Arzela-Ascoli theorem: constructing a  $2\epsilon$ -net.

**Example 4.21** Turning to the symmetric simple random walk, put  $M_n = \max(S_0, \dots, S_n)$ . As we show later in Theorem 5.1, for any  $b \geq 0$ ,

$$\mathbb{P}(M_n \leq b\sqrt{n}) \rightarrow \frac{2}{\sqrt{2\pi}} \int_0^b e^{-u^2/2} du.$$

From  $h(X^n) \Rightarrow h(W)$  with  $h(x) = \sup_t x(t)$  we conclude that  $\sup_{0 \leq t \leq 1} W_t$  is distributed as  $|W_1|$ . The same limit holds for  $M_n = \max(\frac{S_0}{\sigma}, \frac{S_1 - \mu}{\sigma}, \dots, \frac{S_n - n\mu}{\sigma})$  for sums of iid r.v. with mean  $\mu$  and standard deviation  $\sigma$ . For this reason the functional CLT is also called an *invariance principle*: the general limit can be computed via the simplest relevant case.

**Exercise 4.22** Check if the following functionals are continuous on  $\mathcal{C}$ :

$$\sup_{\{0 \leq s, t \leq 1\}} |x(t) - x(s)|, \quad \int_0^1 x(t) dt.$$

### 4.3 Tightness in $\mathcal{C}$

**Theorem 4.23** *The Arzela-Ascoli theorem. The set  $A \subset \mathcal{C}$  is relatively compact if and only if*

- (i)  $\sup_{x \in A} |x(0)| < \infty$ ,
- (ii)  $\lim_{\delta \rightarrow 0} \sup_{x \in A} w_x(\delta) = 0$ .

*Proof. Necessity.* If the closure of  $A$  is compact, then (i) obviously must hold. For a fixed  $x$  the function  $w_x(\delta)$  monotonely converges to zero as  $\delta \downarrow 0$ . Since for each  $\delta$  the function  $w_x(\delta)$  is continuous in  $x$  this convergence is uniform over  $x \in K$  for any compact  $K$ . It remains to see that taking  $K$  to be the closure of  $A$  we obtain (ii).

*Sufficiency.* Suppose now that (i) and (ii) hold. For a given  $\epsilon > 0$ , choose  $n$  large enough for  $\sup_{x \in A} w_x(1/n) < \epsilon$ . Since

$$|x(t)| \leq |x(0)| + \sum_{i=1}^n |x(ti/n) - x((i-1)/n)| \leq |x(0)| + n \sup_{x \in A} w_x(1/n),$$

we derive  $\alpha := \sup_{x \in A} \|x\| < \infty$ . The idea is to use this and (ii) to prove that  $A$  is totally bounded, since  $\mathcal{C}$  is complete, it will follow that  $A$  is relatively compact. In other words, we have to find a finite  $B_\epsilon \subset \mathcal{C}$  forming a  $2\epsilon$ -net for  $A$ .

Let  $-\alpha = \alpha_0 < \alpha_1 < \dots < \alpha_k = \alpha$  be such that  $\alpha_j - \alpha_{j-1} \leq \epsilon$ . Then  $B_\epsilon$  can be taken as a set of the continuous polygonal functions  $y : [0, 1] \rightarrow [-\alpha, \alpha]$  that linearly connect the pairs of points  $(\frac{i-1}{n}, \alpha_{j_{i-1}}), (\frac{i}{n}, \alpha_{j_i})$ . See Figure 3. Let  $x \in A$ . It remains to show that there is a  $y \in B_\epsilon$  such that  $\rho(x, y) \leq 2\epsilon$ . Indeed, since  $|x(i/n)| \leq \alpha$ , there is a  $y \in B_\epsilon$  such that  $|x(i/n) - y(i/n)| < \epsilon$  for all  $i = 0, 1, \dots, n$ . Both  $y(i/n)$  and  $y((i-1)/n)$  are within  $2\epsilon$  of  $x(t)$  for  $t \in [(i-1)/n, i/n]$ . Since  $y(t)$  is a convex combination of  $y(i/n)$  and  $y((i-1)/n)$ , it too is within  $2\epsilon$  of  $x(t)$ . Thus  $\rho(x, y) \leq 2\epsilon$  and  $B_\epsilon$  is a  $2\epsilon$ -net for  $A$ .

**Exercise 4.24** Draw a curve  $x \in A$  (cf Figure 3) for which you can not find a  $y \in B_\epsilon$  such that  $\rho(x, y) \leq \epsilon$ .

The next theorem explains the nature of condition (i) in Theorem 4.12.

**Theorem 4.25** *Let  $P_n$  be probability measures on  $(\mathcal{C}, \mathcal{C})$ . The sequence  $(P_n)$  is tight if and only if the following two conditions hold:*

- (i)  $\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} P_n(x : |x(0)| \geq a) = 0$ ,
- (ii)  $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P_n(x : w_x(\delta) \geq \epsilon) = 0$ , for each positive  $\epsilon$ .

Proof. Suppose  $(P_n)$  is tight. Given a positive  $\eta$ , choose a compact  $K$  such that  $P_n(K) > 1 - \eta$  for all  $n$ . By the Arzela-Ascoli theorem we have  $K \subset \{x : |x(0)| \leq a\}$  for large enough  $a$  and  $K \subset \{x : w_x(\delta) \leq \epsilon\}$  for small enough  $\delta$ . Hence the necessity.

According to condition (i), for each positive  $\eta$ , there exist large  $a_\eta$  and  $n_\eta$  such that

$$P_n(x : |x(0)| \geq a_\eta) \leq \eta, \quad n \geq n_\eta,$$

and condition (ii) implies that for each positive  $\epsilon$  and  $\eta$ , there exist a small  $\delta_{\epsilon, \eta}$  and a large  $n_{\epsilon, \eta}$  such that

$$P_n(x : w_x(\delta_{\epsilon, \eta}) \geq \epsilon) \leq \eta, \quad n \geq n_{\epsilon, \eta},$$

Due to Lemma 3.2 for any finite  $k$  the measure  $P_k$  is tight, and so by the necessity there is a  $a_{k, \eta}$  such that  $P_k(x : |x(0)| \geq a_{k, \eta}) \leq \eta$ , and there is a  $\delta_{k, \epsilon, \eta}$  such that  $P_k(x : w_x(\delta_{k, \epsilon, \eta}) \geq \epsilon) \leq \eta$ .

Thus in proving sufficiency, we may put  $n_\eta = n_{\epsilon, \eta} = 1$  in the above two conditions. Fix an arbitrary small positive  $\eta$ . Given the two improved conditions, we have  $P_n(B) \geq 1 - \eta$  and  $P_n(B_k) \geq 1 - 2^{-k}\eta$  with  $B = \{x : |x(0)| < a_\eta\}$  and  $B_k = \{x : w_x(\delta_{1/k, 2^{-k}\eta}) < 1/k\}$ . If  $K$  is the closure of intersection of  $B \cap B_1 \cap B_2 \cap \dots$ , then  $P_n(K) \geq 1 - 2\eta$ . To finish the proof observe that  $K$  is compact by the Arzela-Ascoli theorem.

**Example 4.26** Consider the Dirac probability measure  $P_n$  concentrated on the point  $z_n \in \mathcal{C}$  from Example 4.3. Referring to Theorem 4.25 verify that the sequence  $(P_n)$  is not tight.

**Proof of Theorem 4.17 part b.** The stated tightness follows from Theorem 4.25. Indeed, condition (i) in Theorem 4.25 is trivially fulfilled as  $X_0^n \equiv 0$ . Furthermore, condition (i) of Theorem 4.12 (established in the proof 2 of Theorem 4.19) translates into (ii) in Theorem 4.25.

## 5 Applications of the functional CLT

### 5.1 The minimum and maximum of the Brownian path

**Theorem 5.1** Consider the standard Wiener process  $W = (W_t, 0 \leq t \leq 1)$  and let

$$m = \inf_{0 \leq t \leq 1} W_t, \quad M = \sup_{0 \leq t \leq 1} W_t.$$

If  $a \leq 0 \leq b$  and  $a \leq a' < b' \leq b$ , then

$$\begin{aligned} & \mathbb{P}(a < m \leq M < b; a' < W_1 < b') \\ &= \sum_{k=-\infty}^{\infty} \left( \Phi(2k(b-a) + b') - \Phi(2k(b-a) + a') \right) \\ & \quad - \sum_{k=-\infty}^{\infty} \left( \Phi(2k(b-a) + 2b - a') - \Phi(2k(b-a) + 2b - b') \right), \end{aligned}$$

so that with  $a' = a$  and  $b' = b$  we get

$$\mathbb{P}(a < m \leq M < b) = \sum_{k=-\infty}^{\infty} (-1)^k \left( \Phi(k(b-a) + b) - \Phi(k(b-a) + a) \right).$$

Proof. Let  $S_n$  be the symmetric simple random walk and put  $m_n = \min(S_0, \dots, S_n)$ ,  $M_n = \max(S_0, \dots, S_n)$ . Since the mapping of  $\mathcal{C}$  into  $\mathbf{R}^3$  defined by

$$x \rightarrow \left( \inf_t x(t), \sup_t x(t), x(1) \right)$$

is continuous, the functional CLT entails  $n^{-1/2}(m_n, M_n, S_n) \Rightarrow (m, M, W_1)$ . The theorem's main statement will be obtained in two steps.

Step 1: show that for integers satisfying  $i < 0 < j$  and  $i \leq i' < j' \leq j$ ,

$$\begin{aligned} & \mathbb{P}(i < m_n \leq M_n < j; i' < S_n < j') \\ &= \sum_{k=-\infty}^{\infty} \mathbb{P}(2k(j-i) + i' < S_n < 2k(j-i) + j') \\ & \quad - \sum_{k=-\infty}^{\infty} \mathbb{P}(2k(j-i) + 2j - j' < S_n < 2k(j-i) + 2j - i'). \end{aligned}$$

In other words, we have to show that for  $i < 0 < j$ ,  $i < l < j$

$$\begin{aligned}\mathbb{P}(i < m_n \leq M_n < j; S_n = l) &= \sum_{k=-\infty}^{\infty} \mathbb{P}(S_n = 2k(j-i) + l) \\ &\quad - \sum_{k=-\infty}^{\infty} \mathbb{P}(S_n = 2k(j-i) + 2j - l).\end{aligned}\quad (*)$$

Observe that here both series are just finite sums as  $|S_n| \leq n$ .

Equality (\*) is proved by induction on  $n$ . For  $n = 1$ , if  $j > 1$ , then

$$\begin{aligned}\mathbb{P}(i < m_1 \leq M_1 < j; S_1 = 1) &= \mathbb{P}(S_1 = 1) \\ &= \sum_{k=-\infty}^{\infty} \mathbb{P}(S_1 = 2k(j-i) + 1) - \sum_{k=-\infty}^{\infty} \mathbb{P}(S_1 = 2k(j-i) + 2j - 1),\end{aligned}$$

and if  $i < -1$ , then

$$\begin{aligned}\mathbb{P}(i < m_1 \leq M_1 < j; S_1 = -1) &= \mathbb{P}(S_1 = -1) \\ &= \sum_{k=-\infty}^{\infty} \mathbb{P}(S_1 = 2k(j-i) - 1) - \sum_{k=-\infty}^{\infty} \mathbb{P}(S_1 = 2k(j-i) + 2j + 1).\end{aligned}$$

Assume as induction hypothesis that the statement holds for  $(n-1, i, j, l)$  with all relevant triplets  $(i, j, l)$ . Conditioning on the first step of the random walk, we get

$$\begin{aligned}\mathbb{P}(i < m_n \leq M_n < j; S_n = l) &= \frac{1}{2} \cdot \mathbb{P}(i-1 < m_{n-1} \leq M_{n-1} < j-1; S_{n-1} = l-1) \\ &\quad + \frac{1}{2} \cdot \mathbb{P}(i+1 < m_{n-1} \leq M_{n-1} < j+1; S_{n-1} = l+1),\end{aligned}$$

which together with the induction hypothesis yields the stated equality (\*)

$$\begin{aligned}2\mathbb{P}(i < m_n \leq M_n < j; S_n = l) &= \sum_{k=-\infty}^{\infty} \left( \mathbb{P}(S_{n-1} = 2k(j-i) + l - 1) + \mathbb{P}(S_{n-1} = 2k(j-i) + l + 1) \right) \\ &\quad - \sum_{k=-\infty}^{\infty} \left( \mathbb{P}(S_{n-1} = 2k(j-i) + 2j - l + 1) + \mathbb{P}(S_{n-1} = 2k(j-i) + 2j - l - 1) \right) \\ &= 2 \sum_{k=-\infty}^{\infty} \mathbb{P}(S_n = 2k(j-i) + l) - 2 \sum_{k=-\infty}^{\infty} \mathbb{P}(S_n = 2k(j-i) + 2j - l).\end{aligned}$$

Step 2: show that for  $c > 0$  and  $a < b$ ,

$$\begin{aligned}\sum_{k=-\infty}^{\infty} \mathbb{P}\left(2k\lfloor c\sqrt{n} \rfloor + \lfloor a\sqrt{n} \rfloor < S_n < 2k\lfloor c\sqrt{n} \rfloor + \lfloor b\sqrt{n} \rfloor\right) \\ \rightarrow \sum_{k=-\infty}^{\infty} \left( \Phi(2kc + b) - \Phi(2kc + a) \right), \quad n \rightarrow \infty.\end{aligned}$$

This is obtained using the CLT. The interchange of the limit with the summation over  $k$  follows from

$$\lim_{k_0 \rightarrow \infty} \sum_{|k| > k_0} \mathbb{P}\left(2k\lfloor c\sqrt{n} \rfloor + \lfloor a\sqrt{n} \rfloor < S_n < 2k\lfloor c\sqrt{n} \rfloor + \lfloor b\sqrt{n} \rfloor\right) = 0,$$

which in turn can be justified by the following series form of Scheffe's theorem. If  $\sum_k s_{kn} = \sum_k s_k = 1$ , the terms being nonnegative, and if  $s_{kn} \rightarrow s_k$  for each  $k$ , then  $\sum_k r_k s_{kn} \rightarrow \sum_k r_k s_k$  provided  $r_k$  is bounded. To apply this in our case we should take

$$s_{kn} = \mathbb{P}\left(2k\lfloor \sqrt{n} \rfloor - \lfloor \sqrt{n} \rfloor < S_n \leq 2k\lfloor \sqrt{n} \rfloor + \lfloor \sqrt{n} \rfloor\right), \quad s_k = \Phi(2k+1) - \Phi(2k-1).$$

**Corollary 5.2** Consider the standard Wiener process  $W$ . If  $a \leq 0 \leq b$ , then

$$\begin{aligned}\mathbb{P}\left(\sup_{0 \leq t \leq 1} W_t < b\right) &= 2\Phi(b) - 1, \\ \mathbb{P}\left(\inf_{0 \leq t \leq 1} W_t > a\right) &= 1 - 2\Phi(a), \\ \mathbb{P}\left(\sup_{0 \leq t \leq 1} |W_t| < b\right) &= 2 \sum_{k=-\infty}^{\infty} \left\{ \Phi((4k+1)b) - \Phi((4k-1)b) \right\}.\end{aligned}$$

## 5.2 The arcsine law

**Lemma 5.3** For  $x \in \mathbf{C}$  and a Borel measurable, bounded  $v : \mathbf{R} \rightarrow \mathbf{R}$ , put  $h(x) = \int_0^1 v(x(t))dt$ . If  $v$  is continuous except on a set  $D_v$  with  $\lambda(D_v) = 0$ , where  $\lambda$  is the Lebesgue measure, then  $h$  is  $\mathcal{C}$ -measurable and is continuous except on a set of Wiener measure 0.

Proof. Since both mappings  $x \rightarrow x(t)$  and  $t \rightarrow x(t)$  are continuous, the mapping  $(x, t) \rightarrow x(t)$  is continuous in the product topology and therefore Borel measurable. It follows that the mapping  $\psi(x, t) = v(x(t))$  is also measurable. Since  $\psi$  is bounded,  $h(x) = \int_0^1 \psi(x, t)dt$  is  $\mathcal{C}$ -measurable, see Fubini's theorem.

Let  $E = \{(x, t) : x(t) \in D_v\}$ . If  $\mathbb{W}$  is Wiener measure on  $(\mathbf{C}, \mathcal{C})$ , then by the hypothesis  $\lambda(D_v) = 0$ ,

$$\mathbb{W}\{x : (x, t) \in E\} = \mathbb{W}\{x : x(t) \in D_v\} = 0 \text{ for each } t \in [0, 1].$$

It follows by Fubini's theorem applied to the measure  $\mathbb{W} \times \lambda$  on  $\mathbf{C} \times [0, 1]$  that  $\lambda\{t : (x, t) \in E\} = 0$  for all  $x$  outside a set  $A_v \in \mathcal{C}$  satisfying  $\mathbb{W}(A_v) = 0$ . Suppose that  $\|x_n - x\| \rightarrow 0$ . If  $x \notin A_v$ , then  $x(t) \notin D_v$  for almost all  $t$  and hence  $v(x_n(t)) \rightarrow v(x(t))$  for almost all  $t$ . It follows by the bounded convergence theorem that

$$\text{if } x \notin A_v \text{ and } \|x_n - x\| \rightarrow 0, \quad \text{then } \int_0^1 v(x_n(t))dt \rightarrow \int_0^1 v(x(t))dt.$$

**Exercise 5.4** Let  $W$  be a standard Wiener process and  $t_0 \in (0, 1)$ . Put  $W'_s = \frac{W_t - W_{t_0}}{\sqrt{1-t_0}}$  for  $s = \frac{t-t_0}{1-t_0}$ ,  $t \in [t_0, 1]$ . Using the Donsker invariance principle show that  $(W'_s, 0 \leq s \leq 1)$  is also distributed as a standard Wiener process.

**Lemma 5.5** Each of the following three mappings  $h_i : \mathbf{C} \rightarrow \mathbf{R}$

$$\begin{aligned}h_1(x) &= \sup\{t : x(t) = 0, t \in [0, 1]\}, \\ h_2(x) &= \lambda\{t : x(t) > 0, t \in [0, 1]\}, \\ h_3(x) &= \lambda\{t : x(t) > 0, t \in [0, h_1(x)]\}\end{aligned}$$

is  $\mathcal{C}$ -measurable and continuous except on a set of Wiener measure 0.

Proof. Using the previous lemma with  $v(z) = 1_{\{z \in (0, \infty)\}}$  we obtain the assertion for  $h_2$ .

Turning to  $h_1$ , observe that

$$\{x : h_1(x) < \alpha\} = \{x : x(t) > 0, t \in [\alpha, 1]\} \cup \{x : x(t) < 0, t \in [\alpha, 1]\}$$

is open and hence  $h_1$  is measurable. If  $h_1$  is discontinuous at  $x$ , then there exist  $0 < t_0 < t_1 < 1$  such that  $x(t_1) = 0$  and

$$\text{either } x(t) > 0 \text{ for all } t \in [t_0, 1] \setminus \{t_1\} \quad \text{or} \quad x(t) < 0 \text{ for all } t \in [t_0, 1] \setminus \{t_1\}.$$

That  $h_1$  is continuous except on a set of Wiener measure 0 will therefore follow if we show that, for each  $t_0$ , the random variables

$$M_0 = \sup\{W_t, t \in [t_0, 1]\} \quad \text{and} \quad \inf\{W_t, t \in [t_0, 1]\}$$

have continuous distributions. By the last exercise and Theorem 5.1,  $M' = M_0 - W_{t_0}$  has a continuous distribution. Because  $M'$  and  $W_{t_0}$  are independent, we conclude that their sum also has a continuous distribution. The infimum is treated the same way.

