Chapter 13
Convergence of Measures

One focus of probability theory is distributions that are the result of an interplay of a large number of random impacts. Often a useful approximation can be obtained by taking a limit of such distributions, for example, a limit where the number of impacts goes to infinity. With the Poisson distribution, we have encountered such a limit distribution that occurs as the number of very rare events when the number of possibilities goes to infinity (see Theorem 3.7). In many cases, it is necessary to rescale the original distributions in order to capture the behavior of the essential fluctuations, e.g., in the central limit theorem. While these theorems work with real random variables, we will also see limit theorems where the random variables take values in more general spaces such as the space of continuous functions when we model the path of the random motion of a particle.

In this chapter, we provide the abstract framework for the investigation of convergence of measures. We introduce the notion of weak convergence of probability measures on general (mostly Polish) spaces and derive the fundamental properties. The reader will profit from a solid knowledge of point set topology. Thus we start with a short overview of some topological definitions and theorems.

We do not strive for the greatest generality but rather content ourselves with the key theorems for probability theory. For further reading, we recommend [14, 82].

At first reading, the reader might wish to skip this rather analytically flavored chapter. In this case, for the time being it suffices to get acquainted with the definitions of weak convergence and tightness (Definitions 13.12 and 13.26), as well as with the statements of the Portemanteau theorem (Theorem 13.16) and Prohorov's theorem (Theorem 13.29).

13.1 A Topology Primer

Excursively, we present some definitions and facts from point set topology. For details, see, e.g., [90].

In the following, let (E, τ) be a topological space with the Borel σ-algebra $\mathcal{E} = \mathcal{B}(E)$ (compare Definitions 1.20 and 1.21). We will also assume that (E, τ) is a

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_13,

Hausdorff space; that is, for any two points $x, y \in E$ with $x \neq y$, there exist disjoint open sets U, V such that $x \in U$ and $y \in V$.

For $A \subset E$, we denote by $\overline{A}$ the *closure* of A, by A° the *interior* and by ∂A the *boundary* of A. A set $A \subset E$ is called *dense* if $\overline{A} = E$.

(E, τ) is called *metrizable* if there exists a metric d on E such that τ is induced by the open balls $B_\varepsilon(x) := \{y \in E : d(x, y) < \varepsilon\}$. A metric d on E is called *complete* if any Cauchy sequence with respect to d converges in E. (E, τ) is called *completely metrizable* if there exists a complete metric on E that induces τ. If (E, d) is a metric space and $A, B \subset E$, then we write $d(A, B) = \inf\{d(x, y) : x \in A, y \in B\}$ and $d(x, B) := d(\{x\}, B)$ for $x \in E$.

A metrizable space (E, τ) is called *separable* if there exists a countable dense subset of E. Separability in metrizable spaces is equivalent to the existence of a *countable base of the topology*; that is, a countable set $\mathcal{U} \subset \tau$ with $A = \bigcup_{U \in \mathcal{U} : U \subset A} U$ for all $A \in \tau$. (For example, choose the ε-balls centered at the points of a countable subset and let ε run through the positive rational numbers.) A compact metric space is always separable (simply choose for each $n \in \mathbb{N}$ a finite cover $\mathcal{U}_n \subset \tau$ comprising balls of radius $\frac{1}{n}$ and then take $\mathcal{U} := \bigcup_{n \in \mathbb{N}} \mathcal{U}_n$).

A set $A \subset E$ is called *compact* if each open cover $\mathcal{U} \subset \tau$ of A (that is, $A \subset \bigcup_{U \in \mathcal{U}} U$) has a finite subcover; that is, a finite $\mathcal{U}' \subset \mathcal{U}$ with $A \subset \bigcup_{U \in \mathcal{U}'} U$. Compact sets are closed. By the Heine–Borel theorem, a subset of $\mathbb{R}^d$ is compact if and only if it is bounded and closed. $A \subset E$ is called *relatively compact* if $\overline{A}$ is compact. On the other hand, A is called *sequentially compact* (respectively *relatively sequentially compact*) if any sequence $(x_n)_{n \in \mathbb{N}}$ with values in A has a subsequence $(x_{n_k})_{k \in \mathbb{N}}$ that converges to some $x \in A$ (respectively $x \in \overline{A}$). In metrizable spaces, the notions *compact* and *sequentially compact* coincide. A set $A \subset E$ is called *σ-compact* if A is a countable union of compact sets. E is called *locally compact* if any point $x \in E$ has an open neighborhood whose closure is compact. A locally compact, separable metric space is manifestly σ-compact. If E is a locally compact metric space and if $U \subset E$ is open and $K \subset U$ is compact, then there exists a compact set L with $K \subset L^\circ \subset L \subset U$. (For example, for any $x \in K$, take an open ball $B_{\varepsilon_x}(x)$ of radius $\varepsilon_x > 0$ that is contained in U and that is relatively compact. By making ε_x smaller (if necessary), one can assume that the closure of this ball is contained in U. As K is compact, there are finitely many points $x_1, \ldots, x_n \in K$ with $K \subset V := \bigcup_{i=1}^n B_{\varepsilon_{x_i}}(x_i)$. By construction, $L = \overline{V} \subset U$ is compact.)

We present one type of topological space that is of particular importance in probability theory in a separate definition.

Definition 13.1 A topological space (E, τ) is called a *Polish space* if it is separable and if there exists a complete metric that induces the topology τ.

Examples of Polish spaces are countable discrete spaces (however, not $\mathbb{Q}$ with the usual topology), the Euclidean spaces $\mathbb{R}^n$, and the space $C([0, 1])$ of continuous functions $[0, 1] \to \mathbb{R}$, equipped with the supremum norm $\| \cdot \|_\infty$. In practice, all spaces that are of importance in probability theory are Polish spaces.

Let (E, d) be a metric space. A set $A \subset E$ is called *totally bounded* if, for any $\varepsilon > 0$, there exist finitely many points $x_1, \ldots, x_n \in A$ such that $A \subset \bigcup_{i=1}^{n} B_\varepsilon(x_i)$. Evidently, compact sets are totally bounded. In Polish spaces, a partial converse is true.

Lemma 13.2 *Let (E, τ) be a Polish space with complete metric d. A subset $A \subset E$ is totally bounded with respect to d if and only if A is relatively compact.*

Proof This is left as an exercise. □

In the following, let (E, τ) be a topological space with Borel σ-algebra $\mathcal{E} = \mathcal{B}(E) := \sigma(\tau)$ and with complete metric d. For measures on $(E, \mathcal{E})$, we introduce the following notions of regularity.

Definition 13.3 A σ-finite measure μ on $(E, \mathcal{E})$ is called

(i) *locally finite* or a *Borel measure* if, for any point $x \in E$, there exists an open neighborhood $U \ni x$ such that $\mu(U) < \infty$,
(ii) *inner regular* if

$$\mu(A) = \sup\{\mu(K) : K \subset A \text{ is compact}\} \quad \text{for all } A \in \mathcal{E}, \tag{13.1}$$

(iii) *outer regular* if

$$\mu(A) = \inf\{\mu(U) : U \supset A \text{ is open}\} \quad \text{for all } A \in \mathcal{E}, \tag{13.2}$$

(iv) *regular* if μ is inner and outer regular, and
(v) a *Radon measure* if μ is an inner regular Borel measure.