Finally, for  $h_3$ , use the representation

$$h_3(x) = \psi(x, h_1(x)), \quad \text{where } \psi(x, t) = \int_0^t v(x(u))du \quad \text{with } v(z) = 1_{\{z \in (0, \infty)\}}.$$

**Theorem 5.6** Consider the standard Wiener process  $W$  and let

$T = h_1(W)$  be the time at which  $W$  last passes through 0,

$U = h_2(W)$  be the total amount of time  $W$  spends above 0, and

$V = h_3(W)$  be the total amount of time  $W$  spends above 0 in the interval  $[0, T]$ .

so that

$$U = V + (1 - T)1_{\{W_1 \geq 0\}}.$$

Then the triplet  $(T, V, W_1)$  has the joint density

$$f(t, v, z) = 1_{\{0 < v < t < 1\}} g(t, z), \quad g(t, z) = \frac{1}{2\pi} \frac{|z|}{t^{3/2}(1-t)^{3/2}} e^{-\frac{z^2}{2(1-t)}}.$$

In particular, the conditional distribution of  $V$  given  $(T, W_1)$  is uniform on  $[0, T]$ , and

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

Proof. The main idea is to apply the invariance principle via the symmetric simple random walk  $S_n$ . We will use three properties of  $S_n$  and its path functionals  $(T_n, U_n, V_n)$ . First, we need the local limit theorem for  $p_n(i) = \mathbb{P}(S_n = i)$  similar to that of Example 1.29:

$$\text{if } \frac{i}{\sqrt{n}} \rightarrow z, \text{ with } n - i \text{ being even, then } \frac{\sqrt{n}}{2} p_n(i) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Second, we need the fact that

$$\mathbb{P}(S_1 \geq 1, \dots, S_{n-1} \geq 1, S_n = i) = \frac{i}{n} p_n(i), \quad i \geq 1.$$

The third fact we need is that if  $S_{2n} = 0$ , then  $U_{2n} = V_{2n}$  and

$$\mathbb{P}(V_{2n} = 2j | S_{2n} = 0) = \frac{1}{n+1}, \quad j = 0, 1, \dots, n.$$

Using these three facts we obtain that for  $0 \leq 2j \leq 2k < n$  and  $i \geq 1$ ,

$$\begin{aligned} \mathbb{P}(T_n = 2k, V_{2n} = 2j, S_n = i) &= \mathbb{P}(S_{2k} = 0, V_{2n} = 2j, S_{2k+1} \geq 1, \dots, S_{n-1} \geq 1, S_n = i) \\ &= \mathbb{P}(S_{2k} = 0) \mathbb{P}(V_{2n} = 2j | S_{2k} = 0) \mathbb{P}(S_{2k+1} \geq 1, \dots, S_{n-1} \geq 1, S_n = i | S_{2k} = 0) \\ &= p_{2k}(0) \frac{1}{k+1} \frac{i}{n-2k} p_{n-2k}(i). \end{aligned}$$

We apply Theorem 1.28 to the three-dimensional lattice of points  $(\frac{2k}{n}, \frac{2j}{n}, \frac{i}{\sqrt{n}})$  for which  $i \equiv n \pmod{2}$ . The volume of the corresponding cell is  $\frac{2}{n} \cdot \frac{2}{n} \cdot \frac{2}{\sqrt{n}} = 8n^{-5/2}$ . If

$$\frac{2k}{n} \rightarrow t, \quad \frac{2j}{n} \rightarrow v, \quad \frac{i}{\sqrt{n}} \rightarrow z, \quad 0 < v < t < 1, \quad z > 0,$$

then

$$\begin{aligned} \frac{n^{5/2}}{8} \mathbb{P}(T_n = 2k, V_{2n} = 2j, S_n = i) &= \frac{\sqrt{n}}{\sqrt{2k}} \frac{\sqrt{2k}}{2} p_{2k}(0) \frac{n}{2(k+1)} \frac{i}{\sqrt{n}} \frac{n}{n-2k} \frac{\sqrt{n}}{\sqrt{n-2k}} \frac{\sqrt{n-2k}}{2} p_{n-2k}(i) \\ &\rightarrow \frac{1}{\sqrt{2\pi}} \frac{1}{t^{3/2}} z \frac{1}{(1-t)^{3/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2(1-t)}} = g(t, z). \end{aligned}$$

The same result holds for negative  $z$  by symmetry.

The joint density of  $(T, W_1)$  is  $tg(t, z)1_{\{0 < t < 1\}}$ , hence the marginal density for  $T$  equals

$$f_T(t) = \int_{-\infty}^{\infty} tg(t, z) dz = \int_0^{\infty} e^{-\frac{z^2}{2(1-t)}} \frac{z dz}{\pi(1-t)^{3/2} t^{1/2}} = \frac{1}{\pi(1-t)^{1/2} t^{1/2}}$$

implying

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

Notice also that

$$\begin{aligned} G(u) &:= \int_{-\infty}^{\infty} \int_u^1 g(t, z) dt dz = \int_u^1 \int_0^{\infty} e^{-\frac{z^2}{2(1-t)}} \frac{z dz}{\pi(1-t)^{3/2} t^{3/2}} dt \\ &= \int_u^1 \frac{dt}{\pi(1-t)^{1/2} t^{3/2}} = -\frac{2}{\pi} \int_u^1 \frac{dt^{-1/2}}{\sqrt{1-t}} = \frac{2}{\pi} \int_1^{u^{-1/2}} \frac{y dy}{\sqrt{y^2 - 1}} = \frac{2}{\pi} \sqrt{u^{-1} - 1}. \end{aligned}$$

If  $(T, W_1) = (t, z)$ , then  $U$  is distributed uniformly over  $[1-t, 1]$  for  $z \geq 0$ , and uniformly over  $[0, t]$  for  $z < 0$ :

$$\mathbb{P}(U \leq u | T = t, W_1 = z) = \frac{u-1+t}{t} 1_{\{u \in [1-t, 1], z \geq 0\}} + \frac{u}{t} 1_{\{u \in [0, t], z < 0\}} + 1_{\{u \in (t, 1], z < 0\}}.$$

Thus the marginal distribution function of  $U$  equals

$$\begin{aligned} \mathbb{P}(U \leq u) &= \mathbb{E} \left( \frac{u-1+T}{T} 1_{\{u \in [1-T, 1], W_1 \geq 0\}} + \frac{u}{T} 1_{\{u \in [0, T], W_1 < 0\}} + 1_{\{u \in (T, 1], W_1 < 0\}} \right) \\ &= \int_0^{\infty} \int_{1-u}^1 (u-1+t) g(t, z) dt dz + \int_{-\infty}^0 \int_u^1 u g(t, z) dt dz + \int_{-\infty}^0 \int_0^u t g(t, z) dt dz \\ &= \frac{1}{2} \int_{1-u}^1 f_T(t) dt + \frac{u-1}{2} G(1-u) + \frac{u}{2} G(u) + \frac{1}{2} \int_0^u f_T(t) dt \\ &= \frac{1}{2} \mathbb{P}(T > 1-u) + \frac{1}{2} \mathbb{P}(T \leq u) = \frac{2}{\pi} \arcsin(\sqrt{u}). \end{aligned}$$

### 5.3 The Brownian bridge

**Definition 5.7** The transformed standard Wiener process  $W_t^\circ = W_t - tW_1$ ,  $t \in [0, 1]$ , is called the standard Brownian bridge.

**Exercise 5.8** Show that the standard Brownian bridge  $W^\circ$  is a Gaussian process with zero mean and covariance  $\mathbb{E}(W_s^\circ W_t^\circ) = s(1-t)$  for  $s \leq t$ .

**Example 5.9** Define  $h : \mathcal{C} \rightarrow \mathcal{C}$  by  $h(x(t)) = x(t) - tx(1)$ . This is a continuous mapping since  $\rho(h(x), h(y)) \leq 2\rho(x, y)$ , and  $h(X^n) \Rightarrow W^\circ$  by Theorem 4.19.

**Theorem 5.10** Let  $P_\epsilon$  be the probability measure on  $(\mathcal{C}, \mathcal{C})$  defined by

$$P_\epsilon(A) = \mathbb{P}(W \in A | 0 \leq W_1 \leq \epsilon), \quad A \in \mathcal{C}.$$

Then  $P_\epsilon \Rightarrow \mathbb{W}^\circ$  as  $\epsilon \rightarrow 0$ , where  $\mathbb{W}^\circ$  is the distribution of the Brownian bridge  $W^\circ$ .

Proof. We will prove that for every closed  $F \in \mathcal{C}$

$$\limsup_{\epsilon \rightarrow 0} \mathbb{P}(W \in F | 0 \leq W_1 \leq \epsilon) \leq \mathbb{P}(W^\circ \in F).$$

Using  $W_t^\circ = W_t - tW_1$  we get  $\mathbb{E}(W_t^\circ W_1) = 0$  for all  $t$ . From the normality we conclude that  $W_1$  is independent of each  $(W_{t_1}^\circ, \dots, W_{t_k}^\circ)$ . Therefore,

$$\mathbb{P}(W^\circ \in A, W_1 \in B) = \mathbb{P}(W^\circ \in A) \mathbb{P}(W_1 \in B), \quad A \in \mathcal{C}_f, B \in \mathcal{R},$$

and since  $\mathcal{C}_f$ , the collection of finite-dimensional sets, see the proof of Lemma 4.13, is a separating class, it follows

$$\mathbb{P}(W^\circ \in A | 0 \leq W_1 \leq \epsilon) = \mathbb{P}(W^\circ \in A), \quad A \in \mathcal{C}, \epsilon > 0.$$

Observe that  $\rho(W, W^\circ) = |W_1|$ . Thus,

$$\{|W_1| \leq \delta\} \cap \{W \in F\} \subset \{W^\circ \in F_\delta\}, \quad \text{where } F_\delta = \{x : \rho(x, F) \leq \delta\}.$$

Therefore, if  $\epsilon < \delta$

$$\mathbb{P}(W \in F | 0 \leq W_1 \leq \epsilon) \leq \mathbb{P}(W^\circ \in F_\delta | 0 \leq W_1 \leq \epsilon) = \mathbb{P}(W^\circ \in F_\delta),$$

leading to the required result

$$\limsup_{\epsilon \rightarrow 0} \mathbb{P}(W \in F | 0 \leq W_1 \leq \epsilon) \leq \limsup_{\delta \rightarrow 0} \mathbb{P}(W^\circ \in F_\delta) = \mathbb{P}(W^\circ \in F).$$

**Theorem 5.11** *Distribution functions for several functionals of the Brownian bridge:*

$$\begin{aligned}\mathbb{P}(a < \inf_t W_t^\circ \leq \sup_t W_t^\circ \leq b) &= \sum_{k=-\infty}^{\infty} \left( e^{-2k^2(b-a)^2} - e^{-2(b+k(b-a))^2} \right), \quad a < 0 < b, \\ \mathbb{P}(\sup_t |W_t^\circ| \leq b) &= 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 b^2}, \quad b > 0, \\ \mathbb{P}(\sup_t W_t^\circ \leq b) &= \mathbb{P}(\inf_t W_t^\circ > -b) = 1 - e^{-2b^2}, \quad b > 0, \\ \mathbb{P}(h_2(W^\circ) \leq u) &= u, \quad u \in [0, 1].\end{aligned}$$

Proof. The main idea of the proof is the following. Suppose that  $h : \mathbf{C} \rightarrow \mathbf{R}^k$  is a measurable mapping and that the set  $D_h$  of its discontinuities satisfies  $\mathbb{W}^\circ(D_h) = 0$ . It follows by Theorem 5.10 and the mapping theorem that

$$\mathbb{P}(h(W^\circ) \leq \alpha) = \lim_{\epsilon \rightarrow 0} \mathbb{P}(h(W) \leq \alpha | 0 \leq W_1 \leq \epsilon).$$

Using either this or alternatively,

$$\mathbb{P}(h(W^\circ) \leq \alpha) = \lim_{\epsilon \rightarrow 0} \mathbb{P}(h(W) \leq \alpha | -\epsilon \leq W_1 \leq 0)$$

one can find explicit forms for distributions connected with  $W^\circ$ .

Turning to Theorem 5.1 with  $a < 0 < b$  and  $a' = 0, b' = \epsilon$  we get

$$\begin{aligned}\mathbb{P}(a < m \leq M < b; 0 < W_1 < \epsilon) \\ &= \sum_{k=-\infty}^{\infty} \left( \Phi(2k(b-a) + \epsilon) - \Phi(2k(b-a)) \right) \\ &\quad - \sum_{k=-\infty}^{\infty} \left( \Phi(2k(b-a) + 2b) - \Phi(2k(b-a) + 2b - \epsilon) \right).\end{aligned}$$

This implies the first statement as

$$\frac{\Phi(z + \epsilon) - \Phi(z)}{\epsilon} \rightarrow \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

As for the last statement, we need to show, in terms of  $U = h_2(W)$ , that

$$\lim_{\epsilon \rightarrow 0} \mathbb{P}(U \leq u | -\epsilon \leq W_1 \leq 0) = u,$$

or, in terms of  $V = h_3(W)$ , that

$$\lim_{\epsilon \rightarrow 0} \mathbb{P}(V \leq u | -\epsilon \leq W_1 \leq 0) = u.$$

Recall that the distribution of  $V$  for given  $T$  and  $W_1$  is uniform on  $(0, T)$ , in other words,  $L = V/T$  is uniformly distributed on  $(0, 1)$  and is independent of  $(T, W_1)$ . Thus,

$$\begin{aligned}\mathbb{P}(V \leq u | -\epsilon \leq W_1 \leq 0) &= \mathbb{P}(TL \leq u | -\epsilon \leq W_1 \leq 0) \\ &= \int_0^1 \mathbb{P}(T \leq u/s | -\epsilon \leq W_1 \leq 0) ds \\ &= u + \int_u^1 \mathbb{P}(T \leq u/s | -\epsilon \leq W_1 \leq 0) ds.\end{aligned}$$

It remains to see that

$$\begin{aligned}\int_u^1 \mathbb{P}(T \leq u/s | -\epsilon \leq W_1 \leq 0) ds &= \frac{1}{\Phi(\epsilon) - \Phi(0)} \int_u^1 \mathbb{P}(T \leq u/s; -\epsilon \leq W_1 \leq 0) ds \\ &\leq c\epsilon^{-1} \int_u^1 \mathbb{P}(T \leq r; -\epsilon \leq W_1 \leq 0) \frac{dr}{r^2} \leq c\epsilon^{-1} u^{-2} \int_u^1 \int_0^r \int_0^\epsilon tg(t, z) dz dt dr \\ &\leq c_1 \epsilon u^{-2} \int_0^1 \frac{dt}{t^{1/2}(1-t)^{1/2}} \rightarrow 0, \quad \epsilon \rightarrow 0.\end{aligned}$$



## 6 The space $D = D[0, 1]$

### 6.1 Cadlag functions

**Definition 6.1** Let  $D = D[0, 1]$  be the space of functions  $x : [0, 1] \rightarrow \mathbf{R}$  that are right continuous and have left-hand limits.

**Exercise 6.2** If  $x_n \in D$  and  $\|x_n - x\| \rightarrow 0$ , then  $x \in D$ .

For  $x \in D$  and  $T \subset [0, 1]$  we will use notation

$$w_x(T) = w(x, T) = \sup_{s, t \in T} |x(t) - x(s)|,$$

and write  $w_x[t, t + \delta]$  instead of  $w_x([t, t + \delta])$ . This should not be confused with the earlier defined modulus of continuity

$$w_x(\delta) = w(x, \delta) = \sup_{0 \leq t \leq 1 - \delta} w_x[t, t + \delta],$$

Clearly, if  $T_1 \subset T_2$ , then  $w_x(T_1) \leq w_x(T_2)$ . Hence  $w_x(\delta)$  is monotone over  $\delta$ .

**Example 6.3** Consider  $x_n(t)$  the fractional part of  $nt$ . It has regular downward jumps of size 1. For example,  $x_1(t) = t$  for  $t \in [0, 1)$ , and  $x_1(1) = 0$ . Another example:  $x_2(t) = 2t$  for  $t \in [0, 1/2)$ ,  $x_2(t) = 2t - 1$  for  $t \in [1/2, 1)$ , and  $x_2(1) = 0$ . Placing an interval  $[t, t + \delta]$  around a jump, we find  $w_{x_n}(\delta) \equiv 1$ .

**Lemma 6.4** Consider an arbitrary  $x \in D$ . For each  $\epsilon > 0$ , there exist points  $0 = t_0 < t_1 < \dots < t_v = 1$  such that

$$w_x[t_{i-1}, t_i] < \epsilon, \quad i = 1, 2, \dots, v.$$

It follows that  $x$  is bounded, and that  $x$  can be uniformly approximated by simple functions constant over intervals, so that it is Borel measurable. It follows also that  $x$  has at most countably many jumps.

*Proof.* To prove the first statement, let  $t^\circ = t^\circ(\epsilon)$  be the supremum of those  $t \in [0, 1]$  for which  $[0, t]$  can be decomposed into finitely many subintervals satisfying  $w_x[t_{i-1}, t_i] < \epsilon$ . We show in three steps that  $t^\circ = 1$ .

Step 1. Since  $x(0) = x(0+)$ , we have  $w_x[0, \eta_0] < \epsilon$  for some small positive  $\eta_0$ . Thus  $t^\circ > 0$ .

Step 2. Since  $x(t^\circ -)$  exists, we have  $w_x[t^\circ - \eta_1, t^\circ] < \epsilon$  for some small positive  $\eta_1$ , which implies that the interval  $[0, t^\circ]$  can itself be so decomposed.

Step 3. Suppose  $t^\circ = \tau$ , where  $\tau < 1$ . From  $x(\tau) = x(\tau+)$  using the argument of the step 1, we see that according to the definition of  $t^\circ$  we should have  $t^\circ > \tau$ .

The last statement of the lemma follows from the fact that for any natural  $n$ , there exist at most finitely many points  $t$  at which  $|x(t) - x(t-)| \geq n^{-1}$ .

**Exercise 6.5** Find a bounded function  $x \notin D$  with the following property: for any set  $0 = t_0 < t_1 < \dots < t_v = 1$  there exists an  $i$  such that  $w_x[t_{i-1}, t_i] \geq 1$ .

**Definition 6.6** Let  $\delta \in (0, 1)$ . A set  $0 = t_0 < t_1 < \dots < t_v = 1$  is called  $\delta$ -sparse if  $t_i - t_{i-1} > \delta$  for  $i = 1, \dots, v$ . Define an analog of the modulus of continuity  $w_x(\delta)$  by

$$w'_x(\delta) = w'(x, \delta) = \inf_{\{t_i\}} \max_{1 \leq i \leq v} w_x[t_{i-1}, t_i],$$

where the infimum extends over all  $\delta$ -sparse sets  $\{t_i\}$ . The function  $w'_x(\delta)$  is called a *cadlag modulus* of  $x$ .

**Exercise 6.7** Using Lemma 6.4 show that a function  $x : [0, 1] \rightarrow \mathbf{R}$  belongs to  $D$  if and only if  $w'_x(\delta) \rightarrow 0$ .

**Exercise 6.8** Compute  $w'_x(\delta)$  for  $x = 1_{[0, a]}$ .

**Lemma 6.9** For any  $x$ ,  $w'_x(\delta)$  is non-decreasing over  $\delta$ , and  $w'_x(\delta) \leq w_x(2\delta)$ . Moreover, for any  $x \in D$ ,

$$j_x \leq w_x(\delta) \leq 2w'_x(\delta) + j_x, \quad j_x = \sup_{0 < t \leq 1} |x(t) - x(t-)|.$$

*Proof.* Taking a  $\delta$ -sparse set with  $t_i - t_{i-1} \leq 2\delta$  we get  $w'_x(\delta) \leq w_x(2\delta)$ . To see that  $w_x(\delta) \leq 2w'_x(\delta) + j_x$  take a  $\delta$ -sparse set such that  $w_x[t_{i-1}, t_i] \leq w'_x(\delta) + \epsilon$  for all  $i$ . If  $|t - s| \leq \delta$ , then  $s, t \in [t_{i-1}, t_{i+1})$  for some  $i$  and  $|x(t) - x(s)| \leq 2(w'_x(\delta) + \epsilon) + j_x$ .

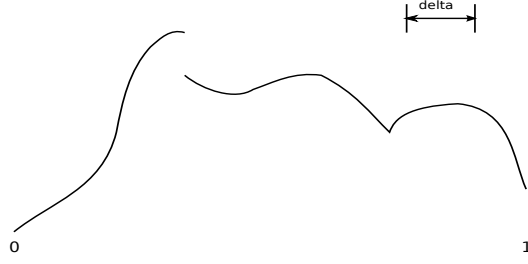


Figure 4: Exercise 6.13.

**Lemma 6.10** Considering triples  $t_1, t, t_2$  in  $[0, 1]$  put

$$w''_x(\delta) = w''(x, \delta) = \sup_{t_1 \leq t_2 \leq t_1 + \delta} \sup_{t_1 \leq t \leq t_2} \{|x(t) - x(t_1)| \wedge |x(t_2) - x(t)|\}.$$

For any  $x$ ,  $w''_x(\delta)$  is non-decreasing over  $\delta$ , and  $w''_x(\delta) \leq w'_x(\delta)$ .

Proof. Suppose that  $w'_x(\delta) < w$  and  $\{\tau_i\}$  be a  $\delta$ -sparse set such that  $w_x[\tau_{i-1}, \tau_i] < w$  for all  $i$ . If  $t_1 \leq t \leq t_2 \leq t_1 + \delta$ , then either  $|x(t) - x(t_1)| < w$  or  $|x(t_2) - x(t)| < w$ . Thus  $w''_x(\delta) < w$  and letting  $w \downarrow w'_x(\delta)$  we obtain  $w''_x(\delta) \leq w'_x(\delta)$ .

**Example 6.11** For the functions  $x_n(t) = 1_{\{t \in [0, n^{-1}]\}}$  and  $y_n = 1_{\{t \in [1-n^{-1}, 1]\}}$  we have  $w''_{x_n}(\delta) = w''_{y_n}(\delta) = 0$ , although  $w'_{x_n}(\delta) = w'_{y_n}(\delta) = 1$  for  $n \geq \delta^{-1}$ .

**Exercise 6.12** Consider  $x_n(t)$  from Example 6.3. Find  $w'_{x_n}(\delta)$  and  $w''_{x_n}(\delta)$  for all  $(n, \delta)$ .

**Exercise 6.13** Compare the values of  $w'_x(\delta)$  and  $w''_x(\delta)$  for the curve  $x$  on the Figure 4.

**Lemma 6.14** For any  $x \in D$  and  $\delta \in (0, 1)$ ,

$$\frac{w'_x(\delta/2)}{24} \leq w''_x(\delta) \vee |x(\delta) - x(0)| \vee |x(1-) - x(1 - \delta)| \leq w'_x(\delta).$$

Proof. The second inequality follows from the definition of  $w'_x(\delta)$  and Lemma 6.10. For the first inequality it suffices to show that

- (i)  $w_x[t_1, t_2] \leq 2(w''_x(\delta) + |x(t_2) - x(t_1)|)$ , if  $t_2 \leq t_1 + \delta$ ,
- (ii)  $w'_x(\delta/2) \leq 6(w''_x(\delta) \vee w_x[0, \delta] \vee w_x[1 - \delta, 1])$ ,

as these two relations imply

$$\begin{aligned} \frac{w'_x(\delta/2)}{6} &\leq w''_x(\delta) \vee w_x[0, \delta] \vee w_x[1 - \delta, 1] \\ &\leq (2w''_x(\delta) + 2|x(\delta) - x(0)|) \vee (2w''_x(\delta) + 2|x(1-) - x(1 - \delta)|) \\ &\leq (4w''_x(\delta)) \vee (4|x(\delta) - x(0)|) \vee (4|x(1-) - x(1 - \delta)|). \end{aligned}$$

Here we used the trick

$$w_x[1 - \delta, 1] = \lim_{t \uparrow 1} w_x[1 - \delta, t] \leq 2w''_x(\delta) + 2 \lim_{t \uparrow 1} |x(t) - x(1 - \delta)|.$$

To see (i), note that, if  $t_1 \leq t < t_2 \leq t_1 + \delta$ , then either  $|x(t) - x(t_1)| \leq w''_x(\delta)$ , or  $|x(t_2) - x(t)| \leq w''_x(\delta)$ . In the latter case, we have

$$|x(t) - x(t_1)| \leq |x(t) - x(t_2)| + |x(t_2) - x(t_1)| \leq w''_x(\delta) + |x(t_2) - x(t_1)|.$$

Therefore, for  $t_2 \leq t_1 + \delta$ ,

$$\sup_{t_1 \leq t < t_2} |x(t) - x(t_1)| \leq w''_x(\delta) + |x(t_2) - x(t_1)|,$$

hence

$$\sup_{t_1 \leq s, t < t_2} |x(t) - x(s)| \leq 2(w''_x(\delta) + |x(t_2) - x(t_1)|).$$

We prove (ii) in four steps.

Step 1. We will need the following inequality

$$|x(s) - x(t_1)| \wedge |x(t_2) - x(t)| \leq 2w_x''(\delta) \text{ if } t_1 \leq s < t \leq t_2 \leq t_1 + \delta. \quad (*)$$

To see this observe that, by the definition of  $w_x''(\delta)$ , either  $|x(s) - x(t_1)| \leq w_x''(\delta)$  or both  $|x(t_2) - x(s)| \leq w_x''(\delta)$  and  $|x(t) - x(s)| \leq w_x''(\delta)$ . In the second case, using the triangular inequality we get  $|x(t_2) - x(t)| \leq 2w_x''(\delta)$ .

Step 2. Putting

$$\alpha := w_x''(\delta) \vee w_x[0, \delta] \vee w_x[1 - \delta, 1], \quad T_{x, \alpha} := \{t : x(t) - x(t-) > 2\alpha\},$$

show that there exist points  $0 = s_0 < s_1 < \dots < s_r = 1$  such that  $s_i - s_{i-1} \geq \delta$  and

$$T_{x, \alpha} \subset \{s_0, \dots, s_r\}.$$

Suppose  $u_1, u_2 \in T_{x, \alpha}$  and  $0 < u_1 < u_2 < u_1 + \delta$ . Then there are disjoint intervals  $(t_1, s)$  and  $(t, t_2)$  such that  $u_1 \in (t_1, s)$ ,  $u_2 \in (t, t_2)$ , and  $t_2 - t_1 < \delta$ . As both these intervals are short enough, we have a contradiction with  $(*)$ . Thus  $(0, 1)$  can not contain two points from  $T_{x, \alpha}$ , within  $\delta$  of one another. And neither  $[0, \delta)$  nor  $[1 - \delta, 1)$  can contain a point from  $T_{x, \alpha}$ .

Step 3. Recursively adding middle points for the pairs  $(s_{i-1}, s_i)$  such that  $s_i - s_{i-1} > \delta$  we get an enlarged set  $\{s_0, \dots, s_r\}$  (with possibly a larger  $r$ ) satisfying

$$T_{x, \alpha} \subset \{s_0, \dots, s_r\}, \quad \delta/2 < s_i - s_{i-1} \leq \delta, \quad i = 1, \dots, r.$$

Step 4. It remains to show that  $w_x'(\delta/2) \leq 6\alpha$ . Since  $\{s_0, \dots, s_r\}$  from step 3 is a  $(\delta/2)$ -sparse set, it suffices to verify that

$$w_x[s_{i-1}, s_i] \leq 6\alpha, \quad i = 1, \dots, r.$$

The proof will be completed after we demonstrate that

$$|x(t_2) - x(t_1)| \leq 6\alpha \quad \text{for } s_{i-1} \leq t_1 < t_2 < s_i.$$

Define  $\sigma_1$  and  $\sigma_2$  by

$$\begin{aligned} \sigma_1 &= \sup\{\sigma \in [t_1, t_2] : \sup_{t_1 \leq u \leq \sigma} |x(u) - x(t_1)| \leq 2\alpha\}, \\ \sigma_2 &= \inf\{\sigma \in [t_1, t_2] : \sup_{\sigma \leq u \leq t_2} |x(t_2) - x(u)| \leq 2\alpha\}. \end{aligned}$$

If  $\sigma_1 < \sigma_2$ , then there are  $\sigma_1 < s < t < \sigma_2$  violating  $(*)$  due to the fact that by definition of  $\alpha$ , we have  $w_x''(\delta) \leq \alpha$ . Therefore,  $\sigma_2 \leq \sigma_1$  and it follows that  $|x(\sigma_1-) - x(t_1)| \leq 2\alpha$  and  $|x(t_2) - x(\sigma_1)| \leq 2\alpha$ . Since  $\sigma_1 \in (s_{i-1}, s_i)$ , we have  $|x(\sigma_1) - x(\sigma_1-)| \leq 2\alpha$  implying

$$|x(t_2) - x(t_1)| \leq |x(t_2) - x(\sigma_1)| + |x(\sigma_1) - x(\sigma_1-)| + |x(\sigma_1-) - x(t_1)| \leq 6\alpha.$$

## 6.2 Two metrics in $D$ and the Skorokhod topology

**Example 6.15** Consider  $x(t) = 1_{\{t \in [a, 1]\}}$  and  $y(t) = 1_{\{t \in [b, 1]\}}$  for  $a, b \in (0, 1)$ . If  $a \neq b$ , then  $\|x - y\| = 1$  even when  $a$  is very close to  $b$ . For the space  $D$ , the uniform metric is not good and we need another metric.

**Definition 6.16** Let  $\Lambda$  denote the class of strictly increasing continuous mappings  $\lambda : [0, 1] \rightarrow [0, 1]$  with  $\lambda 0 = 0$ ,  $\lambda 1 = 1$ . Denote by  $1 \in \Lambda$  the identity map  $1t \equiv t$ , and put  $\|\lambda\|^\circ = \sup_{s < t} \left| \log \frac{\lambda t - \lambda s}{t - s} \right|$ . The smaller  $\|\lambda\|^\circ$  is the closer to 1 are the slopes of  $\lambda$ :

$$e^{-\|\lambda\|^\circ} \leq \frac{\lambda t - \lambda s}{t - s} \leq e^{\|\lambda\|^\circ}.$$

**Exercise 6.17** Let  $\lambda, \mu \in \Lambda$ . Show that

$$\|\lambda\mu - \lambda\| \leq \|\mu - 1\| \cdot e^{\|\lambda\|^\circ}.$$

**Definition 6.18** For  $x, y \in D$  define

$$\begin{aligned} d(x, y) &= \inf_{\lambda \in \Lambda} \{\|\lambda - 1\| \vee \|x - y\lambda\|\}, \\ d^\circ(x, y) &= \inf_{\lambda \in \Lambda} \{\|\lambda\|^\circ \vee \|x - y\lambda\|\}, \end{aligned}$$

**Exercise 6.19** Show that  $d(x, y) \leq \|x - y\|$  and  $d^\circ(x, y) \leq \|x - y\|$ .

**Example 6.20** Consider  $x(t) = 1_{\{t \in [a, 1]\}}$  and  $y(t) = 1_{\{t \in [b, 1]\}}$  for  $a, b \in (0, 1)$ . Clearly, if  $\lambda(a) = b$ , then  $\|x - y\lambda\| = 0$  and otherwise  $\|x - y\lambda\| = 1$ . Thus

$$d(x, y) = \inf\{\|\lambda - 1\| : \lambda \in \Lambda, \lambda(a) = b\} = |a - b|,$$

$$d^\circ(x, y) = \left(\inf\{\|\lambda\|^\circ : \lambda \in \Lambda, \lambda(a) = b\}\right) \wedge 1 = \left(\left|\log \frac{a}{b}\right| \vee \left|\log \frac{1-a}{1-b}\right|\right) \wedge 1,$$

so that  $d(x, y) \rightarrow 0$  and  $d^\circ(x, y) \rightarrow 0$  as  $b \rightarrow a$ .

**Exercise 6.21** Given  $0 < b < a < c < 1$ , find  $d(x, y)$  for

$$x(t) = 2 \cdot 1_{\{t \in [a, 1]\}}, \quad y(t) = 1_{\{t \in [b, 1]\}} + 1_{\{t \in [c, 1]\}}.$$

Does  $d(x, y) \rightarrow 0$  as  $b \rightarrow a$  and  $c \rightarrow a$ ?

**Lemma 6.22** Both  $d$  and  $d^\circ$  are metrics in  $\mathbf{D}$ , and  $d \leq e^{d^\circ} - 1$ .

Proof. Note that  $d(x, y)$  is the infimum of those  $\epsilon > 0$  for which there exists a  $\lambda \in \Lambda$  with

$$\sup_t |\lambda t - t| = \sup_t |t - \lambda^{-1}t| < \epsilon,$$

$$\sup_t |x(t) - y(\lambda t)| = \sup_t |x(\lambda^{-1}t) - y(t)| < \epsilon.$$

Of course  $d(x, y) \geq 0$ ,  $d(x, y) = 0$  implies  $x = y$ , and  $d(x, y) = d(y, x)$ . To see that  $d$  is a metric we have to check the triangle inequality  $d(x, z) \leq d(x, y) + d(y, z)$ . It follows from

$$\|\lambda_1 \lambda_2 - 1\| \leq \|\lambda_1 - 1\| + \|\lambda_2 - 1\|,$$

$$\|x - z\lambda_1 \lambda_2\| \leq \|x - y\lambda_2\| + \|y - z\lambda_1\|.$$

Symmetry and the triangle inequality for  $d^\circ$  follows from  $\|\lambda^{-1}\|^\circ = \|\lambda\|^\circ$  and the inequality

$$\|\lambda_1 \lambda_2\|^\circ \leq \|\lambda_1\|^\circ + \|\lambda_2\|^\circ.$$

That  $d^\circ(x, y) = 0$  implies  $x = y$  follows from  $d \leq e^{d^\circ} - 1$  which is a consequence of  $\|x - y\lambda\| \leq e^{\|x - y\lambda\|} - 1$  and

$$\|\lambda - 1\| = \sup_{0 \leq t \leq 1} t \left| \frac{\lambda t - t}{t} - 1 \right| \leq e^{\|\lambda\|^\circ} - 1.$$

The last inequality uses  $|u - 1| \leq e^{|u|} - 1$  for  $u > 0$ .

**Example 6.23** Consider  $j_x$ , the maximum jump in  $x \in \mathbf{D}$ . Clearly,  $|j_x - j_y| < \epsilon$  if  $\|x - y\| < \epsilon/2$ , and so  $j_x$  is continuous in the uniform topology. It is also continuous in the Skorokhod topology. Indeed, if  $d(x, y) < \epsilon/2$ , then there is a  $\lambda$  such that  $\|\lambda - 1\| < \epsilon/2$  and  $\|x - y\lambda\| < \epsilon/2$ . Since  $j_y = j_{y\lambda}$ , we conclude using continuity in the uniform topology  $|j_x - j_y| = |j_x - j_{y\lambda}| < \epsilon$ .

**Lemma 6.24** If  $d(x, y) = \delta^2$  and  $\delta \leq 1/3$ , then  $d^\circ(x, y) \leq 4\delta + w'_x(\delta)$ .

Proof. We prove that if  $d(x, y) < \delta^2$  and  $\delta \leq 1/3$ , then  $d^\circ(x, y) < 4\delta + w'_x(\delta)$ . Choose  $\mu \in \Lambda$  such that  $\|\mu - 1\| < \delta^2$  and  $\|x\mu^{-1} - y\| < \delta^2$ . Take  $\{t_i\}$  to be a  $\delta$ -sparse set satisfying  $w_x[t_{i-1}, t_i] \leq w'_x(\delta) + \delta$  for each  $i$ . Take  $\lambda$  to agree with  $\mu$  at the points  $\{t_i\}$  and to be linear in between. Since  $\mu^{-1}\lambda t_i = t_i$ , we have  $t \in [t_{i-1}, t_i)$  if and only if  $\mu^{-1}\lambda t \in [t_{i-1}, t_i)$ , and therefore

$$|x(t) - y(\lambda t)| \leq |x(t) - x(\mu^{-1}\lambda t)| + |x(\mu^{-1}\lambda t) - y(\lambda t)| < w'_x(\delta) + \delta + \delta^2 \leq 4\delta + w'_x(\delta).$$

Now it is enough to verify that  $\|\lambda\|^\circ < 4\delta$ . Draw a picture to see that the slopes of  $\lambda$  are always between  $\frac{\delta \pm 2(\delta^2)}{\delta} = 1 \pm 2\delta$ . Since  $|\log(1 \pm 2\delta)| < 4\delta$  for sufficiently small  $\delta$ , we get  $\|\lambda\|^\circ < 4\delta$ .

**Theorem 6.25** The metrics  $d$  and  $d^\circ$  are equivalent and generate the same, so called Skorokhod topology.

Proof. By definition  $d(x_n, x) \rightarrow 0$  ( $d^\circ(x_n, x) \rightarrow 0$ ) if and only if there is a sequence  $\lambda_n \in \Lambda$  such that  $\|\lambda_n - 1\| \rightarrow 0$  ( $\|\lambda_n\|^\circ \rightarrow 0$ ) and  $\|x_n \lambda_n - x\| \rightarrow 0$ . If  $d^\circ(x_n, x) \rightarrow 0$ , then  $d(x_n, x) \rightarrow 0$  due to  $d \leq e^{d^\circ} - 1$ . The reverse implication follows from Lemma 6.33.

**Definition 6.26** Denote by  $\mathcal{D}$  the Borel  $\sigma$ -algebra formed from the open and closed sets in  $(\mathbf{D}, d)$ , or equivalently  $(\mathbf{D}, d^\circ)$ , using the operations of countable intersection, countable union, and set difference.

**Lemma 6.27** Skorokhod convergence  $x_n \rightarrow x$  in  $\mathbf{D}$  implies  $x_n(t) \rightarrow x(t)$  for continuity points  $t$  of  $x$ . Moreover, if  $x$  is continuous on  $[0, 1]$ , then Skorokhod convergence implies uniform convergence.

Proof. Let  $\lambda_n \in \Lambda$  be such that  $\|\lambda_n - 1\| \rightarrow 0$  and  $\|x_n - x\lambda_n\| \rightarrow 0$ . The first assertion follows from

$$|x_n(t) - x(t)| \leq |x_n(t) - x(\lambda_n t)| + |x(\lambda_n t) - x(t)|.$$

The second assertion is obtained from

$$\|x_n - x\| \leq \|x_n - x\lambda_n\| + w_x(\|\lambda_n - 1\|).$$

**Example 6.28** Put  $x_n(t) = 1_{\{t \in [a-2^{-n}, 1]\}} + 1_{\{t \in [a+2^{-n}, 1]\}}$  and  $x(t) = 2 \cdot 1_{\{t \in [a, 1]\}}$  for some  $a \in (0, 1)$ . We have  $x_n(t) \rightarrow x(t)$  for continuity points  $t$  of  $x$ , however  $x_n$  does not converge to  $x$  in the Skorokhod topology.

**Exercise 6.29** Fix  $\delta \in (0, 1)$  and consider  $w'_x(\delta)$  as a function of  $x \in \mathbf{D}$ .

(i) The function  $w'_x(\delta)$  is continuous with respect to the uniform metric since

$$|w'_x(\delta) - w'_y(\delta)| \leq 2\|x - y\|.$$

Hint: show that  $w_x[t_{i-1}, t_i] \leq w_y[t_{i-1}, t_i] + 2\|x - y\|$ .

(ii) However  $w'_x(\delta)$  is not continuous with respect to  $(\mathbf{D}, d)$ . Verify this using  $x_n = 1_{[0, \delta+2^{-n})}$  and  $x = 1_{[0, \delta]}$ .

**Exercise 6.30** Show that  $h(x) = \sup_t x(t)$  is a continuous mapping from  $\mathbf{D}$  to  $\mathbf{R}$ . Hint: show first that for any  $\lambda \in \Lambda$ ,

$$|h(x) - h(y)| \leq \|x - y\|.$$

### 6.3 Separability and completeness of $\mathbf{D}$

**Lemma 6.31** Given  $0 = s_0 < s_1 < \dots < s_k = 1$  define a non-decreasing map  $\kappa : [0, 1] \rightarrow [0, 1]$  by setting

$$\kappa t = \begin{cases} s_{j-1} & \text{for } t \in [s_{j-1}, s_j), \quad j = 1, \dots, k, \\ 1 & \text{for } t = 1. \end{cases}$$

If  $\max_j (s_j - s_{j-1}) \leq \delta$ , then  $d(x\kappa, x) \leq \delta \vee w'_x(\delta)$  for any  $x \in \mathbf{D}$ .

Proof. Given  $\epsilon > 0$  find a  $\delta$ -sparse set  $\{t_i\}$  satisfying  $w_x[t_{i-1}, t_i] \leq w'_x(\delta) + \epsilon$  for each  $i$ . Let  $\lambda \in \Lambda$  be linear between  $\lambda t_i = s_j$  for  $t_i \in (s_{j-1}, s_j]$ ,  $j = 1, \dots, k$  and  $\lambda 0 = 0$ . Since  $\|\lambda - 1\| \leq \delta$ , it suffices to show that  $|x(\kappa t) - x(\lambda^{-1}t)| \leq w'_x(\delta) + \epsilon$ . This holds if  $t$  is 0 or 1, and it is enough to show that, for  $t \in (0, 1)$ , both  $\kappa t$  and  $\lambda^{-1}t$  lie in the same  $[t_{i-1}, t_i]$ , see Figure 5. We prove this by showing that  $\kappa t < t_i$  is equivalent to  $\lambda^{-1}t < t_i$ ,  $i = 1, \dots, k$ . Suppose that  $t_i \in (s_{j-1}, s_j]$ . Then

$$\kappa t < t_i \quad \Rightarrow \quad \kappa t < s_j \quad \Rightarrow \quad \kappa t \leq s_{j-1} \quad \Rightarrow \quad \kappa t < t_i.$$

Thus  $\kappa t < t_i$  is equivalent to  $\kappa t < s_j$  which in turn is equivalent to  $t < s_j$ . On the other hand,  $\lambda t_i = s_j$ , and hence  $t < s_j$  is equivalent to  $t < \lambda t_i$  or  $\lambda^{-1}t < t_i$ .

**Example 6.32** Let  $x_n(t) = 1_{\{t \in [t_0, t_0+2^{-n})\}}$ . Since  $d(x_n, x_{n+1}) = 2^{-n} - 2^{-n-1} = 2^{-n-1}$ , the sequence  $(x_n)$  is fundamental. However, it is not  $d$ -convergent. Indeed,  $x_n(t) \rightarrow 0$  for all  $t \neq t_0$  and Skorokhod convergence  $x_n \rightarrow x$  in  $\mathbf{D}$  by Lemma 6.27, should imply  $x(t) = 0$  for all points of continuity of  $x$ . Since  $x \in \mathbf{D}$  has at most countably many points of discontinuities, by right continuity we conclude that  $x \equiv 0$ . Moreover, since the limit  $x \equiv 0$  is continuous, we must have  $\|x_n - x\| \rightarrow 0$ . But  $\|x_n\| \equiv 1$ .