Definition 13.4 We introduce the following spaces of measures on E:

$$\begin{aligned}
\mathcal{M}(E) &:= \{\text{Radon measures on } (E, \mathcal{E})\}, \\
\mathcal{M}_f(E) &:= \{\text{finite measures on } (E, \mathcal{E})\}, \\
\mathcal{M}_1(E) &:= \{\mu \in \mathcal{M}_f(E) : \mu(E) = 1\}, \\
\mathcal{M}_{\le 1}(E) &:= \{\mu \in \mathcal{M}_f(E) : \mu(E) \le 1\}.
\end{aligned}$$

The elements of $\mathcal{M}_{\le 1}(E)$ are called *sub-probability measures* on E.

Further, we agree on the following notation for spaces of continuous functions:

$$\begin{aligned}
C(E) &:= \{f : E \to \mathbb{R} \text{ is continuous}\}, \\
C_b(E) &:= \{f \in C(E) \text{ is bounded}\}, \\
C_c(E) &:= \{f \in C(E) \text{ has compact support}\} \subset C_b(E).
\end{aligned}$$

Recall that the support of a real function f is $\overline{f^{-1}(\mathbb{R} \setminus \{0\})}$.

Unless otherwise stated, the vector spaces $C(E)$, $C_b(E)$ and $C_c(E)$ are equipped with the supremum norm.

Lemma 13.5 *If E is Polish and $\mu \in \mathcal{M}_f(E)$, then for any $\varepsilon > 0$, there is a compact set $K \subset E$ with $\mu(E \setminus K) < \varepsilon$.*

Proof Let $\varepsilon > 0$. For each $n \in \mathbb{N}$, there exists a sequence $x_1^n, x_2^n, \ldots \in E$ with $E = \bigcup_{i=1}^{\infty} B_{1/n}(x_i^n)$. Fix $N_n \in \mathbb{N}$ such that $\mu(E \setminus \bigcup_{i=1}^{N_n} B_{1/n}(x_i^n)) < \frac{\varepsilon}{2^n}$. Define

$$A := \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{N_n} B_{1/n}\big(x_i^n\big).$$

By construction, A is totally bounded. Since E is Polish, $\overline{A}$ is compact. Furthermore, it follows that $\mu(E \setminus \overline{A}) \le \mu(E \setminus A) < \sum_{n=1}^{\infty} \varepsilon 2^{-n} = \varepsilon$. □

Theorem 13.6 *If E is Polish and if $\mu \in \mathcal{M}_f(E)$, then μ is regular. In particular, in this case, $\mathcal{M}_f(E) \subset \mathcal{M}(E)$.*

Proof (*Outer regularity*) *Step 1*. Let $B \subset E$ be closed and let $\varepsilon > 0$. Let d be a complete metric on E. For $\delta > 0$, let

$$B_\delta := \big\{x \in E : d(x, B) < \delta\big\}$$

be the open δ-neighborhood of B. As B is closed, we have $\bigcap_{\delta > 0} B_\delta = B$. Since μ is upper semicontinuous (Theorem 1.36), there is a $\delta > 0$ such that $\mu(B_\delta) \le \mu(B) + \varepsilon$.

Step 2. Let $B \in \mathcal{E}$ and $\varepsilon > 0$. Consider the class of sets

$$\mathcal{A} := \{V \cap C : V \subset E \text{ open}, \ C \subset E \text{ closed}\}.$$

Clearly, we have $\mathcal{E} = \sigma(\mathcal{A})$. It is easy to check that $\mathcal{A}$ is a semiring. Hence by the approximation theorem for measures (Theorem 1.65), there are mutually disjoint sets $A_n = V_n \cap C_n \in \mathcal{A}$, $n \in \mathbb{N}$, such that $B \subset A := \bigcup_{n=1}^{\infty} A_n$ and $\mu(A) \le \mu(B) + \varepsilon/2$. As shown in the first step, for any $n \in \mathbb{N}$, there is an open set $W_n \supset C_n$ such that $\mu(W_n) \le \mu(C_n) + \varepsilon 2^{-n-1}$. Hence also $U_n := V_n \cap W_n$ is open. Let $B \subset U := \bigcup_{n=1}^{\infty} U_n$. We conclude that $\mu(U) \le \mu(A) + \sum_{n=1}^{\infty} \varepsilon 2^{-n-1} \le \mu(B) + \varepsilon$.

(Inner regularity) Replacing B by B^c, the outer regularity yields the existence of a closed set $D \subset B$ with $\mu(B \setminus D) < \varepsilon/2$. By Lemma 13.5, there exists a compact set K with $\mu(K^c) < \varepsilon/2$. Define $C = D \cap K$. Then $C \subset B$ is compact and $\mu(B \setminus C) < \varepsilon$. Hence μ is also inner regular. □

Corollary 13.7 *The Lebesgue measure λ on $\mathbb{R}^d$ is a regular Radon measure. However, not all σ-finite measures on $\mathbb{R}^d$ are regular.*

Proof Clearly, $\mathbb{R}^d$ is Polish and λ is locally finite. Let $A \in \mathcal{B}(\mathbb{R}^d)$ and $\varepsilon > 0$. There is an increasing sequence $(K_n)_{n \in \mathbb{N}}$ of compact sets with $K_n \uparrow \mathbb{R}^d$. Since any K_n is bounded, we have $\lambda(K_n) < \infty$. Hence, by the preceding theorem, for any $n \in \mathbb{N}$, there exists an open set $U_n \supset A \cap K_n$ with $\lambda(U_n \setminus A) < \varepsilon/2^n$. Thus $\lambda(U \setminus A) < \varepsilon$ for the open set $U := \bigcup_{n\in\mathbb{N}} U_n$.

If $\lambda(A) < \infty$, then there exists an $n \in \mathbb{N}$ with $\lambda(A \setminus K_n) < \varepsilon/2$. By the preceding theorem, there exists a compact set $C \subset A \cap K_n$ with $\lambda((A \cap K_n) \setminus C) < \varepsilon/2$. Therefore, $\lambda(A \setminus C) < \varepsilon$.

If, on the other hand, $\lambda(A) = \infty$, then for any $L > 0$, we have to find a compact set $C \subset A$ with $\lambda(C) > L$. However, $\lambda(A \cap K_n) \stackrel{n\to\infty}{\longrightarrow} \infty$; hence there exists an $n \in \mathbb{N}$ with $\lambda(A \cap K_n) > L + 1$. By what we have shown already, there exists a compact set $C \subset A \cap K_n$ with $\lambda((A \cap K_n) \setminus C) < 1$; hence $\lambda(C) > L$.

Finally, consider the measure $\mu = \sum_{q\in\mathbb{Q}} \delta_q$. Clearly, this measure is σ-finite; however, it is neither locally finite nor outer regular. □

Definition 13.8 Let (E, d_E) and (F, d_F) be metric spaces. A function $f : E \to F$ is called *Lipschitz continuous* if there exists a constant $K < \infty$, the so-called Lipschitz constant, with $d_F(f(x), f(y)) \le K \cdot d_E(x, y)$ for all $x, y \in E$. Denote by $\mathrm{Lip}_K(E; F)$ the space of Lipschitz continuous functions with constant K and by $\mathrm{Lip}(E; F) = \bigcup_{K>0} \mathrm{Lip}_K(E; F)$ the space of Lipschitz continuous functions on E.

We abbreviate $\mathrm{Lip}_K(E) := \mathrm{Lip}_K(E; \mathbb{R})$ and $\mathrm{Lip}(E) := \mathrm{Lip}(E; \mathbb{R})$.

Definition 13.9 Let $\mathcal{F} \subset \mathcal{M}(E)$ be a family of Radon measures. A family $\mathcal{C}$ of measurable maps $E \to \mathbb{R}$ is called a *separating family* for $\mathcal{F}$ if, for any two measures $\mu, \nu \in \mathcal{F}$, the following holds:

$$\left(\int f\, d\mu = \int f\, d\nu \text{ for all } f \in \mathcal{C} \cap \mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu)\right) \implies \mu = \nu.$$

Lemma 13.10 *Let (E, d) be a metric space. For any closed set $A \subset E$ and any $\varepsilon > 0$, there is a Lipschitz continuous map $\rho_{A,\varepsilon} : E \to [0, 1]$ with*

$$\rho_{A,\varepsilon}(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } d(x, A) \ge \varepsilon. \end{cases}$$

Proof Let $\varphi : \mathbb{R} \to [0, 1], t \mapsto (t \vee 0) \wedge 1$. For $x \in E$, define $\rho_{A,\varepsilon}(x) = 1 - \varphi(\varepsilon^{-1} d(x, A))$. □

Theorem 13.11 *Let (E, d) be a metric space.*

(i) $\mathrm{Lip}_1(E; [0, 1])$ *is separating for* $\mathcal{M}(E)$.
(ii) *If, in addition, E is locally compact, then $C_c(E) \cap \mathrm{Lip}_1(E; [0, 1])$ is separating for $\mathcal{M}(E)$.*

Proof (i) Assume $\mu_1, \mu_2 \in \mathcal{M}(E)$ are measures with $\int f\, d\mu_1 = \int f\, d\mu_2$ for all $f \in \mathrm{Lip}_1(E; [0, 1])$. If $A \in \mathcal{E}$, then $\mu_i(A) = \sup\{\mu_i(K) : K \subset A \text{ is compact}\}$ since the Radon measure μ_i is inner regular ($i = 1, 2$). Hence, it is enough to show that $\mu_1(K) = \mu_2(K)$ for any compact set K.

Now let $K \subset E$ be compact. Since μ_1 and μ_2 are locally finite, for every $x \in K$, there exists an open set $U_x \ni x$ with $\mu_1(U_x) < \infty$ and $\mu_2(U_x) < \infty$. Since K is compact, we can find finitely many points $x_1, \ldots, x_n \in K$ such that $K \subset U := \bigcup_{j=1}^n U_{x_j}$. By construction, $\mu_i(U) < \infty$; hence $\mathbb{1}_U \in L^1(\mu_i)$ for $i = 1, 2$. Since U^c is closed and since $U^c \cap K = \emptyset$, we get $\delta := d(U^c, K) > 0$. Let $\rho_{K,\varepsilon}$ be the map from Lemma 13.10. Hence $\mathbb{1}_K \le \rho_{K,\varepsilon} \le \mathbb{1}_U \in L^1(\mu_i)$ if $\varepsilon \in (0, \delta)$. Since $\rho_{K,\varepsilon} \overset{\varepsilon\to 0}{\longrightarrow} \mathbb{1}_K$, we get by dominated convergence (Corollary 6.26) that $\mu_i(K) = \lim_{\varepsilon\to 0} \int \rho_{K,\varepsilon}\, d\mu_i$. However, $\varepsilon\rho_{K,\varepsilon} \in \mathrm{Lip}_1(E; [0, 1])$ for all $\varepsilon > 0$; hence, by assumption,

$$\int \rho_{K,\varepsilon}\, d\mu_1 = \varepsilon^{-1} \int (\varepsilon\rho_{K,\varepsilon})\, d\mu_1 = \varepsilon^{-1} \int (\varepsilon\rho_{K,\varepsilon})\, d\mu_2 = \int \rho_{K,\varepsilon}\, d\mu_2.$$

This implies $\mu_1(K) = \mu_2(K)$; hence $\mu_1 = \mu_2$.

(ii) If E is locally compact, then in (i) we can choose the neighborhoods U_x to be relatively compact. Hence U is relatively compact; thus $\rho_{K,\varepsilon}$ has compact support and is thus in $C_c(E)$ for all $\varepsilon \in (0, \delta)$. □

Exercise 13.1.1

(i) Show that $C([0, 1])$ has a separable dense subset.
(ii) Show that the space $(C_b([0, \infty)), \|\cdot\|_\infty)$ of bounded continuous functions, equipped with the supremum norm, is not separable.
(iii) Show that the space $C_c([0, \infty))$ of continuous functions with compact support, equipped with the supremum norm, is separable.

Exercise 13.1.2 Let μ be a locally finite measure. Show that $\mu(K) < \infty$ for any compact set K.

Exercise 13.1.3 (Lusin's theorem) Let Ω be a Polish space, let μ be a σ-finite measure on $(\Omega, \mathcal{B}(\Omega))$ and let $f : \Omega \to \mathbb{R}$ be a map. Show that the following two statements are equivalent:

(i) There is a Borel measurable map $g : \Omega \to \mathbb{R}$ with $f = g$ μ-almost everywhere.
(ii) For any $\varepsilon > 0$, there is a compact set K_ε with $\mu(\Omega \setminus K_\varepsilon) < \varepsilon$ such that the restricted function $f|_{K_\varepsilon}$ is continuous.

Exercise 13.1.4 Let $\mathcal{U}$ be a family of intervals in $\mathbb{R}$ such that $W := \bigcup_{U\in\mathcal{U}} U$ has finite Lebesgue measure $\lambda(W)$. Show that for any $\varepsilon > 0$, there exist finitely many pairwise disjoint sets $U_1, \ldots, U_n \in \mathcal{U}$ with

$$\sum_{i=1}^n \lambda(U_i) > \frac{1-\varepsilon}{3}\lambda(W).$$

Hint: Choose a finite family $\mathcal{U}' \subset \mathcal{U}$ such that $\bigcup_{U\in\mathcal{U}'} U$ has Lebesgue measure at least $(1-\varepsilon)\lambda(W)$. Choose a maximal sequence $\mathcal{U}''$ (sorted by decreasing lengths) of *disjoint* intervals and show that each $U \in \mathcal{U}'$ is in $(x - 3a, x + 3a)$ for some $(x - a, x + a) \in \mathcal{U}''$.

Exercise 13.1.5 Let $C \subset \mathbb{R}^d$ be an open, bounded and convex set and assume that $\mathcal{U} \subset \{x + rC : x \in \mathbb{R}^d, r > 0\}$ is such that $W := \bigcup_{U \in \mathcal{U}} U$ has finite Lebesgue measure $\lambda^d(W)$. Show that for any $\varepsilon > 0$, there exist finitely many pairwise disjoint sets $U_1, \ldots, U_n \in \mathcal{U}$ such that

$$\sum_{i=1}^{n} \lambda^d(U_i) > \frac{1-\varepsilon}{3^d} \lambda(W).$$

Show by a counterexample that the condition of similarity of the open sets in $\mathcal{U}$ is essential.

Exercise 13.1.6 Let μ be a Radon measure on $\mathbb{R}^d$ and let $A \in \mathcal{B}(\mathbb{R}^d)$ be a μ-null set. Let $C \subset \mathbb{R}^d$ be bounded, convex and open with $0 \in C$. Use Exercise 13.1.5 to show that

$$\lim_{r \downarrow 0} \frac{\mu(x + rC)}{r^d} = 0 \quad \text{for } \lambda^d\text{-almost all } x \in A.$$

Conclude that if F is the distribution function of a Stieltjes measure μ on $\mathbb{R}$ and if $A \in \mathcal{B}(\mathbb{R})$ is a μ-null set, then $\frac{d}{dx} F(x) = 0$ for λ-almost all $x \in A$.