**Theorem 6.33** The space  $\mathbf{D}$  is separable under  $d$  and  $d^\circ$ , and is complete under  $d^\circ$ .

Proof. *Separability* for  $d$ . Put  $s_j = j/k$ ,  $j = 1, \dots, k$ . Let  $B_k$  be the set of functions having a constant, rational value over each  $[s_{j-1}, s_j)$  and a rational value at  $t = 1$ . Then  $B = \cup B_k$  is countable. Now it is enough to prove that given  $x \in \mathbf{D}$  and  $\epsilon > 0$  we can find some  $y \in B$  such that  $d(x, y) < 2\epsilon$ . Choosing  $k$  such that  $k^{-1} < \epsilon$  and  $w'_x(k^{-1}) < \epsilon$  we can find  $y \in B_k$  satisfying  $d(x\kappa, y) < \epsilon$ , for  $\kappa$  defined as in Lemma 6.31. It remains to see that  $d(x\kappa, x) < \epsilon$  according to Lemma 6.31.

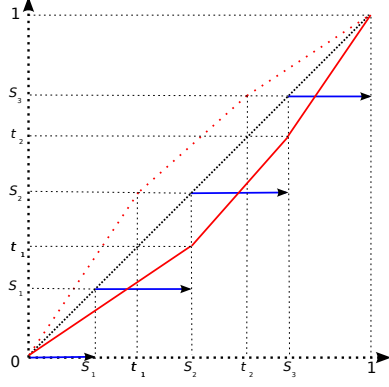


Figure 5: Lemma 6.31. Comparing the blue graph of the function  $\kappa$  to the red graph of the function  $\lambda^{-1}$ .

*Completeness.* We show that any  $d^\circ$ -fundamental sequence  $x_n \in \mathbf{D}$  contains a subsequence  $y_k = x_{n_k}$  that is  $d^\circ$ -convergent. Choose  $n_k$  in such a way that  $d^\circ(y_k, y_{k+1}) < 2^{-k}$ . Then  $\Lambda$  contains  $\mu_k$  such that  $\|\mu_k\|^\circ < 2^{-k}$  and  $\|y_k \mu_k - y_{k+1}\| < 2^{-k}$ .

We suggest a choice of  $\lambda_k \in \Lambda$  such that  $\|\lambda_k\|^\circ \rightarrow 0$  and  $\|y_k \lambda_k - y\| \rightarrow 0$  for some  $y \in \mathbf{D}$ . To this end put  $\mu_{k,m} = \mu_k \mu_{k+1} \dots \mu_{k+m}$ . From

$$\begin{aligned} \|\mu_{k,m+1} - \mu_{k,m}\| &\leq \|\mu_{k+m+1} - 1\| \cdot e^{\|\mu_k \mu_{k+1} \dots \mu_{k+m}\|^\circ} \\ &\leq (e^{\|\mu_{k+m+1}\|^\circ} - 1) \cdot e^{\|\mu_k\|^\circ + \dots + \|\mu_{k+m}\|^\circ} \\ &\leq 2^{-k-m} \cdot e^{2^{-k} + \dots + 2^{-k-m}} < 2^{-k-m+2} \end{aligned}$$

we conclude that for a fixed  $k$  the sequence of functions  $\mu_{k,m}$  is uniformly fundamental. Thus there exists a  $\lambda_k$  such that  $\|\mu_{k,m} - \lambda_k\| \rightarrow 0$  as  $m \rightarrow \infty$ . To prove that  $\lambda_k \in \Lambda$  we use

$$\log \left| \frac{\mu_{k,m} t - \mu_{k,m} s}{t - s} \right| \leq \|\mu_k \mu_{k+1} \dots \mu_{k+m}\|^\circ < 2^{-k+1}.$$

Letting here  $m \rightarrow \infty$  we get  $\|\lambda_k\|^\circ \leq 2^{-k+1}$ . Since  $\|\lambda_k\|^\circ$  is finite we conclude that  $\lambda_k$  is strictly increasing and therefore  $\lambda_k \in \Lambda$ .

Finally, observe that

$$\|y_k \lambda_k - y_{k+1} \lambda_{k+1}\| = \|y_k \mu_k \lambda_{k+1} - y_{k+1} \lambda_{k+1}\| = \|y_k \mu_k - y_{k+1}\| < 2^{-k}.$$

It follows, that the sequence  $y_k \lambda_k \in \mathbf{D}$  is uniformly fundamental and hence  $\|y_k \lambda_k - y\| \rightarrow 0$  for some  $y$ . Observe that  $y$  must lie in  $\mathbf{D}$ . Since  $\|\lambda_k\|^\circ \rightarrow 0$ , we obtain  $d^\circ(y_k, y) \rightarrow 0$ .

## 6.4 Relative compactness in the Skorokhod topology

First comes an analogue of the Arzela-Ascoli theorem in terms of  $w'_x(\delta)$ , and then a convenient alternative in terms of  $w''_x(\delta)$ .

**Theorem 6.34** *A set  $A \subset \mathbf{D}$  is relatively compact in the Skorokhod topology iff*

- (i)  $\sup_{x \in A} \|x\| < \infty$ ,
- (ii)  $\limsup_{\delta \rightarrow 0} w'_x(\delta) = 0$ .

Proof of sufficiency only. Put  $\alpha = \sup_{x \in A} \|x\|$ . For a given  $\epsilon > 0$ ,

put  $H_\epsilon = \{\alpha_i\}$ , where  $-\alpha = \alpha_0 < \alpha_1 < \dots < \alpha_k = \alpha$  and  $\alpha_j - \alpha_{j-1} \leq \epsilon$ ,

and choose  $\delta < \epsilon$  so that  $w'_x(\delta) < \epsilon$  for all  $x \in A$ .

According to Lemma 6.31 for any  $\kappa = \kappa_{\{s_j\}}$  satisfying  $\max_j (s_{j-1} - s_j) \leq \delta$ , we have  $d(x\kappa, x) \leq \epsilon$  for all  $x \in A$ . Take  $B_\epsilon$  be the set of  $y \in \mathbf{D}$  that assume on each  $[s_{j-1}, s_j]$  a constant value from  $H_\epsilon$  and  $y(1) \in H_\epsilon$ . For any  $x \in A$  there is a  $y \in B_\epsilon$  such that  $d(x\kappa, y) \leq \epsilon$ . Thus  $B_\epsilon$  forms a  $2\epsilon$ -net for  $A$  in the sense of  $d$  and  $A$  is totally bounded in the sense of  $d$ .

But we must show that  $A$  is totally bounded in the sense of  $d^\circ$ , since this is the metric under which  $\mathbf{D}$  is complete. This is true as according Lemma 6.33, the set  $B_{\delta^2}$  is an  $\epsilon'$ -net for  $A$ , where  $\epsilon' = 4\delta + \sup_{x \in A} w'_x(\delta)$  can be chosen arbitrary small.

**Theorem 6.35** A set  $A \subset \mathbf{D}$  is relatively compact in the Skorokhod topology iff

$$(i) \quad \sup_{x \in A} \|x\| < \infty,$$

$$(ii) \quad \begin{cases} \lim_{\delta \rightarrow 0} \sup_{x \in A} w_x''(\delta) = 0, \\ \lim_{\delta \rightarrow 0} \sup_{x \in A} |x(\delta) - x(0)| = 0, \\ \lim_{\delta \rightarrow 0} \sup_{x \in A} |x(1-\delta) - x(1)| = 0. \end{cases}$$

Proof. It is enough to show that (ii) of Theorem 6.35 is equivalent to (ii) of Theorem 6.34. This follows from Lemma 6.14.

## 7 Probability measures on $\mathbf{D}$ and random elements

### 7.1 Finite-dimensional distributions on $\mathbf{D}$

Finite-dimensional sets play in  $\mathbf{D}$  the same role as they do in  $\mathbf{C}$ .

**Definition 7.1** Consider projection mappings  $\pi_{t_1, \dots, t_k} : \mathbf{D} \rightarrow \mathbf{R}^k$ . For  $T \subset [0, 1]$ , define in  $\mathcal{D}$  the subclass  $\mathcal{D}_f(T)$  of finite-dimensional sets  $\pi_{t_1, \dots, t_k}^{-1}(H)$ , where  $k$  is arbitrary,  $t_i$  belong to  $T$ , and  $H \in \mathcal{R}^k$ .

**Theorem 7.2** Consider projection mappings  $\pi_{t_1, \dots, t_k} : \mathbf{D} \rightarrow \mathbf{R}^k$ . The following three statements hold.

- (a) The projections  $\pi_0$  and  $\pi_1$  are continuous, and for  $t \in (0, 1)$ ,  $\pi_t$  is continuous at  $x$  if and only if  $x$  is continuous at  $t$ .
- (b) Each  $\pi_{t_1, \dots, t_k}$  is a measurable map.
- (c) If  $T$  contains 1 and is dense in  $[0, 1]$ , then  $\sigma\{\pi_t : t \in T\} = \sigma\{\mathcal{D}_f(T)\} = \mathcal{D}$  and  $\mathcal{D}_f(T)$  is a separating class.

Proof. (a) Since each  $\lambda \in \Lambda$  fixes 0 and 1,  $\pi_0$  and  $\pi_1$  are continuous: for  $i = 0, 1$ ,

$$d(x, y) \geq \inf_{\lambda \in \Lambda} \|x - y\lambda\| \geq |x(i) - y(i)| = |\pi_i(x) - \pi_i(y)|.$$

Suppose that  $0 < t < 1$ . If  $x$  is continuous at  $t$ , then by Lemma 6.27,  $\pi_t$  is continuous at  $x$ . Suppose, on the other hand, that  $x$  is discontinuous at  $t$ . If  $\lambda_n \in \Lambda$  carries  $t$  to  $t - 1/n$  and is linear on  $[0, t]$  and  $[t, 1]$ , and if  $x_n(s) = x(\lambda_n s)$ , then  $d(x_n, x) \rightarrow 0$  but  $x_n(t) \not\rightarrow x(t)$ .

(b) A mapping into  $\mathbf{R}^k$  is measurable if each component mapping is. Therefore it suffices to show that  $\pi_t$  is measurable. Since  $\pi_1$  is continuous, we may assume  $t < 1$ . We use the pointwise convergence

$$h_\epsilon(x) := \epsilon^{-1} \int_t^{t+\epsilon} x_s ds \rightarrow \pi_t(x), \quad \epsilon \rightarrow 0, \quad x \in \mathbf{D}.$$

If  $x^n \rightarrow x$  in the Skorokhod topology, then  $x_t^n \rightarrow x_t$  for continuity points  $t$  of  $x$ , and we conclude that  $x^n \rightarrow x$  almost surely. By the Bounded Convergence Theorem, the almost sure convergence  $x^n \rightarrow x$  and the uniform boundedness of  $(x^n)$  imply  $h_\epsilon(x^n) \rightarrow h_\epsilon(x)$ . Thus for each  $\epsilon$ ,  $h_\epsilon$  is continuous and therefore measurable, implying that its limit  $\pi_t$  is also measurable.

(c) By right-continuity and the assumption that  $T$  is dense, it follows that  $\pi_0$  is measurable with respect to  $\sigma\{\mathcal{D}_f(T)\}$ . So we may as well assume that  $0 \in T$ .

Suppose  $s_0, \dots, s_k$  are points in  $T$  satisfying  $0 = s_0 < \dots < s_k = 1$ . For  $\alpha = (\alpha_0, \dots, \alpha_k) \in \mathbf{R}^{k+1}$  define  $V\alpha \in \mathbf{D}$  by

$$(V\alpha)(t) = \begin{cases} \alpha_{j-1} & \text{for } t \in [s_{j-1}, s_j), \quad j = 1, \dots, k, \\ \alpha_k & \text{for } t = 1. \end{cases}$$

Clearly,  $V : \mathbf{R}^{k+1} \rightarrow \mathbf{D}$  is continuous implying that  $\kappa = V\pi_{s_0, \dots, s_k}$  is measurable  $\sigma\{\mathcal{D}_f(T)\}/\mathcal{D}$ .

Since  $T$  is dense, for any  $n$  we can choose  $s_0^n, \dots, s_k^n$  so that  $\max_i (s_i^n - s_{i-1}^n) < n^{-1}$ . Put  $\kappa^n = V\pi_{s_0^n, \dots, s_k^n}$ . With this choice define a map  $A_n : \mathbf{D} \rightarrow \mathbf{D}$  by  $A_n x = x\kappa^n$ . By Lemma 6.31,  $A_n x \rightarrow x$  for each  $x$ . We conclude that the identity map is measurable  $\sigma\{\mathcal{D}_f(T)\}/\mathcal{D}$  and therefore  $\mathcal{D} \subset \sigma\{\mathcal{D}_f(T)\}$ . Finally, since  $\mathcal{D}_f(T)$  is a  $\pi$ -system, it is a separating class.

**Definition 7.3** Let  $D_c$  be the set of count paths: nondecreasing functions  $x \in \mathbf{D}$  with  $x(t) \in \mathbb{Z}$  for each  $t$ , and  $x(t) - x(t-) = 1$  at points of discontinuity.

**Exercise 7.4** Find  $d(x, y)$  for  $x, y \in D_c$  in terms of the jump points of these two count paths. How does a fundamental sequence  $(x_n)$  in  $D_c$  look for large  $n$ ? Show that  $D_c$  is closed in the Skorokhod topology.

**Lemma 7.5** Let  $T_0 = \{t_1, t_2, \dots\}$  be a countable, dense set in  $[0, 1]$ , and put  $\pi(x) = (x(t_1), x(t_2), \dots)$ .

(a) The mapping  $\pi : \mathbf{D} \rightarrow \mathbf{R}^\infty$  is  $\mathcal{D}/\mathcal{R}^\infty$ -measurable.

(b) If  $x, x_n \in D_c$  are such that  $\pi(x_n) \rightarrow \pi(x)$ , then  $x_n \rightarrow x$  in the Skorokhod topology.

Proof. (a) In terms of notation of Section 2.3,

$$\pi^{-1}(\pi_k^{-1}H) = \pi_{t_1, \dots, t_k}^{-1}H \in \mathcal{D} \quad \text{for } H \in \mathcal{R}^k,$$

and the finite-dimensional sets  $\pi_k^{-1}H$  generate  $\mathcal{R}^\infty$ .

(b) Convergence  $\pi(x_n) \rightarrow \pi(x)$  implies  $x_n(t_i) \rightarrow x(t_i)$ , which in turn means that  $x_n(t_i) = x(t_i)$  for  $n > n_i$ , for all  $i = 1, 2, \dots$ . A function in  $x \in D_c$  has only finitely many discontinuities, say  $0 < s_1 < \dots < s_k \leq 1$ . For a given  $\epsilon$  choose points  $u_i$  and  $v_i$  in  $T_0$  in such a way that  $u_i < s_i \leq v_i < u_i + \epsilon$  and the intervals  $[v_{i-1}, u_i]$ ,  $i = 1, \dots, k$  are disjoint, with  $v_0 = 0$ . Then for  $n$  exceeding some  $n_0$ ,  $x_n$  agrees with  $x$  over each  $[v_{i-1}, u_i]$  and has a single jump in each  $[u_i, v_i]$ . If  $\lambda_n \in \Lambda$  carries  $s_i$  to the point in  $[u_i, v_i]$  where  $x_n$  has a jump and is defined elsewhere by linearity, then  $\|\lambda_n - 1\| \leq \epsilon$  and  $x_n(\lambda_n t) \equiv x(t)$  implying  $d(x_n, x) \leq \epsilon$  for  $n > n_0$ .

**Theorem 7.6** Let  $T_0$  be a countable, dense set in  $[0, 1]$ . If  $P_n(D_c) = P(D_c) = 1$  and  $P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow P \pi_{t_1, \dots, t_k}^{-1}$  for all  $k$ -tuples in  $T_0$ , then  $P_n \Rightarrow P$ .

Proof. The idea is, in effect, to embed  $D_c$  in  $\mathbf{R}^\infty$  and apply Theorem 2.14. By hypothesis,  $P_n \pi^{-1} \pi_k^{-1} \Rightarrow P \pi^{-1} \pi_k^{-1}$ , but since in  $\mathbf{R}^\infty$  weak convergence is the same thing as weak convergence of finite-dimensional distributions, it follows that  $P_n \pi^{-1} \Rightarrow P \pi^{-1}$  in  $\mathbf{R}^\infty$ . For  $A \subset \mathbf{D}$ , define  $A^* = \pi^{-1}(\pi A)^-$ . If  $A \in \mathcal{D}$ , then

$$\limsup_n P_n(A) \leq \limsup_n P_n(A^*) = \limsup_n P_n(\pi^{-1}(\pi A)^-) \leq P(\pi^{-1}(\pi A)^-) = P(A^*).$$

Therefore, if  $F \in \mathcal{D}$  is closed, then

$$\limsup_n P_n(F) = \limsup_n P_n(F \cap D_c) \leq P((F \cap D_c)^*) = P((F \cap D_c)^* \cap D_c).$$

It remains to show that if  $F \in \mathcal{D}$  is closed, then  $(F \cap D_c)^* \cap D_c \subset F$ . Take an  $x \in (F \cap D_c)^* \cap D_c$ . Since  $x \in (F \cap D_c)^*$ , we have  $\pi(x) \in (\pi(F \cap D_c))^-$  and there is a sequence  $x_n \in F \cap D_c$  such that  $\pi(x_n) \rightarrow \pi(x)$ . Because  $x \in D_c$ , the previous lemma gives  $x_n \rightarrow x$  in the Skorokhod topology. Since  $x_n \in F$  and  $F \in \mathcal{D}$  is closed, we conclude that  $x \in F$ .

**Corollary 7.7** Suppose for each  $n$ ,  $\xi_{n1}, \dots, \xi_{nn}$  are iid indicator r.v. with  $\mathbb{P}(\xi_{ni} = 1) = \alpha/n$ . If  $X_t^n = \sum_{i \leq nt} \xi_{ni}$ , then  $X^n \Rightarrow X$  in  $\mathbf{D}$ , where  $X$  is the Poisson process with parameter  $\alpha$ .

Proof. The random process  $X_t^n = \sum_{i \leq nt} \xi_{ni}$  has independent increments. Its finite-dimensional distributions weakly converge to that of the Poisson process  $X$  with

$$\mathbb{P}(X_t - X_s = k) = \frac{\alpha^k (t-s)^k}{k!} e^{-\alpha(t-s)} \quad \text{for } 0 \leq s < t \leq 1.$$

**Exercise 7.8** Suppose that  $\xi$  is uniformly distributed over  $[\frac{1}{3}, \frac{2}{3}]$ , and consider the random functions

$$X_t = 2 \cdot 1_{\{t \in [\xi, 1]\}}, \quad X_t^n = 1_{\{t \in [\xi - n^{-1}, 1]\}} + 1_{\{t \in [\xi + n^{-1}, 1]\}}.$$

Show that  $X^n \not\Rightarrow X$ , even though  $(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow (X_{t_1}, \dots, X_{t_k})$  for all  $(t_1, \dots, t_k)$ . Why does Theorem 7.6 not apply?

**Lemma 7.9** Let  $P$  be a probability measure on  $(\mathbf{D}, \mathcal{D})$ . Define  $T_P \subset [0, 1]$  as the collection of  $t$  such that the projection  $\pi_t$  is  $P$ -almost surely continuous. The set  $T_P$  contains 0 and 1, and its complement in  $[0, 1]$  is at most countable. For  $t \in (0, 1)$ ,  $t \in T_P$  is equivalent to  $P\{x : x(t) \neq x(t-)\} = 0$ .

Proof. Recall Theorem 7.2 (a) and put  $J_t = \{x : x(t) \neq x(t-)\}$  for a  $t \in (0, 1)$ . We have to show that  $P(J_t) > 0$  is possible for at most countably many  $t$ . Let  $J_t(\epsilon) = \{x : |x(t) - x(t-)| > \epsilon\}$ . For fixed, positive  $\epsilon$  and  $\delta$ , there can be at most finitely many  $t$  for which  $P(J_t(\epsilon)) \geq \delta$ . Indeed, if  $P(J_{t_n}(\epsilon)) \geq \delta$  for infinitely many distinct  $t_n$ , then

$$P(J_{t_n}(\epsilon) \text{ i.o.}) = P(\limsup_n J_{t_n}(\epsilon)) \geq \limsup_n P(J_{t_n}(\epsilon)) \geq \delta,$$

contradicting the fact that for a single  $x \in \mathbf{D}$  the jumps can exceed  $\epsilon$  at only finitely many points, see Lemma 6.4. Thus  $P(J_t(\epsilon)) > 0$  is possible for at most countably many  $t$ . The desired result follows from

$$\{t : P(J_t) > 0\} = \bigcup_n \{t : P(J_{t(n^{-1})}) > 0\},$$

which in turn is a consequence of  $P(J_t(\epsilon)) \uparrow P(J_t)$  as  $\epsilon \downarrow 0$ .



**Theorem 7.10** Let  $P_n, P$  be probability measures on  $(\mathbf{D}, \mathcal{D})$ . If the sequence  $(P_n)$  is tight and  $P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow P \pi_{t_1, \dots, t_k}^{-1}$  holds whenever  $t_1, \dots, t_k$  lie in  $T_P$ , then  $P_n \Rightarrow P$ .