Exercise 13.1.7 (Fundamental theorem of calculus) (Compare [37].) Let $f \in \mathcal{L}^1(\mathbb{R}^d)$, $\mu = f\lambda^d$ and let $C \subset \mathbb{R}^d$ be open, convex and bounded with $0 \in C$. Show that

$$\lim_{r \downarrow 0} \frac{\mu(x + rC)}{r^d \lambda^d(C)} = f(x) \quad \text{for } \lambda^d\text{-almost all } x \in \mathbb{R}^d.$$

For the case $d = 1$, conclude the fundamental theorem of calculus:

$$\frac{d}{dx} \int_{[0,x]} f \, d\lambda = f(x) \quad \text{for } \lambda\text{-almost all } x \in \mathbb{R}.$$

Hint: Use Exercise 13.1.6 with

$$\mu_q(dx) = (f(x) - q)^+ \lambda^d(dx) \quad \text{for } q \in \mathbb{Q},$$

as well as the inequality

$$\frac{\mu(x + rC)}{r^d \lambda^d(C)} \le q + \frac{\mu_q(x + rC)}{r^d \lambda^d(C)}.$$

Exercise 13.1.8 Similarly as in Corollary 13.7, show the following: Let E be a σ-compact polish space and let μ be a measure on E. Then μ is a Radon measure if and only if $\mu(K) < \infty$ for any compact $K \subset E$.

13.2 Weak and Vague Convergence

In Theorem 13.11, we saw that integrals of bounded continuous functions f determine a Radon measure on a metric space (E, d). If E is locally compact, it is enough to consider f with compact support. This suggests that we can use $C_b(E)$ and $C_c(E)$ as classes of test functions in order to define the convergence of measures.

Definition 13.12 (Weak and vague convergence) Let E be a metric space.

(i) Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_f(E)$. We say that $(\mu_n)_{n\in\mathbb{N}}$ *converges weakly* to μ, formally $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ (weakly) or $\mu = \text{w-lim}_{n\to\infty} \mu_n$, if

$$\int f\, d\mu_n \overset{n\to\infty}{\longrightarrow} \int f\, d\mu \quad \text{for all } f \in C_b(E).$$

(ii) Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}(E)$. We say that $(\mu_n)_{n\in\mathbb{N}}$ *converges vaguely* to μ, formally $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ (vaguely) or $\mu = \text{v-lim}_{n\to\infty} \mu_n$, if

$$\int f\, d\mu_n \overset{n\to\infty}{\longrightarrow} \int f\, d\mu \quad \text{for any } f \in C_c(E).$$

Remark 13.13 If E is Polish, then by Theorems 13.6 and 13.11, the weak limit is unique. The same holds for the vague limit if E is locally compact. ◊

Remark 13.14

(i) In functional analysis the notion of weak convergence is somewhat different. Starting from a normed vector space X (here the space of finite signed measures with the total variation norm), consider the space X' of continuous linear functionals $X \to \mathbb{R}$. The sequence (μ_n) in X converges *weakly* to $\mu \in X$, if $\Phi(\mu_n) \overset{n\to\infty}{\longrightarrow} \Phi(\mu)$ for every $\Phi \in X'$. In the case of finite signed measures this is equivalent to: (μ_n) is bounded and $\mu_n(A) \overset{n\to\infty}{\longrightarrow} \mu(A)$ for any measurable A (see [38, Theorem IV.9.5]). Comparing this to Theorem 13.16(vi), we see that the functional analysis notion of weak convergence is stronger than ours in Definition 13.12.

(ii) Weak convergence (as introduced in Definition 13.12) induces on $\mathcal{M}_f(E)$ the *weak topology* τ_w. This is the coarsest topology such that for all $f \in C_b(E)$, the map $\mathcal{M}_f(E) \to \mathbb{R}$, $\mu \mapsto \int f\, d\mu$ is continuous. In functional analysis, τ_w corresponds to the so-called weak*-topology. Starting from a normed vector space X (here $X = C_b(E)$ with the norm $\|\cdot\|_\infty$), we define the weak*-topology on the dual space X' by writing $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ if and only if $\mu_n(x) \overset{n\to\infty}{\longrightarrow} \mu(x)$ for all $x \in X$. Clearly, each μ defines a continuous linear form on $C_b(E)$ by $f \mapsto \mu(f) := \int f\, d\mu$. Hence $\mathcal{M}_f(E) \subset C_b(E)'$. This implies that τ_w is the trace of the weak*-topology on $\mathcal{M}_f(E)$.

(iii) If E is separable, then it can be shown that $(\mathcal{M}_f(E), \tau_w)$ is metrizable; for example, by virtue of the so-called *Prohorov metric*. This is defined by

$$d_P(\mu, \nu) := \max\{d'_P(\mu, \nu), d'_P(\nu, \mu)\}, \tag{13.3}$$

where

$$d'_P(\mu, \nu) := \inf\{\varepsilon > 0 : \mu(B) \le \nu(B^\varepsilon) + \varepsilon \text{ for any } B \in \mathcal{B}(E)\}, \tag{13.4}$$

and where $B^\varepsilon = \{x : d(x, B) < \varepsilon\}$; see, e.g., [14, Appendix III, Theorem 5]. (It can be shown that $d'_P(\mu, \nu) = d'_P(\nu, \mu)$ if $\mu, \nu \in \mathcal{M}_1(E)$.) If E is locally compact and Polish, then $(\mathcal{M}_f(E), \tau_w)$ is again Polish (see [136, p. 167]).

(iv) Similarly, the *vague topology* τ_v on $\mathcal{M}(E)$ is the coarsest topology such that for all $f \in C_c(E)$, the map $\mathcal{M}(E) \to \mathbb{R}$, $\mu \mapsto \int f \, d\mu$ is continuous. If E is locally compact, then $(\mathcal{M}(E), \tau_v)$ is a Hausdorff space. If, in addition, E is Polish, then $(\mathcal{M}(E), \tau_v)$ is again Polish (see, e.g., [82, Section 15.7]). ◊

While weak convergence implies convergence of the total masses (since $1 \in C_b(E)$), with vague convergence a mass defect (but not a mass gain) can be experienced in the limit.

Lemma 13.15 *Let E be a locally compact Polish space and let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}(E)$ be measures such that $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ vaguely. Then*

$$\mu(E) \le \liminf_{n\to\infty} \mu_n(E).$$

Proof Let $(f_N)_{N\in\mathbb{N}}$ be a sequence in $C_c(E; [0, 1])$ with $f_N \uparrow 1$. Then

$$\begin{aligned}\mu(E) &= \sup_{N\in\mathbb{N}} \int f_N \, d\mu \\ &= \sup_{N\in\mathbb{N}} \lim_{n\to\infty} \int f_N \, d\mu_n \\ &\le \liminf_{n\to\infty} \sup_{N\in\mathbb{N}} \int f_N \, d\mu_n \\ &= \liminf_{n\to\infty} \mu_n(E).\end{aligned}$$

□

Clearly, the sequence $(\delta_{1/n})_{n\in\mathbb{N}}$ of probability measures on $\mathbb{R}$ converges weakly to δ_0; however, not in total variation norm. Indeed, for the closed set $(-\infty, 0]$, we have $\lim_{n\to\infty} \delta_{1/n}((-\infty, 0]) = 0 < 1 = \delta_0((-\infty, 0])$. Loosely speaking, at the boundaries of closed sets, mass can immigrate but not emigrate. The opposite is true for open sets: $\lim_{n\to\infty} \delta_{1/n}((0, \infty)) = 1 > 0 = \delta_0((0, \infty))$. Here mass can emigrate but not immigrate. In fact, weak convergence can be characterized by this property. In the following theorem, a whole bunch of such statements will be hung on a coat hanger (French: *portemanteau*).

For measurable $g : \Omega \to \mathbb{R}$, let U_g be the set of points of discontinuity of g. Recall from Exercise 1.1.3 that U_g is Borel measurable.

Theorem 13.16 (Portemanteau) *Let E be a metric space and let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\leq 1}(E)$. The following are equivalent.*

(i) $\mu = \text{w-lim}_{n\to\infty} \mu_n$.
(ii) $\int f\, d\mu_n \overset{n\to\infty}{\longrightarrow} \int f\, d\mu$ *for all bounded Lipschitz continuous f.*
(iii) $\int f\, d\mu_n \overset{n\to\infty}{\longrightarrow} \int f\, d\mu$ *for all bounded measurable f with $\mu(U_f) = 0$.*
(iv) $\liminf_{n\to\infty} \mu_n(E) \geq \mu(E)$ *and* $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for all closed $F \subset E$.*
(v) $\limsup_{n\to\infty} \mu_n(E) \leq \mu(E)$ *and* $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *for all open $G \subset E$.*
(vi) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for all measurable A with $\mu(\partial A) = 0$.*

If E is locally compact and Polish, then in addition each of the following is equivalent to the previous statements.

(vii) $\mu = \text{v-lim}_{n\to\infty} \mu_n$ *and* $\mu(E) = \lim_{n\to\infty} \mu_n(E)$.
(viii) $\mu = \text{v-lim}_{n\to\infty} \mu_n$ *and* $\mu(E) \geq \limsup_{n\to\infty} \mu_n(E)$.

Proof "(iv) $\Longleftrightarrow$ (v)$\Longrightarrow$(vi)" This is trivial.

"(iii) $\Longrightarrow$(i)$\Longrightarrow$(ii)" This is trivial.

"(ii) $\Longrightarrow$(iv)" Convergence of the total masses follows by using the test function $1 \in \text{Lip}(E; [0, 1])$. Let F be closed and let $\rho_{F,\varepsilon}$ be as in Lemma 13.10. Then

$$\limsup_{n\to\infty} \mu_n(F) \leq \inf_{\varepsilon>0} \lim_{n\to\infty} \int \rho_{F,\varepsilon}\, d\mu_n = \inf_{\varepsilon>0} \int \rho_{F,\varepsilon}\, d\mu = \mu(F)$$

since $\rho_{F,\varepsilon}(x) \overset{\varepsilon\to 0}{\longrightarrow} \mathbb{1}_F(x)$ for all $x \in E$.

"(viii) $\Longrightarrow$(vii)" This is obvious by Lemma 13.15.

"(i) $\Longrightarrow$(vii)" This is clear since $C_c(E) \subset C_b(E)$ and $1 \in C_b(E)$.

"(vii) $\Longrightarrow$(v)" Let G be open and $\varepsilon > 0$. Since μ is inner regular (Theorem 13.6), there is a compact set $K \subset G$ with $\mu(G) - \mu(K) < \varepsilon$. As E is locally compact, there is a compact set L with $K \subset L^\circ \subset L \subset G$. Let $\delta := d(K, L^c) > 0$ and let $\rho_{K,\delta}$ be as in Lemma 13.10. Then $\mathbb{1}_K \leq \rho_{K,\delta} \leq \mathbb{1}_L$; hence $\rho_{K,\delta} \in C_c(E)$ and thus

$$\liminf_{n\to\infty} \mu_n(G) \geq \liminf_{n\to\infty} \int \rho_{K,\delta}\, d\mu_n = \int \rho_{K,\delta}\, d\mu \geq \mu(K) \geq \mu(G) - \varepsilon.$$

Letting $\varepsilon \to 0$, we get (v).

"(vi) $\Longrightarrow$(iii)" Let $f : E \to \mathbb{R}$ be bounded and measurable with $\mu(U_f) = 0$. We make the elementary observation that for all $D \subset \mathbb{R}$,

$$\partial f^{-1}(D) \subset f^{-1}(\partial D) \cup U_f. \tag{13.5}$$

Indeed, if f is continuous at $x \in E$, then for any $\delta > 0$, there is an $\varepsilon(\delta) > 0$ with $f(B_{\varepsilon(\delta)}(x)) \subset B_\delta(f(x))$. If $x \in \partial f^{-1}(D)$, then there are $y \in f^{-1}(D) \cap B_{\varepsilon(\delta)}(x)$

and $z \in f^{-1}(D^c) \cap B_{\varepsilon(\delta)}(x)$. Therefore, $f(y) \in B_\delta(f(x)) \cap D \neq \emptyset$ and $f(z) \in B_\delta(f(x)) \cap D^c \neq \emptyset$; hence $f(x) \in \partial D$.

Let $\varepsilon > 0$. Evidently, the set $A := \{y \in \mathbb{R} : \mu(f^{-1}(\{y\})) > 0\}$ of atoms of the finite measure $\mu \circ f^{-1}$ is at most countable. Hence, there exist $N \in \mathbb{N}$ and $y_0 \le -\|f\|_\infty < y_1 < \ldots < y_{N-1} < \|f\|_\infty < y_N$ such that

$$y_i \in \mathbb{R} \setminus A \quad \text{and} \quad |y_{i+1} - y_i| < \varepsilon \quad \text{for all } i.$$

Let $E_i = f^{-1}([y_{i-1}, y_i))$ for $i = 1, \ldots, N$. Then $E = \biguplus_{i=1}^N E_i$ and by (13.5),

$$\mu(\partial E_i) \le \mu\big(f^{-1}\big(\{y_{i-1}\}\big)\big) + \mu\big(f^{-1}\big(\{y_i\}\big)\big) + \mu(U_f) = 0.$$

Therefore,

$$\begin{aligned}\limsup_{n\to\infty} \int f\,d\mu_n &\le \limsup_{n\to\infty} \sum_{i=1}^N \mu_n(E_i)\cdot y_i = \sum_{i=1}^N \mu(E_i)\cdot y_i \\ &\le \varepsilon + \int f\,d\mu.\end{aligned}$$

We let $\varepsilon \to 0$ and obtain $\limsup_{n\to\infty} \int f\,d\mu_n \le \int f\,d\mu$. Finally, consider $(-f)$ to obtain the reverse inequality $\liminf_{n\to\infty} \int f\,d\mu_n \ge \int f\,d\mu$. □

Definition 13.17 Let $X, X_1, X_2, \ldots$ be random variables with values in E. We say that $(X_n)_{n\in\mathbb{N}}$ *converges in distribution* to X, formally $X_n \xrightarrow{\mathcal{D}} X$ or $X_n \overset{n\to\infty}{\Longrightarrow} X$, if the distributions converge weakly and hence if $P_X = \text{w-lim}_{n\to\infty} \mathbf{P}_{X_n}$. Sometimes we write $X_n \xrightarrow{\mathcal{D}} \mathbf{P}_X$ or $X_n \overset{n\to\infty}{\Longrightarrow} \mathbf{P}_X$ if we want to specify only the distribution $\mathbf{P}_X$ but not the random variable X.