Proof. We will show that if a subsequence  $(P_{n'}) \subset (P_n)$  converges weakly to some  $Q$ , then  $Q = P$ . Indeed, if  $t_1, \dots, t_k$  lie in  $T_Q$ , then  $\pi_{t_1, \dots, t_k}$  is continuous on a set of  $Q$ -measure 1, and therefore,  $P_{n'} \Rightarrow Q$  implies by the mapping theorem that  $P_{n'} \pi_{t_1, \dots, t_k}^{-1} \Rightarrow Q \pi_{t_1, \dots, t_k}^{-1}$ . On the other hand, if  $t_1, \dots, t_k$  lie in  $T_P$ , then  $P_{n'} \pi_{t_1, \dots, t_k}^{-1} \Rightarrow P \pi_{t_1, \dots, t_k}^{-1}$  by the assumption. Therefore, if  $t_1, \dots, t_k$  lie in  $T_Q \cap T_P$ , then  $Q \pi_{t_1, \dots, t_k}^{-1} = P \pi_{t_1, \dots, t_k}^{-1}$ . It remains to see that  $\mathcal{D}_f(T_Q \cap T_P)$  is a separating class by applying Lemma 7.9 and Theorem 7.2.

## 7.2 Tightness criteria in $\mathbf{D}$

**Theorem 7.11** Let  $P_n$  be probability measures on  $(\mathbf{D}, \mathcal{D})$ . The sequence  $(P_n)$  is tight if and only if the following two conditions hold:

- (i)  $\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} P_n(x : \|x\| \geq a) = 0$ ,
- (ii)  $\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P_n(x : w'_x(\delta) \geq \epsilon) = 0$ , for each positive  $\epsilon$ .

Condition (ii) is equivalent to

$$(ii') \quad \forall \epsilon, \eta > 0; \exists \delta, n_0 > 0 : \quad P_n(x : w'_x(\delta) \geq \epsilon) \leq \eta, \text{ for } n > n_0.$$

Proof. This theorem is proven similarly to Theorem 4.25 using Theorem 6.34. Equivalence of (ii) and (ii') is due to monotonicity of  $w'_x(\delta)$ .

**Theorem 7.12** Let  $P_n$  be probability measures on  $(\mathbf{D}, \mathcal{D})$ . The sequence  $(P_n)$  is tight if and only if the following two conditions hold:

- (i)  $\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} P_n(x : \|x\| \geq a) = 0$ ,
- (ii)  $\forall \epsilon, \eta > 0; \exists \delta, n_0 > 0 : \quad \begin{cases} P_n(x : w''_x(\delta) \geq \epsilon) \leq \eta, \\ P_n(x : |x(\delta) - x(0)| \geq \epsilon) \leq \eta, \\ P_n(x : |x(1-\delta) - x(1)| \geq \epsilon) \leq \eta, \end{cases} \text{ for } n > n_0.$

Proof. This theorem follows from Theorem 7.11 with (i) and (ii') using Lemma 6.14. (Recall how Theorem 6.35 was obtained from Theorem 6.34 using Lemma 6.14.)

**Lemma 7.13** Turn to Theorems 7.11 and 7.12. Under (ii) condition (i) is equivalent to the following weaker version:

(i') for each  $t$  in a set  $T$  that is dense in  $[0, 1]$  and contains 1,

$$\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} P_n(x : |x(t)| \geq a) = 0,$$

Proof. The implication (i)  $\Rightarrow$  (i') is trivial. Assume (ii) of Theorem 7.11 and (i'). For a given  $\delta \in (0, 1)$  choose from  $T$  points  $0 < s_1 < \dots < s_k = 1$  such that  $\max\{s_1, s_2 - s_1, \dots, s_k - s_{k-1}\} < \delta$ . By hypothesis (i'), there exists an  $a$  such that

$$P_n(x : \max_j |x(s_j)| \geq a) < \eta, \quad n > n_0 \quad (*)$$

For a given  $x$ , take a  $\delta$ -sparse set  $(t_0, \dots, t_v)$  such that all  $w_x[t_{i-1}, t_i] < w'_x(\delta) + 1$ . Since each  $[t_{i-1}, t_i]$  contains an  $s_j$ , we have

$$\|x\| \leq \max_j |x(s_j)| + w'_x(\delta) + 1.$$

Using (ii') of Theorem 7.11 and (\*), we get  $P_n(x : \|x\| \geq a + 2) < 2\eta$  implying (i).

## 7.3 A key condition on 3-dimensional distributions

The following condition plays an important role.

**Definition 7.14** For a probability measure  $P$  on  $\mathbf{D}$ , we will write  $P \in H_{\alpha, \beta}$ , if there exist  $\alpha > 1$ ,  $\beta \geq 0$ , and a nondecreasing continuous  $H : [0, 1] \rightarrow \mathbf{R}$  such that for all  $\epsilon > 0$  and all  $0 \leq t_1 \leq t_2 \leq t_3 \leq 1$ ,

$$P \pi_{t_1, t_2, t_3}^{-1} \{(z_1, z_2, z_3) : |z_2 - z_1| \geq \epsilon, |z_3 - z_2| \geq \epsilon\} \leq \epsilon^{-2\beta} (H(t_3) - H(t_1))^\alpha.$$

For a random element  $X$  on  $\mathbf{D}$  with probability distribution  $P$ , this condition  $P \in H_{\alpha, \beta}$  means that for all  $\epsilon > 0$  and all  $0 \leq r \leq s \leq t \leq 1$

$$\mathbb{P}(|X_s - X_r| \geq \epsilon, |X_t - X_s| \geq \epsilon) \leq \epsilon^{-2\beta} (H(t) - H(r))^\alpha.$$

**Lemma 7.15** *Let a random element  $X$  on  $\mathbf{D}$  have a probability distribution  $P \in H_{\alpha,\beta}$ . Then there is a constant  $K_{\alpha,\beta}$  depending only on  $\alpha$  and  $\beta$  such that*

$$\mathbb{P}\left(\sup_{r \leq s \leq t} (|X_s - X_r| \wedge |X_t - X_s|) \geq \epsilon\right) \leq \frac{K_{\alpha,\beta}}{\epsilon^{2\beta}} (H(1) - H(0))^\alpha.$$

Proof. The stated estimate is obtained in four consecutive steps.

Step 1. Let  $T_k = \{i/2^k, 0 \leq i \leq 2^k\}$  and

$$\begin{aligned} A_k &= \max(|X_s - X_r| \wedge |X_t - X_s| \text{ over the adjacent triplets } r \leq s \leq t \text{ in } T_k), \\ B_k &= \max(|X_s - X_r| \wedge |X_t - X_s| \text{ over } r \leq s \leq t \text{ from } T_k). \end{aligned}$$

We will show that  $B_k \leq 2(A_1 + \dots + A_k)$ . To this end, for each  $t \in T_k$  define a  $t_n \in T_{k-1}$  by

$$t_n = \begin{cases} t & \text{if } t \in T_{k-1}, \\ t - 2^{-k} & \text{if } t \notin T_{k-1} \text{ and } |X_t - X_{t-2^{-k}}| \leq |X_t - X_{t+2^{-k}}|, \\ t + 2^{-k} & \text{if } t \notin T_{k-1} \text{ and } |X_t - X_{t-2^{-k}}| > |X_t - X_{t+2^{-k}}|, \end{cases}$$

so that  $|X_t - X_{t_n}| \leq A_k$ . Then for any triplet  $r \leq s \leq t$  from  $T_k$ ,

$$\begin{aligned} |X_s - X_r| &\leq |X_s - X_{s_n}| + |X_{s_n} - X_{r_n}| + |X_{r_n} - X_r| \leq |X_{s_n} - X_{r_n}| + 2A_k, \\ |X_t - X_s| &\leq |X_{t_n} - X_{s_n}| + 2A_k. \end{aligned}$$

Since here  $r_n \leq s_n \leq t_n$  lie in  $T_{k-1}$ , it follows that  $|X_s - X_r| \wedge |X_t - X_s| \leq B_{k-1} + 2A_k$ , and therefore,

$$B_k \leq B_{k-1} + 2A_k \leq 2(A_1 + \dots + A_k), \quad k \geq 1.$$

Step 2. Consider a special case when  $H(t) \equiv t$ . Using the right continuity of the paths we get from step 1 that

$$\sup_{r \leq s \leq t} (|X_s - X_r| \wedge |X_t - X_s|) \leq 2 \sum_{k=1}^{\infty} A_k.$$

This implies that for any  $\theta \in (0, 1)$ ,

$$\begin{aligned} \mathbb{P}\left(\sup_{r \leq s \leq t} (|X_s - X_r| \wedge |X_t - X_s|) \geq 2\epsilon\right) &\leq \mathbb{P}\left(\sum_{k=1}^{\infty} A_k \geq \epsilon\right) \\ &\leq \mathbb{P}\left(\sum_{k=1}^{\infty} A_k \geq \epsilon(1-\theta) \sum_{k=1}^{\infty} \theta^k\right) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k \geq \epsilon(1-\theta)\theta^k) \\ &\leq \sum_{k=1}^{\infty} \sum_{i=1}^{2^k-1} \mathbb{P}(|X_{i/2^k} - X_{(i-1)/2^k}| \wedge |X_{(i+1)/2^k} - X_{i/2^k}| \geq \epsilon(1-\theta)\theta^k). \end{aligned}$$

Applying the key condition with  $H(t) \equiv t$  we derive from the previous relation choosing a  $\theta \in (0, 1)$  satisfying  $\theta^{2\beta} > 2^{1-\alpha}$ , that the stated estimate holds in the special case

$$\begin{aligned} \mathbb{P}\left(\sup_{r \leq s \leq t} (|X_s - X_r| \wedge |X_t - X_s|) \geq 2\epsilon\right) &\leq \sum_{k=1}^{\infty} 2^k \frac{2^{(1-k)\alpha}}{(\epsilon(1-\theta)\theta^k)^{2\beta}} \\ &= \frac{2^\alpha}{\epsilon^{2\beta}(1-\theta)^{2\beta}} \sum_{k=1}^{\infty} (\theta^{-2\beta} 2^{1-\alpha})^k. \end{aligned}$$

Step 3. For a strictly increasing  $H(t)$  take  $a$  so that  $a^{2\beta} H(1)^\alpha = 1$ , and define a new process  $Y_t$  by  $Y_t = aX_{b(t)}$ , where the time change  $b(t)$  is such that  $H(b(t)) = tH(1)$ . Since

$$\begin{aligned} \mathbb{P}\left(|Y_s - Y_r| \geq \epsilon; |Y_t - Y_s| \geq \epsilon\right) &= \mathbb{P}\left(|X_{b(s)} - X_{b(r)}| \geq a^{-1}\epsilon; |X_{b(t)} - X_{b(s)}| \geq a^{-1}\epsilon\right) \\ &\leq \epsilon^{-2\beta} (t-r)^\alpha, \end{aligned}$$

we can apply the result of the step 2 to the new process and prove the statement of the lemma under the assumption of step 3.

Step 4. If  $H(t)$  is not strictly increasing, put  $H_v(t) = H(t) + vt$  for an arbitrary small positive  $v$ . We have

$$\mathbb{P}\left(|X_s - X_r| \geq \epsilon; |X_t - X_s| \geq \epsilon\right) \leq \epsilon^{-2\beta} (H_v(t) - H_v(r))^\alpha,$$

and according to step 3

$$\mathbb{P}\left(\sup_{r \leq s \leq t} (|X_s - X_r| \wedge |X_t - X_s|) \geq \epsilon\right) \leq \frac{K_{\alpha,\beta}}{\epsilon^{2\beta}} (H(1) + v - H(0))^\alpha.$$

It remains to let  $v$  go to 0. Lemma 7.15 is proven.

**Lemma 7.16** *If  $P \in H_{\alpha,\beta}$ , then given a positive  $\epsilon$ ,*

$$P(x : w_x''(\delta) \geq \epsilon) \leq \frac{2K_{\alpha,\beta}}{\epsilon^{2\beta}} (H(1) - H(0))(w_H(2\delta))^{\alpha-1},$$

so that  $P(x : w_x''(\delta) \geq \epsilon) \rightarrow 0$  as  $\delta \rightarrow 0$ .

Proof. Take  $t_i = i\delta$  for  $0 \leq i \leq \lceil 1/\delta \rceil$  and  $t_{\lceil 1/\delta \rceil} = 1$ . If  $|t - r| \leq \delta$ , then  $r$  and  $t$  lie in the same  $[t_{i-1}, t_{i+1}]$  for some  $1 \leq i \leq \lceil 1/\delta \rceil - 1$ . According to Lemma 7.15, for  $X$  with distribution  $P$ ,

$$\mathbb{P}\left(\sup_{t_{i-1} \leq r \leq s \leq t \leq t_{i+1}} (|X_s - X_r| \wedge |X_t - X_s|) \geq \epsilon\right) \leq \frac{K_{\alpha,\beta}}{\epsilon^{2\beta}} (H(t_{i+1}) - H(t_{i-1}))^\alpha,$$

and due to monotonicity of  $H$ ,

$$\begin{aligned} \mathbb{P}(w''(X, \delta) \geq \epsilon) &\leq \sum_{i=1}^{\lceil 1/\delta \rceil - 1} \frac{K_{\alpha,\beta}}{\epsilon^{2\beta}} (H(t_{i+1}) - H(t_{i-1}))^\alpha \\ &\leq \frac{K_{\alpha,\beta}}{\epsilon^{2\beta}} \left( \sup_{0 \leq t \leq 1-2\delta} (H(t+2\delta) - H(t))^{\alpha-1} \right) 2(H(1) - H(0)) \\ &= \frac{2K_{\alpha,\beta}}{\epsilon^{2\beta}} (H(1) - H(0))(w_H(2\delta))^{\alpha-1}. \end{aligned}$$

It remains to recall that the modulus of continuity  $w_H(2\delta)$  of the uniformly continuous function  $H$  converges to 0 as  $\delta \rightarrow 0$ .

## 7.4 A criterion for existence

**Theorem 7.17** *There exists in  $\mathbf{D}$  a random element with finite dimensional distributions  $\mu_{t_1, \dots, t_k}$  provided the following three conditions:*

- (i) *the finite dimensional distributions  $\mu_{t_1, \dots, t_k}$  are consistent, see Definition 2.15,*
- (ii) *there exist  $\alpha > 1$ ,  $\beta \geq 0$ , and a nondecreasing continuous  $H : [0, 1] \rightarrow \mathbf{R}$  such that for all  $\epsilon > 0$  and all  $0 \leq t_1 \leq t_2 \leq t_3 \leq 1$ ,*

$$\mu_{t_1, t_2, t_3} \{(z_1, z_2, z_3) : |z_2 - z_1| \geq \epsilon, |z_3 - z_2| \geq \epsilon\} \leq \epsilon^{-2\beta} (H(t_3) - H(t_1))^\alpha,$$

- (iii)  *$\mu_{t, t+\delta} \{(z_1, z_2) : |z_2 - z_1| \geq \epsilon\} \rightarrow 0$  as  $\delta \downarrow 0$  for each  $t \in [0, 1]$ .*

Proof. The main idea, as in the proof of Theorem 4.15 (a), is to construct a sequence  $(X^n)$  of random elements in  $\mathbf{D}$  such that the corresponding sequence of distributions  $(P_n)$  is tight and has the desired limit finite dimensional distributions  $\mu_{t_1, \dots, t_k}$ .

Let vector  $(X_{n,0}, \dots, X_{n,2^n})$  have distribution  $\mu_{t_0, \dots, t_{2^n}}$ , where  $t_i \equiv t_i^n = i2^{-n}$ , and define

$$X_t^n = \begin{cases} X_{n,i} & \text{for } t \in [i2^{-n}, (i+1)2^{-n}), \quad i = 0, \dots, 2^n - 1, \\ X_{n,2^n} & \text{for } t = 1. \end{cases}$$

The rest of the proof uses Theorem 7.12 and is divided in four steps.

Step 1. For all  $\epsilon > 0$  and  $r, s, t \in T_n = \{t_0, \dots, t_{2^n}\}$  we have by (ii),

$$\mathbb{P}\left(|X_s^n - X_r^n| \geq \epsilon, |X_t^n - X_s^n| \geq \epsilon\right) \leq \epsilon^{-2\beta} (H(t) - H(r))^\alpha.$$

It follows that in general for  $0 \leq r \leq s \leq t \leq 1$

$$\mathbb{P}\left(|X_s^n - X_r^n| \geq \epsilon, |X_t^n - X_s^n| \geq \epsilon\right) \leq \epsilon^{-2\beta} (H(t) - H(r - 2^{-n}))^\alpha,$$

where  $H(t) = H(0)$  for  $t < 0$ . Slightly modifying Lemma 7.16 we obtain that given a positive  $\epsilon$ , there is a constant  $K_{\alpha,\beta,\epsilon}$

$$P_n(x : w_x''(\delta) \geq \epsilon) \leq K_{\alpha,\beta,\epsilon} (H(1) - H(0))(w_H(3\delta))^{\alpha-1}, \quad \text{for } \delta \leq 2^{-n}.$$

This gives the first part of (ii) in Theorem 7.12.

Step 2. If  $2^{-k} \leq \delta$ , then

$$\|X^n\| \leq \max_{t \in T_k} |X_t^n| + w''(X^n, \delta).$$

Since the distributions of the first term on the right all coincide for  $n \geq k$ , it follows by step 1 that condition (i) in Theorem 7.12 is satisfied.

Step 3. To take care of the second and third parts of (ii) in Theorem 7.12, we fix some  $\delta_0 \in (0, 1/2)$ , and temporarily assume that for  $\delta \in (0, \delta_0)$ ,

$$\mu_{0,\delta}\{(z_1, z_2) : z_1 = z_2\} = 1, \quad \mu_{1-\delta,1}\{(z_1, z_2) : z_1 = z_2\} = 1. \quad (*)$$

In this special case, the second and third parts of (ii) in Theorem 7.12 hold and we conclude that the sequence of distributions  $P_n$  of  $X_n$  is tight.

By Prokhorov's theorem,  $(X^n)$  has a subsequence weakly converging in distribution to a random element  $X$  of  $\mathbf{D}$  with some distribution  $P$ . We want to show that  $P\pi_{t_1, \dots, t_k}^{-1} = \mu_{t_1, \dots, t_k}$ . Because of the consistency hypothesis, this holds for dyadic rational  $t_i \in \cup_n T_n$ . The general case is obtained using the following facts:

$$\begin{aligned} P_n \pi_{t_1, \dots, t_k}^{-1} &\Rightarrow P \pi_{t_1, \dots, t_k}^{-1}, \\ P_n \pi_{t_1, \dots, t_k}^{-1} &= \mu_{t_{n1}, \dots, t_{nk}}, \quad \text{for some } t_{ni} \in T_n, \text{ provided } k \leq 2^n, \\ \mu_{t_{n1}, \dots, t_{nk}} &\Rightarrow \mu_{t_1, \dots, t_k}. \end{aligned}$$

The last fact is a consequence of (iii). Indeed, by Kolmogorov's extension theorem, there exists a stochastic process  $Z$  with vectors  $(Z_{t_1}, \dots, Z_{t_k})$  having distributions  $\mu_{t_1, \dots, t_k}$ . Then by (iii),  $Z_{t+\delta} \xrightarrow{P} Z_t$  as  $\delta \downarrow 0$ . Using Exercise 1.22 we derive  $(Z_{t_{n1}}, \dots, Z_{t_{nk}}) \xrightarrow{P} (Z_{t_1}, \dots, Z_{t_k})$  implying  $\mu_{t_{n1}, \dots, t_{nk}} \Rightarrow \mu_{t_1, \dots, t_k}$ .

Step 4. It remains to remove the restriction (\*). To this end take

$$\lambda t = \begin{cases} 0 & \text{for } t \leq \delta_0, \\ \frac{t-\delta_0}{1-2\delta_0} & \text{for } \delta_0 < t < 1-\delta_0, \\ 1 & \text{for } t \geq 1-\delta_0. \end{cases}$$

Define  $\nu_{s_1, \dots, s_k}$  as  $\mu_{t_1, \dots, t_k}$  for  $s_i = \lambda t_i$ . Then the  $\nu_{s_1, \dots, s_k}$  satisfy the conditions of the theorem with a new  $H$ , as well as (\*), so that there is a random element  $Y$  of  $\mathbf{D}$  with these finite-dimensional distributions. Finally, setting  $X_t = Y_{\delta_0 + t(1-2\delta_0)}$  we get a process  $X$  with the required finite dimensional distributions  $P\pi_{t_1, \dots, t_k}^{-1} = \mu_{t_1, \dots, t_k}$ .

**Example 7.18** Construction of a Levy process. Let  $\nu_t$  be a measure on the line for which  $\nu_t(\mathbf{R}) = H(t)$  is nondecreasing and continuous,  $t \in [0, 1]$ . Suppose for  $s \leq t$ ,  $\nu_s(A) \leq \nu_t(A)$  for all  $A \in \mathcal{R}$  so that  $\nu_t - \nu_s$  is a measure with total mass  $H(t) - H(s)$ . Then there is an infinitely divisible distribution having mean 0, variance  $H(t) - H(s)$ , and characteristic function

$$\phi_{s,t}(u) = \exp \int_{-\infty}^{\infty} \frac{e^{iuz} - 1 - iuz}{z^2} (\nu_t - \nu_s)(dz).$$

We can use Theorem 7.17 to construct a random element  $X$  of  $\mathbf{D}$  with  $X_0 = 0$ , for which the increments are independent and

$$\mathbb{E}(e^{iu_1 X_{t_1}} e^{iu_2 (X_{t_2} - X_{t_1})} \dots e^{iu_k (X_{t_k} - X_{t_{k-1}})}) = \phi_{0,t_1}(u_1) \phi_{t_1,t_2}(u_2) \dots \phi_{t_{k-1},t_k}(u_k).$$

Indeed, since  $\phi_{r,t}(u) = \phi_{r,s}(u) \phi_{s,t}(u)$  for  $r \leq s \leq t$ , the implied finite-dimensional distributions  $\mu_{t_1, \dots, t_k}$  are consistent. Further, by Chebyshev's inequality and independence, condition (ii) of Theorem 7.17 is valid with  $\alpha = \beta = 2$ :

$$\begin{aligned} \mu_{t_1, t_2, t_3} \{(z_1, z_3, z_3) : |z_2 - z_1| \geq \epsilon, |z_3 - z_2| \geq \epsilon\} &\leq \frac{H(t_2) - H(t_1)}{\epsilon^2} \cdot \frac{H(t_3) - H(t_2)}{\epsilon^2} \\ &\leq \epsilon^{-4} (H(t_3) - H(t_1))^2. \end{aligned}$$

Another application of Chebyshev's inequality gives

$$\mu_{t, t+\delta} \{(z_1, z_2) : |z_2 - z_1| \geq \epsilon\} \leq \frac{H(t+\delta) - H(t)}{\epsilon^2} \rightarrow 0, \quad \delta \downarrow 0.$$

## 8 Weak convergence on $\mathbf{D}$

Recall that the subset  $T_P \subset [0, 1]$ , introduced in Lemma 7.9, is the collection of  $t$  such that the projection  $\pi_t$  is  $P$ -almost surely continuous.

## 8.1 Criteria for weak convergence in $\mathbf{D}$

**Lemma 8.1** *Let  $P$  be a probability measure on  $(\mathbf{D}, \mathcal{D})$  and  $\epsilon > 0$ . By right continuity of the paths we have  $\lim_{\delta \rightarrow 0} P(x : |x(\delta) - x(0)| \geq \epsilon) = 0$ .*

Proof. Put  $A_\delta = \{x : |x(\delta) - x(0)| \geq \epsilon\}$ . Let  $\delta_n \rightarrow 0$ . It suffices to show that  $P(A_{\delta_n}) \rightarrow 0$ . To see this observe that right continuity of the paths entails

$$\bigcap_{n \geq 1} \bigcup_{k \geq n} A_{\delta_k} = \emptyset,$$

and therefore,  $P(A_{\delta_n}) \leq P(\bigcup_{k \geq n} A_{\delta_k}) \rightarrow 0$  as  $n \rightarrow \infty$ .

**Theorem 8.2** *Let  $P_n, P$  be probability measures on  $(\mathbf{D}, \mathcal{D})$ . Suppose  $P_n \pi_{t_1, \dots, t_k}^{-1} \Rightarrow P \pi_{t_1, \dots, t_k}^{-1}$  holds whenever  $t_1, \dots, t_k$  lie in  $T_P$ . If for every positive  $\epsilon$*

$$\begin{aligned} (i) \quad & \lim_{\delta \rightarrow 0} P(x : |x(1) - x(1 - \delta)| \geq \epsilon) = 0, \\ (ii) \quad & \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P_n(x : w_x''(\delta) \geq \epsilon) = 0, \end{aligned}$$

then  $P_n \Rightarrow P$ .

Proof. This result should be compared with Theorem 4.12 dealing with the space  $\mathbf{C}$ .

Recall Theorem 7.10. We prove tightness by checking conditions (i') in Lemma 7.13 and (ii) in Theorem 7.12. For each  $t \in T_P$ , the weakly convergent sequence  $P_n \pi_t^{-1}$  is tight which implies (i') with  $T_P$  in the role of  $T$ .

As to (ii) in Theorem 7.12 we have to verify only the second and third parts. By hypothesis,  $P_n \pi_{0, \delta}^{-1} \Rightarrow P \pi_{0, \delta}^{-1}$  so that for  $\delta \in T_P$ ,

$$\limsup_{n \rightarrow \infty} P_n(x : |x(\delta) - x(0)| \geq \epsilon) \leq P(x : |x(\delta) - x(0)| \geq \epsilon),$$

and the second part follows from Lemma 8.1.

Turning to the third part of (ii) in Theorem 7.12, the symmetric to the last argument brings for  $1 - \delta \in T_P$ ,

$$\limsup_{n \rightarrow \infty} P_n(x : |x(1) - x(1 - \delta)| \geq \epsilon) \leq P(x : |x(1) - x(1 - \delta)| \geq \epsilon).$$

Now, suppose that\*

$$|x(1-) - x(1 - \delta)| \geq \epsilon.$$

Since

$$|x(1-) - x(1 - \delta)| \wedge |x(1) - x(1-)| \leq w_x''(\delta),$$

we have either  $w_x''(\delta) \geq \epsilon$  or  $|x(1) - x(1-)| \leq w_x''(\delta)$ . Moreover, in the latter case, it is either  $|x(1) - x(1 - \delta)| \geq \epsilon/2$  or  $w_x''(\delta) \geq \epsilon/2$  or both. The last two observations yield

$$\begin{aligned} & \limsup_{n \rightarrow \infty} P_n(x : |x(1-) - x(1 - \delta)| \geq \epsilon) \\ & \leq P(x : |x(1) - x(1 - \delta)| \geq \epsilon/2) + \limsup_{n \rightarrow \infty} P_n(x : w_x''(\delta) \geq \epsilon/2), \end{aligned}$$

and the third part readily follows from conditions (i) and (ii).

**Theorem 8.3** *For  $X^n \Rightarrow X$  on  $\mathbf{D}$  it suffices that*

- (i)  $(X_{t_1}^n, \dots, X_{t_k}^n) \Rightarrow (X_{t_1}, \dots, X_{t_k})$  for points  $t_i \in T_P$ , where  $P$  is the probability distribution of  $X$ ,
- (ii)  $X_1 - X_{1-\delta} \Rightarrow 0$  as  $\delta \rightarrow 0$ ,
- (iii) there exist  $\alpha > 1$ ,  $\beta > 0$ , and a nondecreasing continuous function  $H : [0, 1] \rightarrow \mathbf{R}$  such that

$$\mathbb{E}(|X_s^n - X_r^n|^\beta |X_t^n - X_s^n|^\beta) \leq (H(t) - H(r))^\alpha \quad \text{for } 0 \leq r \leq s \leq t \leq 1.$$

Proof. By Theorem 8.2, it is enough to show that

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P}(w''(X^n, \delta) \geq \epsilon) = 0.$$

This follows from Lemma 7.16 as (iii) implies that  $X_n$  has a distribution  $P_n \in H_{\alpha, \beta}$ .

---

\*Here we use an argument suggested by Timo Hirscher.

## 8.2 Functional CLT on $\mathbf{D}$

The identity map  $c : \mathbf{C} \rightarrow \mathbf{D}$  is continuous and therefore measurable  $\mathcal{C}/\mathcal{D}$ . If  $\mathbb{W}$  is a Wiener measure on  $(\mathbf{C}, \mathcal{C})$ , then  $\mathbb{W}c^{-1}$  is a Wiener measure on  $(\mathbf{D}, \mathcal{D})$ . We denote this new measure by  $\mathbb{W}$  rather than  $\mathbb{W}c^{-1}$ . Clearly,  $\mathbb{W}(\mathbf{C}) = 1$ . Let also denote by  $W$  a random element of  $\mathbf{D}$  with distribution  $\mathbb{W}$ .

**Theorem 8.4** *Let  $\xi_1, \xi_2, \dots$  be iid r.v. defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ . If  $\xi_i$  have zero mean and variance  $\sigma^2$  and  $X_t^n = \frac{\xi_1 + \dots + \xi_{\lfloor nt \rfloor}}{\sigma\sqrt{n}}$ , then  $X^n \Rightarrow W$ .*

Proof. We apply Theorem 8.3. Following the proof of Theorem 4.17 (a) one gets the convergence of the fdd (i) even for the  $X^n$  as they are defined here. Condition (ii) follows from the fact that the Wiener process  $W_t$  has no jumps. We finish the proof by showing that (iii) holds with  $\alpha = \beta = 2$  and  $H(t) = 2t$ . Indeed,

$$\mathbb{E}(|X_s^n - X_r^n|^2 |X_t^n - X_s^n|^2) = 0 \quad \text{for } 0 \leq t - r < n^{-1},$$

as either  $X_s^n = X_r^n$  or  $X_t^n = X_s^n$ . On the other hand, for  $t - r \geq n^{-1}$ , by independence,

$$\begin{aligned} \mathbb{E}(|X_s^n - X_r^n|^2 |X_t^n - X_s^n|^2) &= \frac{\lfloor ns \rfloor - \lfloor nr \rfloor}{n} \cdot \frac{\lfloor nt \rfloor - \lfloor ns \rfloor}{n} \\ &\leq \left( \frac{\lfloor nt \rfloor - \lfloor nr \rfloor}{n} \right)^2 \leq (2(t - r))^2. \end{aligned}$$

**Example 8.5** Define  $(\xi_n)$  on  $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$  using the Rademacher functions  $\xi_n(\omega) = 2\omega_n - 1$  in terms of the dyadic (binary) representation  $\omega = \omega_1\omega_2\dots$ . Then  $(\xi_n)$  is a sequence of independent coin tossing outcomes with values  $\pm 1$ . Theorem 8.4 holds with  $\sigma = 1$ :  $\left( \frac{\xi_1 + \dots + \xi_{\lfloor nt \rfloor}}{\sqrt{n}} \right)_{t \in [0,1]} \Rightarrow W$ .

**Lemma 8.6** *Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and let  $\mathbb{P}_0$  be a probability measure absolutely continuous with respect to  $\mathbb{P}$ . Let  $\mathcal{F}_0 \subset \mathcal{F}$  be an algebra of events such that for some  $A_n \in \sigma(\mathcal{F}_0)$*

$$\mathbb{P}(A_n | E) \rightarrow \alpha, \quad \text{for all } E \in \mathcal{F}_0 \text{ with } \mathbb{P}(E) > 0.$$

*Then  $\mathbb{P}_0(A_n) \rightarrow \alpha$ .*

Proof. We have  $\mathbb{P}_0(A) = \int_A g_0(\omega) \mathbb{P}(d\omega)$ , where  $g_0 = d\mathbb{P}_0/d\mathbb{P}$ . It suffices to prove that

$$\int_{A_n} g(\omega) \mathbb{P}(d\omega) \rightarrow \alpha \int_{\Omega} g(\omega) \mathbb{P}(d\omega) \quad (*)$$

if  $g$  is  $\mathcal{F}$ -measurable and  $\mathbb{P}$ -integrable. We prove  $(*)$  in three steps.

Step 1. Write  $\mathcal{F}_1 = \sigma(\mathcal{F}_0)$  and denote by  $\mathcal{F}_2$  the class of events  $E$  for which

$$\mathbb{P}(A_n \cap E) \rightarrow \alpha \mathbb{P}(E).$$

We show that  $\mathcal{F}_1 \subset \mathcal{F}_2$ . To be able to apply Theorem 1.3 we have to show that  $\mathcal{F}_2$  is a  $\lambda$ -system. Indeed, suppose for a sequence of disjoint sets  $E_i$  we have

$$\mathbb{P}(A_n \cap E_i) \rightarrow \alpha \mathbb{P}(E_i).$$

Let  $E = \cup_i E_i$ , then by Lemma 2.10,

$$\mathbb{P}(A_n \cap E) = \sum_i \mathbb{P}(A_n \cap E_i) \rightarrow \alpha \sum_i \mathbb{P}(E_i) = \alpha \mathbb{P}(E).$$

Step 2. Show that  $(*)$  holds for  $\mathcal{F}_1$ -measurable functions  $g$ . Indeed, due to step 1, relation  $(*)$  holds if  $g$  is the indicator of an  $\mathcal{F}_1$ -set. and hence if it is a simple  $\mathcal{F}_1$ -measurable function. If  $g$  is  $\mathcal{F}_1$ -measurable and  $\mathbb{P}$ -integrable function, choose simple  $\mathcal{F}_1$ -measurable functions  $g_k$  that satisfy  $|g_k| \leq |g|$  and  $g_k \rightarrow g$ . Now

$$\left| \int_{A_n} g(\omega) \mathbb{P}(d\omega) - \alpha \int_{\Omega} g(\omega) \mathbb{P}(d\omega) \right| \leq \left| \int_{A_n} g_k(\omega) \mathbb{P}(d\omega) - \alpha \int_{\Omega} g_k(\omega) \mathbb{P}(d\omega) \right| + (1 + \alpha) \mathbb{E}|g - g_k|.$$

Let first  $n \rightarrow \infty$  and then  $k \rightarrow \infty$  and apply the dominated convergence theorem.

Step 3. Finally, take  $g$  to be a  $\mathcal{F}$ -measurable and  $\mathbb{P}$ -integrable. We use conditional expectation  $g_1 = \mathbb{E}(g | \mathcal{F}_1)$

$$\int_{A_n} g(\omega) \mathbb{P}(d\omega) = \mathbb{E}(g 1_{\{A_n\}}) = \mathbb{E}(g_1 1_{\{A_n\}}) \rightarrow \alpha \mathbb{E}(g_1) = \alpha \int_{\Omega} g(\omega) \mathbb{P}(d\omega).$$

**Theorem 8.7** Let  $\xi_1, \xi_2, \dots$  be iid r.v. defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  having zero mean and variance  $\sigma^2$ . Put  $X_t^n = \frac{\xi_1 + \dots + \xi_{\lfloor nt \rfloor}}{\sigma\sqrt{n}}$ . If  $\mathbb{P}_0$  is a probability measure absolutely continuous with respect to  $\mathbb{P}$ , then  $X^n \Rightarrow W$  with respect to  $\mathbb{P}_0$ .

Proof. Step 1. Choose  $k_n$  such that  $k_n \rightarrow \infty$  and  $k = o(n)$  as  $n \rightarrow \infty$  and put

$$\bar{X}_t^n = \frac{1}{\sigma\sqrt{n}} \sum_{i=k_n}^{\lfloor nt \rfloor} \xi_i,$$

$$Y_n = \frac{1}{\sigma\sqrt{n}} \max_{1 \leq k \leq k_n} |\xi_1 + \dots + \xi_k|.$$

By Kolmogorov's inequality, for any  $a > 0$ ,

$$\mathbb{P}(Y_n \geq a) \rightarrow 0, \quad n \rightarrow \infty,$$

and therefore

$$d(X^n, \bar{X}^n) \leq \|X^n - \bar{X}^n\| = Y_n \Rightarrow 0 \quad \text{with respect to } \mathbb{P}.$$

Applying Theorem 8.4 and Corollary 1.24 we conclude that  $\bar{X}^n \Rightarrow W$  with respect to  $\mathbb{P}$ .

Step 2: show using Lemma 8.6, that  $\bar{X}^n \Rightarrow W$  with respect to  $\mathbb{P}_0$ . If  $A \in \mathcal{D}$  is a  $\mathbb{W}$ -continuity set, then  $\mathbb{P}(A_n) \rightarrow \alpha$  for  $A_n = \{\bar{X}^n \in A\}$  and  $\alpha = \mathbb{W}(A)$ . Let  $\mathcal{F}_0$  be the algebra of the cylinder sets  $\{(\xi_1, \dots, \xi_k) \in H\}$ . If  $E \in \mathcal{F}_0$ , then  $A_n$  are independent of  $E$  for large  $n$  and by Lemma 8.6,  $\mathbb{P}_0(\bar{X}^n \in A) \rightarrow \mathbb{W}(A)$ .

Step 3. Since  $1_{\{Y_n \geq a\}} \rightarrow 0$  almost surely with respect to  $\mathbb{P}$ , the dominated convergence theorem gives

$$\mathbb{P}_0(Y_n \geq a) = \int g_0(\omega) 1_{\{Y_n \geq a\}} \mathbb{P}(d\omega) \rightarrow 0.$$

Arguing as in step 1 we conclude that  $d(X^n, \bar{X}^n) \Rightarrow 0$  with respect to  $\mathbb{P}_0$ .

Step 4. Applying once again Corollary 1.24 we conclude that  $X^n \Rightarrow W$  with respect to  $\mathbb{P}_0$ .

**Example 8.8** Define  $\xi_n$  on  $([0, 1], \mathcal{B}_{[0,1]}, \lambda_p)$  with

$$\lambda_p(du) = au^{a-1}du, \quad a = -\log_2(1-p), \quad p \in (0, 1)$$

again, as in Example 8.5, using the Rademacher functions. If  $p = \frac{1}{2}$ , then  $\lambda_p = \lambda$  and we are back to Example 8.5. With  $p \neq \frac{1}{2}$ , this corresponds to dependent  $p$ -coin tossings with

$$\int_0^{2^{-n}} \lambda_p(du) = (1-p)^n$$

being the probability of having  $n$  failures in the first  $n$  tossings, and

$$\int_{1-2^{-n}}^1 \lambda_p(du) = 1 - (1-p)^{-\log_2(1-2^{-n})}$$

being the probability of having  $n$  successes in the first  $n$  tossings. By Theorem 8.7, even in this case  $\left(\frac{\xi_1 + \dots + \xi_{\lfloor nt \rfloor}}{\sqrt{n}}\right)_{t \in [0,1]} \Rightarrow W$ .

### 8.3 Empirical distribution functions

**Definition 8.9** Let  $\xi_1(\omega), \dots, \xi_n(\omega)$  be iid with a distribution function  $F$  over  $[0, 1]$ . The corresponding empirical process is defined by  $Y_t^n = \sqrt{n}(F_n(t) - F(t))$ , where

$$F_n(t) = n^{-1}(1_{\{\xi_1 \leq t\}} + \dots + 1_{\{\xi_n \leq t\}})$$

is the empirical distribution function.

**Lemma 8.10** Let  $(Z_1^n, \dots, Z_r^n)$  have a multinomial distribution  $Mn(n, p_1, \dots, p_r)$ . Then the normalized vector  $\left(\frac{Z_1^n - np_1}{\sqrt{n}}, \dots, \frac{Z_r^n - np_r}{\sqrt{n}}\right)$  converges in distribution to a multivariate normal distribution with zero means and a covariance matrix

$$\mathbf{V} = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 & \dots & -p_1p_r \\ -p_2p_1 & p_2(1-p_2) & -p_2p_3 & \dots & -p_2p_r \\ -p_3p_1 & -p_3p_2 & p_3(1-p_3) & \dots & -p_3p_r \\ \dots & \dots & \dots & \dots & \dots \\ -p_rp_1 & -p_rp_2 & -p_rp_3 & \dots & p_r(1-p_r) \end{pmatrix}.$$

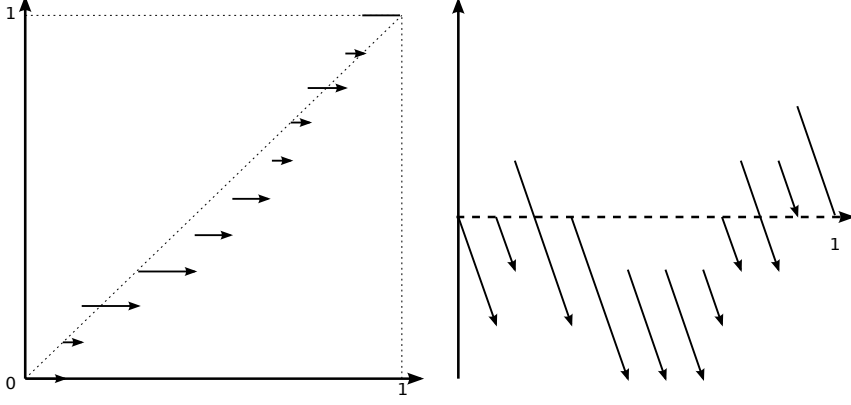


Figure 6: An example of the empirical distribution function with  $n = 10$  for the uniform distribution (left panel) and the corresponding empirical process (right panel).