Theorem 13.18 (Slutzky's theorem) *Let $X, X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be random variables with values in E. Assume $X_n \xrightarrow{\mathcal{D}} X$ and $d(X_n, Y_n) \xrightarrow{n\to\infty} 0$ in probability. Then $Y_n \xrightarrow{\mathcal{D}} X$.*

Proof Let $f : E \to \mathbb{R}$ be bounded and Lipschitz continuous with constant K. Then

$$\big|f(x) - f(y)\big| \le K d(x,y) \wedge 2\|f\|_\infty \quad \text{for all } x, y \in E.$$

Dominated convergence yields $\limsup_{n\to\infty} \mathbf{E}[|f(X_n) - f(Y_n)|] = 0$. Hence we have

$$\begin{aligned}&\limsup_{n\to\infty} \big|\mathbf{E}\big[f(Y_n)\big] - \mathbf{E}\big[f(X)\big]\big| \\ &\quad\le \limsup_{n\to\infty} \big|\mathbf{E}\big[f(X)\big] - \mathbf{E}\big[f(X_n)\big]\big| + \limsup_{n\to\infty} \big|\mathbf{E}\big[f(X_n) - f(Y_n)\big]\big| = 0.\end{aligned}$$

□

Corollary 13.19 *If* $X_n \overset{n\to\infty}{\longrightarrow} X$ *in probability, then* $X_n \overset{\mathcal{D}}{\longrightarrow} X$, $n \to \infty$. *The converse is false in general.*

Example 13.20 If $X, X_1, X_2, \ldots$ are i.i.d. (with nontrivial distribution), then trivially $X_n \overset{\mathcal{D}}{\longrightarrow} X$ but not $X_n \overset{n\to\infty}{\longrightarrow} X$ in probability.

Recall the definition of a distribution function of a probability measure from Definition 1.59. ♢

Definition 13.21 Let $F, F_1, F_2, \ldots$ be distribution functions of probability measures on $\mathbb{R}$. We say that $(F_n)_{n\in\mathbb{N}}$ *converges weakly* to F, formally $F_n \overset{n\to\infty}{\Longrightarrow} F$, $F_n \overset{\mathcal{D}}{\longrightarrow} F$ or $F = \text{w-lim}_{n\to\infty} F_n$, if

$$F(x) = \lim_{n\to\infty} F_n(x) \quad \text{for all points of continuity } x \text{ of } F. \tag{13.6}$$

If $F, F_1, F_2, \ldots$ are distribution functions of sub-probability measures, then we define $F(\infty) := \lim_{x\to\infty} F(x)$ and for weak convergence require in addition $F(\infty) \geq \limsup_{n\to\infty} F_n(\infty)$.

Note that (13.6) implies $F(\infty) \leq \liminf_{n\to\infty} F_n(\infty)$. Hence, if $F_n \overset{\mathcal{D}}{\longrightarrow} F$, then $F(\infty) = \lim_{n\to\infty} F_n(\infty)$.

Example 13.22 If F is the distribution function of a probability measure on $\mathbb{R}$ and $F_n(x) := F(x+n)$ for $x \in \mathbb{R}$, then $(F_n)_{n\in\mathbb{N}}$ converges pointwise to 1. However, this is not a distribution function, as 1 does not converge to 0 for $x \to -\infty$. On the other hand, if $G_n(x) = F(x-n)$, then $(G_n)_{n\in\mathbb{N}}$ converges pointwise to $G \equiv 0$. However, $G(\infty) = 0 < \limsup_{n\to\infty} G_n(\infty) = 1$; hence we do not have weak convergence here either. Indeed, in each case, there is a mass defect in the limit (in the case of the F_n on the left and in the case of the G_n on the right). However, the definition of weak convergence of distribution functions is constructed so that no mass defect occurs in the limit. ♢

Theorem 13.23 *Let* $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\leq 1}(\mathbb{R})$ *with corresponding distribution functions* $F, F_1, F_2, \ldots$. *The following are equivalent.*

(i) $\mu = \text{w-lim}_{n\to\infty} \mu_n$.
(ii) $F_n \overset{\mathcal{D}}{\longrightarrow} F$.

Proof "(i)$\Longrightarrow$(ii)" Let F be continuous at x. Then $\mu(\partial(-\infty, x]) = \mu(\{x\}) = 0$. By Theorem 13.16, $F_n(x) = \mu_n((-\infty, x]) \overset{n\to\infty}{\longrightarrow} \mu((-\infty, x]) = F(x)$.

"(ii)$\Longrightarrow$(i)" Let $f \in \text{Lip}_1(\mathbb{R}; [0, 1])$. By Theorem 13.16, it is enough to show that

$$\int f\, d\mu_n \overset{n\to\infty}{\longrightarrow} \int f\, d\mu. \tag{13.7}$$

Let $\varepsilon > 0$. Fix $N \in \mathbb{N}$ and choose $N+1$ points of continuity $y_0 < y_1 < \ldots < y_N$ of F such that $F(y_0) < \varepsilon$, $F(y_N) > F(\infty) - \varepsilon$ and $y_i - y_{i-1} < \varepsilon$ for all i. Then

$$\int f\,d\mu_n \le \big(F_n(y_0) + F_n(\infty) - F_n(y_N)\big) \\ + \sum_{i=1}^{N} \big(f(y_i) + \varepsilon\big)\big(F_n(y_i) - F_n(y_{i-1})\big).$$

By assumption, $\lim_{n\to\infty} F_n(\infty) = F(\infty)$ and $F_n(y_i) \overset{n\to\infty}{\longrightarrow} F(y_i)$ for every $i = 0, \ldots, N$; hence

$$\limsup_{n\to\infty} \int f\,d\mu_n \le 3\varepsilon + \sum_{i=1}^{N} f(y_i)\big(F(y_i) - F(y_{i-1})\big) \le 4\varepsilon + \int f\,d\mu.$$

Therefore,

$$\limsup_{n\to\infty} \int f\,d\mu_n \le \int f\,d\mu.$$

Replacing f by $(1-f)$, we get (13.7). □

Corollary 13.24 *Let $X, X_1, X_2, \ldots$ be real random variables with distribution functions $F, F_1, F_2, \ldots$. Then the following are equivalent.*

(i) $X_n \overset{\mathcal{D}}{\longrightarrow} X$.
(ii) $\mathbf{E}[f(X_n)] \overset{n\to\infty}{\longrightarrow} \mathbf{E}[f(X)]$ *for all* $f \in C_b(\mathbb{R})$.
(iii) $F_n \overset{\mathcal{D}}{\longrightarrow} F$.

How stable is weak convergence if we pass to image measures under some map φ? Clearly, we need a certain continuity of φ at least at those points where the limit measure puts mass. The following theorem formalizes this idea and will come in handy in many applications.

Theorem 13.25 (Continuous mapping theorem) *Let (E_1, d_1) and (E_2, d_2) be metric spaces and let $\varphi : E_1 \to E_2$ be measurable. Denote by U_φ the set of points of discontinuity of φ.*

(i) *If $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\le 1}(E_1)$ with $\mu(U_\varphi) = 0$ and $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ weakly, then $\mu_n \circ \varphi^{-1} \overset{n\to\infty}{\longrightarrow} \mu \circ \varphi^{-1}$ weakly.*
(ii) *If $X, X_1, X_2, \ldots$ are E_1-valued random variables with $\mathbf{P}[X \in U_\varphi] = 0$ and $X_n \overset{\mathcal{D}}{\longrightarrow} X$, then $\varphi(X_n) \overset{\mathcal{D}}{\longrightarrow} \varphi(X)$.*

Proof First note that $U_\varphi \subset E_1$ is Borel measurable by Exercise 1.1.3. Hence the conditions make sense.