Proof. To apply the continuity property of the multivariate characteristic functions consider

$$\mathbb{E} \exp \left( i\theta_1 \frac{Z_1^n - np_1}{\sqrt{n}} + \dots + i\theta_r \frac{Z_r^n - np_r}{\sqrt{n}} \right) = \left( \sum_{j=1}^r p_j e^{i\tilde{\theta}_j/\sqrt{n}} \right)^n,$$

where  $\tilde{\theta}_j = \theta_j - (\theta_1 p_1 + \dots + \theta_r p_r)$ . Similarly to the classical case we have

$$\left( \sum_{j=1}^r p_j e^{i\tilde{\theta}_j/\sqrt{n}} \right)^n = \left( 1 - \frac{1}{2n} \sum_{j=1}^r p_j \tilde{\theta}_j^2 + o(n^{-1}) \right)^n \rightarrow e^{-\frac{1}{2} \sum_{j=1}^r p_j \tilde{\theta}_j^2} = e^{-\frac{1}{2} (\sum_{j=1}^r p_j \theta_j^2 - (\sum_{j=1}^r p_j \theta_j)^2)}.$$

It remains to see that the right hand side equals  $e^{-\frac{1}{2} \boldsymbol{\theta} \mathbf{V} \boldsymbol{\theta}}$  which follows from the representation

$$\mathbf{V} = \begin{pmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_r \end{pmatrix} - \begin{pmatrix} p_1 \\ \vdots \\ p_r \end{pmatrix} \begin{pmatrix} p_1, \dots, p_r \end{pmatrix}.$$

**Theorem 8.11** *If  $\xi_1, \xi_2, \dots$  are iid  $[0, 1]$ -valued r.v. with a distribution function  $F$ , then the empirical process weakly converges  $Y^n \Rightarrow Y$  to a random element  $(Y_t)_{t \in [0, 1]} = (W_{F(t)}^\circ)_{t \in [0, 1]}$ , where  $W^\circ$  is the standard Brownian bridge. The limit  $Y$  is a Gaussian process specified by  $\mathbb{E}(Y_t) = 0$  and  $\mathbb{E}(Y_s Y_t) = F(s)(1 - F(t))$  for  $s \leq t$ .*

Proof. We start with the uniform case,  $F(t) \equiv t$  for  $t \in [0, 1]$ , by showing  $Y^n \Rightarrow W^\circ$ , where  $W^\circ$  is the Brownian bridge with  $\mathbb{E}(W_s^\circ W_t^\circ) = s(1 - t)$  for  $s \leq t$ . Let

$$U_t^n = nF_n(t) = 1_{\{\xi_1 \leq t\}} + \dots + 1_{\{\xi_n \leq t\}}$$

be the number of  $\xi_1, \dots, \xi_n$  falling inside  $[0, t]$ . Since the increments of  $U_t^n$  are described by multinomial joint distributions, by the previous lemma, the fdd of  $Y_t^n = \frac{U_t^n - nt}{\sqrt{n}}$  converge to those of  $W^\circ$ . Indeed, for  $t_1 < t_2 < \dots$  and  $i < j$ ,

$$\mathbb{E}(W_{t_i}^\circ - W_{t_{i-1}}^\circ)(W_{t_j}^\circ - W_{t_{j-1}}^\circ) = -(t_i - t_{i-1})(t_j - t_{j-1}) = -p_i p_j.$$

By Theorem 8.3 it suffices to prove for  $t_1 \leq t \leq t_2$  that

$$\mathbb{E} \left( (Y_t^n - Y_{t_1}^n)^2 (Y_{t_2}^n - Y_t^n)^2 \right) \leq (t - t_1)(t_2 - t) \leq (t_2 - t_1)^2.$$

In terms of  $\alpha_i = 1_{\{\xi_i \in (t_1, t]\}} + t_1 - t$  and  $\beta_i = 1_{\{\xi_i \in (t, t_2]\}} + t - t_2$  the first inequality is equivalent to

$$\mathbb{E} \left( \left( \sum_{i=1}^n \alpha_i \right)^2 \left( \sum_{i=1}^n \beta_i \right)^2 \right) \leq n^2 (t - t_1)(t_2 - t).$$

As we show next, this follows from  $\mathbb{E}(\alpha_i) = \mathbb{E}(\beta_i) = 0$ , independence  $(\alpha_i, \beta_i) \perp (\alpha_j, \beta_j)$  for  $i \neq j$ , and the following formulas for the second order moments. Let us write  $p_1 = t - t_1$ , and  $p_2 = t_2 - t$ . Since

$$\alpha_i = \begin{cases} 1 - p_1 & \text{w.p. } p_1, \\ -p_1 & \text{w.p. } 1 - p_1, \end{cases} \quad \beta_i = \begin{cases} 1 - p_2 & \text{w.p. } p_2, \\ -p_2 & \text{w.p. } 1 - p_2, \end{cases}$$



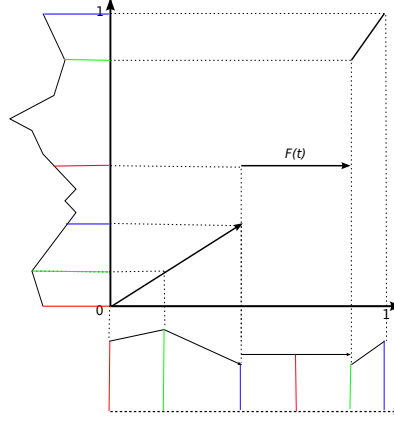


Figure 7: The time transformation  $\psi : \mathbf{D} \rightarrow \mathbf{D}$ . Along the  $y$ -axis an original path  $x(t)$  is depicted, along the  $x$ -axis the time transformed path  $(\psi x)(t) = x(F(t))$  is given. The jumps of  $F$  translate into path jumps, the constant parts of  $F$  translate into the horizontal pieces of the transformed path.

and

$$\alpha_i \beta_i = \begin{cases} -(1-p_1)p_2 & \text{w.p. } p_1, \\ -p_1(1-p_2) & \text{w.p. } p_2, \\ p_1 p_2 & \text{w.p. } 1-p_1-p_2, \end{cases}$$

we have

$$\begin{aligned} \mathbb{E}(\alpha_i^2) &= p_1(1-p_1), \quad \mathbb{E}(\beta_i^2) = p_2(1-p_2), \quad \mathbb{E}(\alpha_i \beta_i) = -p_1 p_2, \\ \mathbb{E}(\alpha_i^2 \beta_i^2) &= p_1(1-p_1)^2 p_2^2 + p_2 p_1^2 (1-p_2)^2 + (1-p_1-p_2) p_1^2 p_2^2 = p_1 p_2 (p_1 + p_2 - 3p_1 p_2) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\left(\left(\sum_{i=1}^n \alpha_i\right)^2 \left(\sum_{i=1}^n \beta_i\right)^2\right) &= n\mathbb{E}(\alpha_i^2 \beta_i^2) + n(n-1)\mathbb{E}(\alpha_i^2)\mathbb{E}(\beta_i^2) + 2n(n-1)(\mathbb{E}(\alpha_i \beta_i))^2 \\ &\leq n^2 p_1 p_2 (p_1 + p_2 - 3p_1 p_2 + 1 - p_1 - p_2 + 3p_1 p_2) \\ &= n^2 p_1 p_2 = n^2(t-t_1)(t_2-t). \end{aligned}$$

This proves the theorem for the uniform case. For a general continuous and strictly increasing  $F(t)$  we use the transformation  $\eta_i = F(\xi_i)$  into uniformly distributed r.v. If  $G_n(t) = G_n(t, \omega)$  is the empirical distribution function of  $(\eta_1(\omega), \dots, \eta_n(\omega))$  and  $Z_t^n = \sqrt{n}(G_n(t) - t)$ , then  $Z^n \Rightarrow W^\circ$ .

Observe that

$$G_n(F(t)) = \frac{1_{\{\eta_1 \leq F(t)\}} + \dots + 1_{\{\eta_n \leq F(t)\}}}{n} = \frac{1_{\{F(\xi_1) \leq F(t)\}} + \dots + 1_{\{F(\xi_n) \leq F(t)\}}}{n} = F_n(t).$$

Define  $\psi : \mathbf{D} \rightarrow \mathbf{D}$  by  $(\psi x)(t) = x(F(t))$ . If  $x_n \rightarrow x$  in the Skorokhod topology and  $x \in \mathbf{C}$ , then the convergence is uniform, so that  $\psi x_n \rightarrow \psi x$  uniformly and hence in the Skorokhod topology. By the mapping theorem  $\psi(Z^n) \Rightarrow \psi(W^\circ)$ . Therefore,

$$Y^n = (Y_t^n)_{t \in [0,1]} = (Z_{F(t)}^n)_{t \in [0,1]} = \psi(Z^n) \Rightarrow \psi(W^\circ) = (W_{F(t)}^\circ)_{t \in [0,1]} = Y.$$

Finally, for  $F(t)$  with jumps and constant parts (see Figure 7) the previous argument works provided there exists an iid sequence  $\eta_1, \eta_2, \dots$  of uniformly distributed r.v. as well as iid  $\xi'_1, \xi'_2, \dots$  with distribution function  $F$ , such that

$$\{\eta_i \leq F(t)\} \equiv \{\xi'_i \leq t\}, \quad t \in [0,1], i \geq 1.$$

This is achieved by starting with uniform  $\eta_1, \eta_2, \dots$  on possibly another probability space and putting  $\xi'_i = \phi(\eta_i)$ , where  $\phi(u) = \inf\{t : u \leq F(t)\}$  is the quantile function satisfying

$$\{\phi(u) \leq t\} = \{u \leq F(t)\}.$$

**Example 8.12** Kolmogorov-Smirnov test. Let  $F$  be continuous. By the mapping theorem we obtain

$$\sqrt{n} \sup_t |F_n(t) - F(t)| = \sup_t |Y_t^n| \Rightarrow \sup_t |W_{F(t)}^\circ| = \sup_t |W_t^\circ|,$$

where the limit distribution is given by Theorem 5.11.

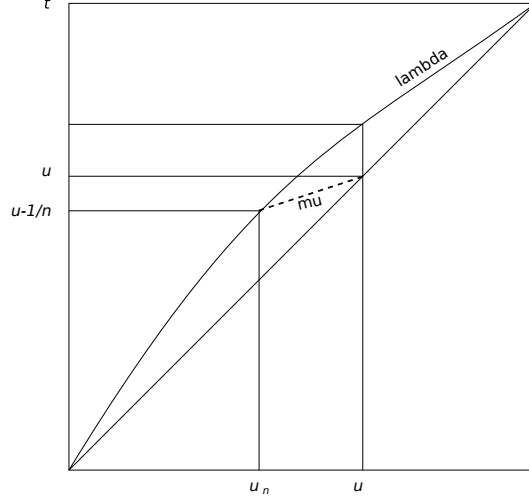


Figure 8: A detail of the proof of Lemma 9.2 and Theorem 9.8.

## 9 The space $\mathbf{D}_\infty = \mathbf{D}[0, \infty)$

### 9.1 Two metrics on $\mathbf{D}_\infty$

To extend the Skorokhod theory to the space  $\mathbf{D}_\infty = \mathbf{D}[0, \infty)$  of the cadlag functions on  $[0, \infty)$ , consider for each  $t > 0$  the space  $\mathbf{D}_t = \mathbf{D}[0, t]$  of the same cadlag functions restricted on  $[0, t]$ . All definitions for  $\mathbf{D} = \mathbf{D}[0, 1]$  have obvious analogues for  $\mathbf{D}_t$ : for example we denote by  $d_t(x, y)$  the analogue of  $d_1(x, y) := d(x, y)$ . Denote  $\|x\|_t = \sup_{u \in [0, t]} |x(u)|$ .

**Example 9.1** One might try to define Skorokhod convergence  $x_n \rightarrow x$  on  $\mathbf{D}_\infty$  by requiring that  $d_t(x_n, x) \rightarrow 0$  for each finite  $t > 0$ . This does not work: if  $x_n(t) = 1_{\{t \in [0, 1 - n^{-1}]\}}$ , the natural limit would be  $x(t) = 1_{\{t \in [0, 1]\}}$  but  $d_1(x_n, x) = 1$  for all  $n$ . The problem here is that  $x$  is discontinuous at  $t = 1$ , and the definition must accommodate discontinuities.

**Lemma 9.2** *Let  $0 < u < t < \infty$ . If  $d_t(x_n, x) \rightarrow 0$  and  $x$  is continuous at  $u$ , then  $d_u(x_n, x) \rightarrow 0$ .*

*Proof.* By hypothesis, there are time transforms  $\lambda_n \in \Lambda_t$  such that  $\|\lambda_n - 1\|_t \rightarrow 0$  and  $\|x_n - x\lambda_n\|_t \rightarrow 0$  as  $n \rightarrow \infty$ . Given  $\epsilon$ , choose  $\delta$  so that  $|v - u| \leq 2\delta$  implies  $|x(v) - x(u)| \leq \epsilon/2$ . Now choose  $n_0$  so that, if  $n \geq n_0$  and  $v \leq t$ , then  $|\lambda_n v - v| \leq \delta$  and  $|x_n(v) - (x\lambda_n)(v)| \leq \epsilon/2$ . Then, if  $n \geq n_0$  and  $|v - u| \leq \delta$ , we have

$$|\lambda_n v - u| \leq |\lambda_n v - v| + |v - u| \leq 2\delta$$

and hence

$$|x_n(v) - x(u)| \leq |x_n(v) - (x\lambda_n)(v)| + |(x\lambda_n)(v) - x(u)| \leq \epsilon.$$

Thus

$$\sup_{|v-u| \leq \delta} |x(v) - x(u)| \leq \epsilon, \quad \sup_{|v-u| \leq \delta} |x_n(v) - x(u)| \leq \epsilon \quad \text{for } n \geq n_0.$$

Let

$$u_n = \begin{cases} u - n^{-1} & \text{if } \lambda_n u > u, \\ u & \text{if } \lambda_n u = u, \\ \lambda_n^{-1}(u - n^{-1}) & \text{if } \lambda_n u < u, \end{cases}$$

so that  $u_n \leq u$ . Since

$$|u_n - u| \leq |\lambda_n^{-1}(u - n^{-1}) - (u - n^{-1})| + n^{-1},$$

we have  $u_n \rightarrow u$ , and since

$$|\lambda_n u_n - u| \leq |\lambda_n u_n - u_n| + |u_n - u|,$$

we also have  $\lambda_n u_n \rightarrow u$ .

Define  $\mu_n \in \Lambda_u$  so that  $\mu_n v = \lambda_n v$  for  $v \in [0, u_n]$  and interpolate linearly on  $(u_n, u]$  aiming at the diagonal point  $\mu_n u = u$ , see Figure 8. By linearity,  $|\mu_n v - v| \leq |\lambda_n u_n - u_n|$  for  $v \in [u_n, u]$  and we have  $\|\mu_n - 1\|_u \rightarrow 0$ .

It remains to show that  $\|x_n - x\mu_n\|_u \rightarrow 0$ . To do this we choose  $n_1$  so that

$$u_n > u - \delta \quad \text{and} \quad \lambda_n u_n > u - \delta \quad \text{for } n \geq n_1.$$

If  $v \leq u_n$ , then

$$|x_n(v) - (x\mu_n)(v)| = |x_n(v) - (x\lambda_n)(v)| \leq \|x_n - x\lambda_n\|_t.$$

On the other hand, if  $v \in [u_n, u]$  and  $n \geq n_1$ , then  $v \in [u - \delta, u]$  and  $\mu_n v \in [u - \delta, u]$  implying for  $n \geq n_1 \vee n_0$

$$|x_n(v) - (x\mu_n)(v)| \leq |x_n(v) - x(u)| + |x(u) - (x\mu_n)(v)| \leq 2\epsilon.$$

The proof is finished.

**Definition 9.3** For any natural  $i$ , define a map  $\psi_i : \mathbf{D}_\infty \rightarrow \mathbf{D}_i$  by

$$(\psi_i x)(t) = x(t)1_{\{t \leq i-1\}} + (i-t)x(t)1_{\{i-1 < t \leq i\}}$$

making the transformed function  $(\psi_i x)(t)$  continuous at  $t = i$ .

**Definition 9.4** Two topologically equivalent metrics  $d_\infty(x, y)$  and  $d_\infty^\circ(x, y)$  are defined on  $\mathbf{D}_\infty$  in terms of  $d(x, y)$  and  $d^\circ(x, y)$  by

$$d_\infty(x, y) = \sum_{i=1}^{\infty} \frac{1 \wedge d_i(\psi_i x, \psi_i y)}{2^i}, \quad d_\infty^\circ(x, y) = \sum_{i=1}^{\infty} \frac{1 \wedge d_i^\circ(\psi_i x, \psi_i y)}{2^i}.$$

The metric properties of  $d_\infty(x, y)$  and  $d_\infty^\circ(x, y)$  follow from those of  $d_i(x, y)$  and  $d_i^\circ(x, y)$ . In particular, if  $d_\infty(x, y) = 0$ , then  $d_i(\psi_i x, \psi_i y) = 0$  and  $\psi_i x = \psi_i y$  for all  $i$ , and this implies  $x = y$ .

**Lemma 9.5** The map  $\psi_i : \mathbf{D}_\infty \rightarrow \mathbf{D}_i$  is continuous.

Proof. It follows from the fact that  $d_\infty(x_n, x) \rightarrow 0$  implies  $d_i(\psi_i x_n, \psi_i x) \rightarrow 0$ .

## 9.2 Characterization of Skorokhod convergence on $\mathbf{D}_\infty$

Let  $\Lambda_\infty$  be the set of continuous, strictly increasing maps  $\lambda : [0, \infty) \rightarrow [0, \infty)$  such that  $\lambda 0 = 0$  and  $\lambda t \rightarrow \infty$  as  $t \rightarrow \infty$ . Denote  $\|x\|_\infty = \sup_{u \in [0, \infty)} |x(u)|$ .

**Exercise 9.6** Let  $\lambda \in \Lambda_t$  where  $t \in (0, \infty]$ . Show that the inverse transformation  $\lambda^{-1} \in \Lambda_t$  is such that  $\|\lambda^{-1} - 1\|_t = \|\lambda - 1\|_t$ .

**Example 9.7** Consider the sequence  $x_n(t) = 1_{\{t \geq n\}}$  of elements of  $\mathbf{D}_\infty$ . Its natural limit is  $x \equiv 0$  as  $\|x_n - x\|_t = 0$  for all  $n > t > 0$ . However,  $\|x_n \lambda_n - x\|_\infty = \|x_n \lambda_n\|_\infty = 1$  for any choice of  $\lambda_n \in \Lambda_\infty$ .

**Theorem 9.8** Convergence  $d_\infty(x_n, x) \rightarrow 0$  takes place if and only if there is a sequence  $\lambda_n \in \Lambda_\infty$  such that

$$\|\lambda_n - 1\|_\infty \rightarrow 0 \quad \text{and} \quad \|x_n \lambda_n - x\|_i \rightarrow 0 \quad \text{for each } i.$$

Proof. Necessity. Suppose  $d_\infty(x_n, x) \rightarrow 0$ . Then  $d_i(\psi_i x_n, \psi_i x) \rightarrow 0$  and there exist  $\lambda_n^{(i)} \in \Lambda_i$  such that

$$\epsilon_n^{(i)} = \|\lambda_n^{(i)} - 1\|_i \vee \|(\psi_i x_n) \lambda_n^{(i)} - \psi_i x\|_i \rightarrow 0, \quad n \rightarrow \infty \quad \text{for each } i.$$

Choose  $n_i > 1$  such that  $n \geq n_i$  implies  $\epsilon_n^{(i)} < i^{-1}$ . Arrange that  $n_i < n_{i+1}$ , and let

$$i_1 = \dots = i_{n_1-1} = 1, \quad i_{n_1} = \dots = i_{n_2-1} = 2, \quad i_{n_2} = \dots = i_{n_3-1} = 3, \quad \dots$$

so that  $i_n \rightarrow \infty$ . Define  $\lambda_n \in \Lambda_\infty$  by

$$\lambda_n t = \begin{cases} \lambda_n^{(i_n)} t & \text{if } t \leq i_n, \\ t & \text{if } t > i_n. \end{cases}$$

Then

$$\|\lambda_n - 1\|_\infty = \|\lambda_n^{(i_n)} - 1\|_{i_n} \leq \epsilon_n^{(i_n)} < i_n^{-1} \rightarrow 0.$$

Now fix  $i$ . If  $n$  is large enough, then  $i < i_n - 1$  and

$$\begin{aligned} \|x_n \lambda_n - x\|_i &= \|(\psi_{i_n} x_n) \lambda_n - \psi_{i_n} x\|_i \\ &\leq \|(\psi_{i_n} x_n) \lambda_n - \psi_{i_n} x\|_{i_n} = \|(\psi_{i_n} x_n) \lambda_n^{(i_n)} - \psi_{i_n} x\|_{i_n} \leq \epsilon_n^{(i_n)} < i_n^{-1} \rightarrow 0. \end{aligned}$$

Sufficiency. Suppose that there is a sequence  $\lambda_n \in \Lambda_\infty$  such that, firstly,  $\|\lambda_n - 1\|_\infty \rightarrow 0$ , and secondly,  $\|x_n \lambda_n - x\|_i \rightarrow 0$  for each  $i$ . Observe that for some  $C_i$ ,

$$\|x\|_i \leq C_i \text{ and } \|x_n\|_i \leq C_i \text{ for all } (n, i).$$

Indeed, by the first assumption, for large  $n$  we have  $\lambda_n(2i) > i$  implying  $\|x_n\|_i \leq \|x_n \lambda_n\|_{2i}$ , where by the second assumption,  $\|x_n \lambda_n\|_{2i} \rightarrow \|x\|_{2i}$ .