(i) Let $f \in C_b(E_2)$. Then $f \circ \varphi$ is bounded and measurable and $U_{f\circ\varphi} \subset U_\varphi$; hence $\mu(U_{f\circ\varphi}) = 0$. By Theorem 13.16,

$$\lim_{n\to\infty} \int f \, d\big(\mu_n \circ \varphi^{-1}\big) = \lim_{n\to\infty} \int (f \circ \varphi) \, d\mu_n$$

$$= \int (f \circ \varphi) \, d\mu = \int f \, d\big(\mu \circ \varphi^{-1}\big).$$

(ii) This is obvious since $\mathbf{P}_{\varphi(X)} = \mathbf{P}_X \circ \varphi^{-1}$. □

Exercise 13.2.1 Recall d'_P from (13.4). Show that $d_P(\mu, \nu) = d'_P(\mu, \nu) = d'_P(\nu, \mu)$ for all $\mu, \nu \in \mathcal{M}_1(E)$.

Exercise 13.2.2 Show that the topology of weak convergence on $\mathcal{M}_f(E)$ is coarser than the topology induced on $\mathcal{M}_f(E)$ by the total variation norm (see Corollary 7.45). That is, $\|\mu_n - \mu\|_{TV} \overset{n\to\infty}{\longrightarrow} 0$ implies $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ weakly.

Exercise 13.2.3 Let $E = \mathbb{R}$ and $\mu_n = \frac{1}{n} \sum_{k=0}^{n} \delta_{k/n}$. Let $\mu = \lambda|_{[0,1]}$ be the Lebesgue measure restricted to $[0, 1]$. Show that $\mu = \text{w-lim}_{n\to\infty} \mu_n$.

Exercise 13.2.4 Let $E = \mathbb{R}$ and λ be the Lebesgue measure on $\mathbb{R}$. For $n \in \mathbb{N}$, let $\mu_n = \lambda|_{[-n,n]}$. Show that $\lambda = \text{v-lim}_{n\to\infty} \mu_n$ but that $(\mu_n)_{n\in\mathbb{N}}$ does not converge weakly.

Exercise 13.2.5 Let $E = \mathbb{R}$ and $\mu_n = \delta_n$ for $n \in \mathbb{N}$. Show that $\text{v-lim}_{n\to\infty} \mu_n = 0$ but that $(\mu_n)_{n\in\mathbb{N}}$ does not converge weakly.

Exercise 13.2.6 (Lévy metric) For two probability distribution functions F and G on $\mathbb{R}$, define the Lévy distance by

$$d(F, G) = \inf\big\{\varepsilon \geq 0 : G(x - \varepsilon) - \varepsilon \leq F(x) \leq G(x + \varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R}\big\}.$$

Show the following:

(i) d is a metric on the set of distribution functions.
(ii) $F_n \overset{n\to\infty}{\Longrightarrow} F$ if and only if $d(F_n, F) \overset{n\to\infty}{\longrightarrow} 0$.
(iii) For every $P \in \mathcal{M}_1(\mathbb{R})$, there is a sequence $(P_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_1(\mathbb{R})$ such that each P_n has finite support and such that $P_n \overset{n\to\infty}{\Longrightarrow} P$.

Exercise 13.2.7 We can extend the notions of *weak convergence* and *vague convergence* to signed measures; that is, to differences $\varphi := \mu^+ - \mu^-$ of measures from $\mathcal{M}_f(E)$ and $\mathcal{M}(E)$, respectively, by repeating the words of Definition 13.12 for these classes. Show that the topology of weak convergence is not metrizable in general.

Hint: Consider $E = [0, 1]$.

(i) For $n \in \mathbb{N}$, define $\varphi_n = \delta_{1/n} - \delta_{2/n}$. Show that, for any $C > 0$, $(C\varphi_n)_{n\in\mathbb{N}}$ converges weakly to the zero measure.
(ii) Assume there is a metric that induces weak convergence. Show that then there would be a sequence $(C_n)_{n\in\mathbb{N}}$ with $C_n \uparrow \infty$ and $0 = \text{w-lim}_{n\to\infty}(C_n\varphi_n)$.
(iii) Choose an $f \in C([0,1])$ with $f(2^{-n}) = (-1)^n C_n^{-1/2}$ for any $n \in \mathbb{N}$, and show that $(\int f\, d(C_n\varphi_n))_{n\in\mathbb{N}}$ does not converge to zero.
(iv) Use this construction to contradict the assumption of metrizability.

Exercise 13.2.8 Show that (13.3) defines a metric on $\mathcal{M}_1(E)$ and that this metric induces the topology of weak convergence.

Exercise 13.2.9 Show the implication "(vi)$\Longrightarrow$(iv)" of Theorem 13.16 directly.

Exercise 13.2.10 Let $X, X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be real random variables. Assume $\mathbf{P}_{Y_n} = \mathcal{N}_{0,1/n}$ for all $n \in \mathbb{N}$. Show that $X_n \xrightarrow{\mathcal{D}} X$ if and only if $X_n + Y_n \xrightarrow{\mathcal{D}} X$.

Exercise 13.2.11 For each $n \in \mathbb{N}$, let X_n be a geometrically distributed random variable with parameter $p_n \in (0,1)$. How must we choose the sequence $(p_n)_{n\in\mathbb{N}}$ in order that $\mathbf{P}_{X_n/n}$ converges weakly to the exponential distribution with parameter $\alpha > 0$?

Exercise 13.2.12 Let $X, X_1, X_2, \ldots$ be real random variables with $X_n \overset{n\to\infty}{\Longrightarrow} X$. Show the following.

(i) $\mathbf{E}[|X|] \le \liminf_{n\to\infty} \mathbf{E}[|X_n|]$.
(ii) Let $r > p > 0$. If $\sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|^r] < \infty$, then $\mathbf{E}[|X|^p] = \lim_{n\to\infty} \mathbf{E}[|X_n|^p]$.

Exercise 13.2.13 Let $F, F_1, F_2, \ldots$ be probability distribution functions on $\mathbb{R}$, and assume $F_n \overset{n\to\infty}{\Longrightarrow} F$. Let $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \ge u\}$, $u \in (0,1)$, be the left continuous inverse of F (see the proof of Theorem 1.104). Show that

$$F_n^{-1}(u) \overset{n\to\infty}{\longrightarrow} F^{-1}(u) \text{ at every point of continuity } u \text{ of } F^{-1}.$$

Conclude that $F^{-1}(u) \overset{n\to\infty}{\longrightarrow} F^{-1}(u)$ for Lebesgue almost all $u \in (0,1)$.

Exercise 13.2.14 Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_1(\mathbb{R})$ with $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ weakly. Show that there exists a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and real random variables $X, X_1, X_2, \ldots$ on $(\Omega, \mathcal{A}, \mathbf{P})$ with distributions $\mathbf{P}_X = \mu$ and $\mathbf{P}_{X_n} = \mu_n$, $n \in \mathbb{N}$, such that

$$X_n \overset{n\to\infty}{\longrightarrow} X \quad \mathbf{P}\text{-a.s.}$$

Hint: Use Exercise 13.2.13.