Fix an  $i$ . It is enough to show that  $d_i(\psi_i x_n, \psi_i x) \rightarrow 0$ . As in the proof of Lemma 9.2 define

$$u_n = \begin{cases} i - n^{-1} & \text{if } \lambda_n i < i, \\ i & \text{if } \lambda_n i = i, \\ \lambda_n^{-1}(i - n^{-1}) & \text{if } \lambda_n i > i, \end{cases}$$

and  $\mu_n \in \Lambda_i$  so that  $\mu_n v = \lambda_n v$  for  $v \in [0, u_n]$  interpolating linearly on  $(u_n, i]$  towards  $\mu_n i = i$ . As before,  $\|\mu_n - 1\|_i \rightarrow 0$  and it suffices to check that

$$\|(\psi_i x_n) \mu_n - \psi_i x\|_i \rightarrow 0, \quad n \rightarrow \infty.$$

To see that the last relation holds suppose  $j := \lambda_n^{-1}(i - 1) \leq i - 1$  (the other case  $j > i - 1$  is treated similarly) and observe that

$$\begin{aligned} \|(\psi_i x_n) \lambda_n - \psi_i x\|_i &= \left( \|x_n \lambda_n - x\|_j \right) \vee \left( \sup_{j < t \leq i-1} |(i - \lambda_n t) x_n(\lambda_n t) - x(t)| \right) \\ &\vee \left( \sup_{i-1 < t \leq i} |(i - \lambda_n t) x_n(\lambda_n t) - (i - t) x(t)| \right) \\ &\leq \left( \|x_n \lambda_n - x\|_i + C_i \sup_{j < t \leq i-1} |(i - 1 - \lambda_n t)| \right) \\ &\vee \left( \sup_{i-1 < t \leq i} |i - \lambda_n t| \cdot |x_n(\lambda_n t) - x(t)| + \sup_{i-1 < t \leq i} |(\lambda_n t - t) x(t)| \right) \\ &\leq \|x_n \lambda_n - x\|_i + C_i \|\lambda_n - 1\|_i \rightarrow 0. \end{aligned}$$

It follows that for  $t \leq u_n$ ,

$$|(\psi_i x_n)(\mu_n t) - (\psi_i x)(t)| \leq \|(\psi_i x_n) \lambda_n - \psi_i x\|_i \rightarrow 0.$$

Turning to the case  $u_n < t \leq i$ , given an  $\epsilon \in (0, 1)$  choose  $n_0$  such that for  $n > n_0$ ,  $u_n$  and  $\mu_n u_n$  both lie in  $[i - \epsilon, i]$ . Then

$$|(\psi_i x_n)(\mu_n t) - (\psi_i x)(t)| \leq \sup_{u_n < t \leq i} |(i - \mu_n t) x_n(\mu_n t) - (i - t) x(t)| \leq 2C_i \epsilon.$$

**Theorem 9.9** *Convergence  $d_\infty(x_n, x) \rightarrow 0$  takes place if and only if  $d_t(x_n, x) \rightarrow 0$  for each continuity point  $t$  of  $x$ .*

Proof. Necessity. If  $d_\infty(x_n, x) \rightarrow 0$ , then  $d_i(\psi_i x_n, \psi_i x) \rightarrow 0$  for each  $i$ . Given a continuity point  $t$  of  $x$ , take an integer  $i$  for which  $t < i - 1$ . According to Lemma 9.2,  $d_t(x_n, x) = d_t(\psi_i x_n, \psi_i x) \rightarrow 0$ .

Sufficiency. Choose continuity points  $t_i$  of  $x$  in such a way that  $t_i \uparrow \infty$  as  $i \rightarrow \infty$ . By hypothesis,

$$d_{t_i}(x_n, x) \rightarrow 0, \quad n \rightarrow \infty, \quad i \geq 1.$$

Choose  $\lambda_n^{(i)} \in \Lambda_{t_i}$  so that

$$\epsilon_n^{(i)} = \|\lambda_n^{(i)} - 1\|_{t_i} \vee \|x_n \lambda_n^{(i)} - x\|_{t_i} \rightarrow 0, \quad n \rightarrow \infty \quad \text{for each } i.$$

Using the argument from the first part of the proof of Theorem 9.8, define integers  $i_n$  in such a way that  $i_n \rightarrow \infty$  and  $\epsilon_n^{(i_n)} < i_n^{-1}$ . Put

$$\lambda_n t = \begin{cases} \lambda_n^{(i_n)} t & \text{if } t \leq t_{i_n}, \\ t & \text{if } t > t_{i_n}, \end{cases}$$

so that  $\lambda_n \in \Lambda_\infty$ . We have  $\|\lambda_n - 1\|_\infty \leq i_n^{-1}$ , and for any given  $i$ , if  $n$  is sufficiently large so that  $i < t_{i_n}$ , then

$$\|x_n \lambda_n - x\|_i = \|x_n \lambda_n^{(i_n)} - x\|_i \leq \|x_n \lambda_n^{(i_n)} - x\|_{t_{i_n}} \leq \epsilon_n^{(i_n)} < i_n^{-1} \rightarrow 0.$$

Applying Theorem 9.8 we get  $d_\infty(x_n, x) \rightarrow 0$ .

**Exercise 9.10** Show that the mapping  $h(x) = \sup_{t \geq 0} x(t)$  is not continuous on  $\mathbf{D}_\infty$ .

### 9.3 Separability and completeness of $D_\infty$

**Lemma 9.11** Suppose  $(S_i, \rho_i)$  are metric spaces and consider  $S = S_1 \times S_2 \times \dots$  together with the metric of coordinate-wise convergence

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

If each  $S_i$  is separable, then  $S$  is separable. If each  $S_i$  is complete, then  $S$  is complete.

Proof. Separability. For each  $i$ , let  $B_i$  be a countable dense subset in  $S_i$  and  $x_i^\circ \in S_i$  be a fixed point. We will show that the countable set  $B = B(x_1^\circ, x_2^\circ, \dots)$  defined by

$$B = \{x \in S : x = (x_1, \dots, x_k, x_{k+1}^\circ, x_{k+2}^\circ, \dots), x_1 \in B_1, \dots, x_k \in B_k, k \in \mathbf{N}\}$$

is dense in  $S$ . Given an  $\epsilon$  and a point  $y \in S$ , choose  $k$  so that  $\sum_{i>k} 2^{-i} < \epsilon$  and then choose points  $x_i \in B_i$  so that  $\rho_i(x_i, y_i) < \epsilon$ . With this choice the corresponding point  $x \in B$  satisfies  $\rho(x, y) < 2\epsilon$ .

Completeness. Suppose that  $x^n = (x_1^n, x_2^n, \dots)$  are points of  $S$  forming a fundamental sequence. Then each sequence  $(x_i^n)$  is fundamental in  $S_i$  and hence  $\rho_i(x_i^n, x_i) \rightarrow 0$  for some  $x_i \in S_i$ . By the M-test, Lemma 2.10,  $\rho(x^n, x) \rightarrow 0$ , where  $x = (x_1, x_2, \dots)$ .

**Definition 9.12** Consider the product space  $\overline{D} = D_1 \times D_2 \times \dots$  with the coordinate-wise convergence metric (cf Definition 9.4)

$$\rho(\bar{x}, \bar{y}) = \sum_{i=1}^{\infty} \frac{1 \wedge d_i^\circ(\bar{x}_i, \bar{y}_i)}{2^i}, \quad \bar{x} = (\bar{x}_1, \bar{x}_2, \dots), \bar{y} = (\bar{y}_1, \bar{y}_2, \dots) \in \overline{D}.$$

Put  $\psi x = (\psi_1 x, \psi_2 x, \dots)$  for  $x \in D_\infty$ . Then  $\psi x \in \overline{D}$  and  $d_\infty^\circ(x, y) = \rho(\psi x, \psi y)$  so that  $\psi$  is an isometry of  $(D_\infty, d_\infty^\circ)$  into  $(\overline{D}, \rho)$ .

**Lemma 9.13** The image  $\overline{D}_\infty := \psi D_\infty$  is closed in  $\overline{D}$ .

Proof. Suppose that  $x_n \in D_\infty$ ,  $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots) \in \overline{D}$ , and  $\rho(\psi x_n, \bar{x}) \rightarrow 0$ , then  $d_i(\psi_i x_n, \bar{x}_i) \rightarrow 0$  for each  $i$ . We must find an  $x \in D_\infty$  such that  $\bar{x} = \psi x$ .

The sequence of functions  $\bar{x}_i \in D_i$ ,  $i = 1, 2, \dots$  has at most countably many points of discontinuity. Therefore, there is a dense set  $T \in [0, \infty)$  such that for every  $i \geq t \in T$ , the function  $\bar{x}_i(\cdot)$  is continuous at  $t$ . Since  $d_i(\psi_i x_n, \bar{x}_i) \rightarrow 0$ , we have  $\psi_i x_n(t) \rightarrow \bar{x}_i(t)$  for all  $t \in T \cap [0, i]$ . This means that for every  $t \in T$  there exists the limit  $x(t) = \lim_n x_n(t)$ , since  $\psi_i x_n(t) = x_n(t)$  for  $i > t + 1$ .

Now  $\psi_i x(t) = \bar{x}_i(t)$  on  $T \cap [0, i]$ . Hence  $x(t) = \bar{x}_i(t)$  on  $T \cap [0, i - 1]$ , so that  $x$  can be extended to a cadlag function on each  $[0, i - 1]$  and then to a cadlag function on  $[0, \infty)$ . We conclude, using right continuity, that  $\psi_i x(t) = \bar{x}_i(t)$  for all  $t \in [0, i]$ .

**Theorem 9.14** The metric space  $(D_\infty, d_\infty^\circ)$  is separable and complete.

Proof. According Lemma 9.11 the space  $\overline{D}$  is separable and complete, so are the closed subspace  $\overline{D}_\infty$  and its isometric copy  $D_\infty$ .

### 9.4 Weak convergence on $D_\infty$

**Definition 9.15** For any natural  $i$  and any  $s \geq i$ , define a map  $\psi_{s,i} : D_s \rightarrow D_i$  by

$$(\psi_{s,i} x)(t) = x(t)1_{\{t \leq i-1\}} + (i-t)x(t)1_{\{i-1 < t \leq i\}}.$$

**Exercise 9.16** Show that the mapping  $\psi_{s,i}$  is continuous.

**Lemma 9.17** A necessary and sufficient condition for  $P_n \Rightarrow P$  on  $D_\infty$  is that  $P_n \psi_k^{-1} \Rightarrow P \psi_k^{-1}$  on  $D_k$  for every  $k \in \mathbf{N}$ .

Proof. Since  $\psi_k$  is continuous, the necessity follows from the mapping theorem.

For the sufficiency we need the isometry  $\psi$  from Definition 9.12 and the inverse isometry  $\psi^{-1}$ :

$$D_\infty \xrightarrow{\psi_k} D_k, \quad D_\infty \xrightarrow{\psi} \overline{D}, \quad \overline{D}_\infty \xrightarrow{\psi^{-1}} D_\infty.$$

Define two more mappings

$$\overline{D} \xrightarrow{\zeta_k} D_1 \times \dots \times D_k, \quad D_k \xrightarrow{\chi_k} D_1 \times \dots \times D_k$$

by

$$\zeta_k(\bar{x}) = (\bar{x}_1, \dots, \bar{x}_k), \quad \chi_k(x) = (\psi_{k,1}x, \dots, \psi_{k,k}x).$$

Consider the Borel  $\sigma$ -algebra  $\bar{\mathcal{D}}$  for  $(\bar{\mathcal{D}}, \rho)$  and let  $\bar{\mathcal{D}}_f \subset \bar{\mathcal{D}}$  be the class of sets of the form  $\zeta_k^{-1}H$  where  $k \geq 1$  and  $H \in \mathcal{D}_1 \times \dots \times \mathcal{D}_k$ , see Definition 2.5. The remainder of the proof is split into four steps.

Step 1. Applying Theorem 2.4 show that  $\mathcal{A} = \bar{\mathcal{D}}_f$  is a convergence-determining class. Given a ball  $B(\bar{x}, \epsilon) \subset \bar{\mathcal{D}}$ , take  $k$  so that  $2^{-k} < \epsilon/2$  and consider the cylinder sets

$$A_\eta = \{\bar{y} \in \bar{\mathcal{D}} : d_i^\circ(\bar{x}_i, \bar{y}_i) < \eta, i = 1, \dots, k\} \quad \text{for } 0 < \eta < \epsilon/2.$$

Then  $\bar{x} \in A_\eta^\circ = A_\eta \subset B(\bar{x}, \epsilon)$  implies  $A_\eta \in \mathcal{A}_{x, \epsilon}$ . It remains to see that the boundaries of  $A_\eta$  for different  $\eta$  are disjoint.

Step 2. For probability measures  $Q_n$  and  $Q$  on  $\bar{\mathcal{D}}$  show that if  $Q_n \zeta_k^{-1} \Rightarrow Q \zeta_k^{-1}$  for every  $k$ , then  $Q_n \Rightarrow Q$ .

This follows from the equality  $\partial(\zeta_k^{-1}H) = \zeta_k^{-1}\partial H$  for  $H \in \mathcal{D}_1 \times \dots \times \mathcal{D}_k$ , see the proof of Theorem 2.14.

Step 3. Assume that  $P_n \psi_k^{-1} \Rightarrow P \psi_k^{-1}$  on  $\mathcal{D}_k$  for every  $k$  and show that  $P_n \psi^{-1} \Rightarrow P \psi^{-1}$  on  $\bar{\mathcal{D}}$ .

The map  $\chi_k$  is continuous: if  $x_n \rightarrow x$  in  $\mathcal{D}_k$ , then  $\psi_{k,i}x_n \rightarrow \psi_{k,i}x$  in  $\mathcal{D}_i$ ,  $i \leq k$ . By the mapping theorem,  $P_n \psi_k^{-1} \chi_k^{-1} \Rightarrow P \psi_k^{-1} \chi_k^{-1}$ , and since  $\chi_k \psi_k = \zeta_k \psi$ , we get  $P_n \psi^{-1} \zeta_k^{-1} \Rightarrow P \psi^{-1} \zeta_k^{-1}$ . Referring to step 2 we conclude  $P_n \psi^{-1} \Rightarrow P \psi^{-1}$ .

Step 4. Show that  $P_n \psi^{-1} \Rightarrow P \psi^{-1}$  on  $\bar{\mathcal{D}}$  implies  $P_n \Rightarrow P$  on  $\mathcal{D}_\infty$ .

Extend the isometry  $\psi^{-1}$  to a map  $\eta : \bar{\mathcal{D}} \rightarrow \mathcal{D}_\infty$  by putting  $\eta(\bar{x}) = x_0 \in \mathcal{D}_\infty$  for all  $\bar{x} \notin \bar{\mathcal{D}}_\infty$ . Then  $\eta$  is continuous when restricted to  $\bar{\mathcal{D}}_\infty$ , and since  $\bar{\mathcal{D}}_\infty$  supports  $P \psi^{-1}$  and the  $P_n \psi^{-1}$ , it follows that

$$P_n = P_n \psi^{-1} \eta^{-1} \Rightarrow P \psi^{-1} \eta^{-1} = P.$$

**Definition 9.18** For a probability measure  $P$  on  $\mathcal{D}_\infty$  define  $T_P \subset [0, \infty)$  as the set of  $t$  for which  $P(J_t) = 0$ , where  $J_t = \{x : x \text{ is discontinuous at } t\}$ . (See Lemma 7.9.)

**Exercise 9.19** Let  $P$  be the probability measure on  $\mathcal{D}_\infty$  generated by the Poisson process with parameter  $\lambda$ . Show that  $T_P = [0, \infty)$ .

**Lemma 9.20** For  $x \in \mathcal{D}_\infty$  let  $r_t x$  be the restriction of  $x$  on  $[0, t]$ . The function  $r_t : \mathcal{D}_\infty \rightarrow \mathcal{D}_t$  is measurable. The set of points at which  $r_t$  is discontinuous belongs to  $J_t$ .

Proof. Denote  $\delta_k = t/k$ . Define the function  $r_t^k x \in \mathcal{D}_t$  as having the value  $x(i\delta_k)$  on  $[i\delta_k, (i+1)\delta_k)$  for  $0 \leq i \leq k-1$  and the value  $x(t)$  at  $t$ . Since the  $\pi_{i\delta_k}$  are measurable  $\mathcal{D}_\infty/\mathcal{R}_1$ , it follows as in the proof of Theorem 7.2 (b) that  $r_t^k$  is measurable  $\mathcal{D}_\infty/\mathcal{D}_t$ . By Lemma 6.31,

$$d_t(r_t^k x, r_t x) \leq \delta_k \vee w_t'(x, \delta_k) \rightarrow 0 \text{ as } k \rightarrow \infty \text{ for each } x \in \mathcal{D}_\infty.$$

Now, to show that  $r_t$  is measurable take a closed  $F \in \mathcal{D}_t$ . We have  $F = \bigcap_\epsilon F^{2\epsilon}$ , where the intersection is over positive rational  $\epsilon$ . From

$$r_t^{-1}F \subset \liminf_k (r_t^k)^{-1}F^\epsilon = \bigcup_{j=1}^\infty \bigcap_{k=j}^\infty (r_t^k)^{-1}F^\epsilon \subset r_t^{-1}F^{2\epsilon}$$

we deduce that  $r_t^{-1}F = \bigcap_\epsilon \liminf_k (r_t^k)^{-1}F^\epsilon$  is measurable. Thus  $r_t$  is measurable.

To prove the second assertion take an  $x \in \mathcal{D}_\infty$  which is continuous at  $t$ . If  $d_\infty(x_n, x) \rightarrow 0$ , then by Theorem 9.9,

$$d_t(r_t x_n, r_t x) = d_t(x_n, x) \rightarrow 0.$$

In other words, if  $x \notin J_t$ , then  $r_t$  is continuous at  $x$ .

**Theorem 9.21** A necessary and sufficient condition for  $P_n \Rightarrow P$  on  $\mathcal{D}_\infty$  is that  $P_n r_t^{-1} \Rightarrow P r_t^{-1}$  for each  $t \in T_P$ .

Proof. If  $P_n \Rightarrow P$  on  $\mathcal{D}_\infty$ , then  $P_n r_t^{-1} \Rightarrow P r_t^{-1}$  for each  $t \in T_P$  due to the mapping theorem and Lemma 9.20.

For the reverse implication, it is enough, by Lemma 9.17, to show that  $P_n \psi_i^{-1} \Rightarrow P \psi_i^{-1}$  on  $\mathcal{D}_i$  for every  $i$ . Given an  $i$  choose a  $t \in T_P$  so that  $t \geq i$ . Since  $\psi_i = \psi_{t,i} \circ r_t$ , the mapping theorem gives

$$P_n \psi_i^{-1} = (P_n r_t^{-1}) \psi_{t,i}^{-1} \Rightarrow (P r_t^{-1}) \psi_{t,i}^{-1} = P \psi_i^{-1}.$$

**Exercise 9.22** Let  $W^\circ$  be the standard Brownian bridge. For  $t \in [0, \infty)$  put  $W_t = (1+t)W_{\frac{t}{1+t}}^\circ$ . Show that such defined random element  $W$  of  $\mathcal{D}_\infty$  is a Gaussian process with zero means and covariance function  $\mathbb{E}(W_s W_t) = s$  for  $0 \leq s \leq t < \infty$ . This is a Wiener process  $W = (W_t, 0 \leq t < \infty)$ . Clearly,  $r_t(W)$  is a Wiener process which is a random element of  $\mathcal{D}_t$ .

**Corollary 9.23** *Let  $\xi_1, \xi_2, \dots$  be iid r.v. defined on  $(\Omega, \mathcal{F}, \mathbb{P})$ . If  $\xi_i$  have zero mean and variance  $\sigma^2$  and  $X_t^n = \frac{\xi_1 + \dots + \xi_{\lfloor nt \rfloor}}{\sigma\sqrt{n}}$ , then  $X^n \Rightarrow W$  on  $\mathbf{D}_\infty$ .*

Proof. By Theorem 8.4,  $X^n \Rightarrow W$  on  $\mathbf{D}_1$ . The same proof gives  $X^n \Rightarrow W$  on  $\mathbf{D}_t$  for each  $t \in [0, \infty)$ . In other words,  $r_t(X^n) \Rightarrow r_t(W)$  for each  $t \in [0, \infty)$ , and it remains to apply Theorem 9.21.

**Corollary 9.24** *Suppose for each  $n$ ,  $\xi_{n1}, \dots, \xi_{nn}$  are iid indicator r.v. with  $\mathbb{P}(\xi_{ni} = 1) = \alpha/n$ . If  $X_t^n = \sum_{i \leq nt} \xi_{ni}$ , then  $X^n \Rightarrow X$  on  $\mathbf{D}_\infty$ , where  $X$  is the Poisson process with parameter  $\alpha$ .*

Proof. Combine Corollary 7.7 and Theorem 9.21.