Exercise 13.2.15 Let (E, d) be a metric space and let $\mu, \mu_1, \mu_2, \ldots$ be probability measures on E. A measurable map $f : E \to \mathbb{R}$ is called uniformly integrable with respect to $(\mu_n)_{n\in\mathbb{N}}$, if

$$\inf_{a>0} \sup_{n\in\mathbb{N}} \int_{\{|f|>a\}} |f| \, d\mu_n = 0.$$

Let f be continuous and uniformly integrable with respect to $(\mu_n)_{n\in\mathbb{N}}$ and assume that $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ weakly. Show that $\int |f| \, d\mu < \infty$ and that

$$\int f \, d\mu_n \overset{n\to\infty}{\longrightarrow} \int f \, d\mu.$$

Hint: Apply Exercise 13.2.14 to the image measures $\mu_n \circ f^{-1}$.

13.3 Prohorov's Theorem

In the following, let E be a Polish space with Borel σ-algebra $\mathcal{E}$. A fundamental question is: When does a sequence $(\mu_n)_{n\in\mathbb{N}}$ of measures on $(E, \mathcal{E})$ converge weakly or does at least have a weak limit point? Evidently, a necessary condition is that $(\mu_n(E))_{n\in\mathbb{N}}$ is bounded. Hence, without loss of generality, we will consider only sequences in $\mathcal{M}_{\leq 1}(E)$. However, this condition is not sufficient for the existence of weak limit points, as for example the sequence $(\delta_n)_{n\in\mathbb{N}}$ of probability measures on $\mathbb{R}$ does not have a weak limit point (although it converges vaguely to the zero measure). This example suggests that we also have to make sure that no mass "vanishes at infinity". The idea will be made precise by the notion of *tightness*.

We start this section by presenting as the main result Prohorov's theorem [136]. We give the proof first for the special case $E = \mathbb{R}$ and then come to a couple of applications. The full proof of the general case is deferred to the end of the section.

Definition 13.26 (Tightness) A family $\mathcal{F} \subset \mathcal{M}_f(E)$ is called *tight* if, for any $\varepsilon > 0$, there exists a compact set $K \subset E$ such that

$$\sup\big\{\mu(E \setminus K) : \mu \in \mathcal{F}\big\} < \varepsilon.$$

Remark 13.27 If E is Polish, then by Lemma 13.5, every singleton $\{\mu\} \subset \mathcal{M}_f(E)$ is tight and thus so is every finite family. ◊

Example 13.28

(i) If E is compact, then $\mathcal{M}_1(E)$ and $\mathcal{M}_{\leq 1}(E)$ are tight.
(ii) If $(X_i)_{i\in I}$ is an arbitrary family of random variables with

$$C := \sup\big\{\mathbf{E}\big[|X_i|\big] : i \in I\big\} < \infty,$$

then $\{\mathbf{P}_{X_i} : i \in I\}$ is tight. Indeed, for $\varepsilon > 0$ and $K = [-C/\varepsilon, C/\varepsilon]$, by Markov's inequality, $\mathbf{P}_{X_i}(\mathbb{R} \setminus K) = \mathbf{P}[|X_i| > C/\varepsilon] \leq \varepsilon$.

(iii) The family $(\delta_n)_{n\in\mathbb{N}}$ of probability measures on $\mathbb{R}$ is not tight.
(iv) The family $(\mathcal{U}_{[-n,n]})_{n\in\mathbb{N}}$ of uniform distributions on the intervals $[-n,n]$, regarded as measures on $\mathbb{R}$, is not tight. ◊

Recall that a family $\mathcal{F}$ of measures is called weakly relatively sequentially compact if every sequence in $\mathcal{F}$ has a weak limit point (in the closure of $\mathcal{F}$).

Theorem 13.29 (Prohorov's theorem (1956)) *Let (E,d) be a metric space and $\mathcal{F}\subset\mathcal{M}_{\le 1}(E)$. Then*:

(i) *$\mathcal{F}$ is tight $\implies$ $\mathcal{F}$ is weakly relatively sequentially compact.*
(ii) *If E is Polish, then also the converse holds*:

$$\mathcal{F} \text{ is tight} \quad\Longleftarrow\quad \mathcal{F} \text{ is weakly relatively sequentially compact.}$$

Corollary 13.30 *Let E be a compact metric space. Then the sets $\mathcal{M}_{\le 1}(E)$ and $\mathcal{M}_1(E)$ are weakly sequentially compact.*

Corollary 13.31 *If E is a locally compact separable metric space, then $\mathcal{M}_{\le 1}(E)$ is vaguely sequentially compact.*

Proof Let $x_1,x_2,\ldots$ be dense in E. As E is locally compact, for each $n\in\mathbb{N}$, there exists an open neighborhood $U_n\ni x_n$ whose closure $\overline{U}_n$ is compact. Hence, also $V_n:=\bigcup_{i=1}^n V_i$ is relatively compact for each $n\in\mathbb{N}$. This implies that $N(K):=\min\{m : K\subset V_m\}$ is finite for any compact $K\subset E$. Inductively define $W_1:=V_1$ and $W_{n+1}:=W_{N(\overline{V}_n)}$, $n\in\mathbb{N}$. Then W_n is open, $\overline{W}_n$ is compact, and $\overline{W}_n\subset W_{n+1}$ for all $n\in\mathbb{N}$. Furthermore, we have $W_n\uparrow E$.

Applying Prohorov's theorem (i.e., Corollary 13.30) to the measures $(\mu_k\mathbb{1}_{\overline{W}_n})_{k\in\mathbb{N}}$, for each $n\in\mathbb{N}$, we can choose a sequence $(k_l^n)_{l\in\mathbb{N}}$ and a measure $\tilde{\mu}_n:=\text{w-lim}_{l\to\infty}\mu_{k_l^n}\mathbb{1}_{\overline{W}_n}$ whose support lies in $\overline{W}_n$. We may assume that the sequences $(k_l^n)_{l\in\mathbb{N}}$ were chosen successively such that (k_l^{n+1}) is a subsequence of (k_l^n).

Note that we have $\tilde{\mu}_n(\overline{W}_n)\le\tilde{\mu}_{n+1}(\overline{W}_n)$, but equality does not hold in general.

For $f\in C_c(E)$, there exists an $n_0\in\mathbb{N}$ such that the support of f is contained in W_{n_0}. Hence, for $m\ge n\ge n_0$, we have

$$\begin{aligned}\int f\,d\tilde{\mu}_n &= \lim_{l\to\infty}\int f\mathbb{1}_{\overline{W}_n}\,d\mu_{k_l^n}\\ &= \lim_{l\to\infty}\int f\mathbb{1}_{\overline{W}_n}\,d\mu_{k_l^m}\\ &= \lim_{l\to\infty}\int f\mathbb{1}_{\overline{W}_m}\,d\mu_{k_l^m}=\int f\,d\tilde{\mu}_m\end{aligned}$$

and thus

$$\int f\,d\tilde{\mu}_n=\lim_{m\to\infty}\int f\,d\mu_{k_m^m}.$$

This implies that for any measurable relatively compact set $A \subset E$, we have

$$\tilde{\mu}_m(A) = \tilde{\mu}_{N(\overline{A})} \quad \text{for any } m \geq N(\overline{A}).$$

For any measurable set $A \subset E$, define

$$\mu(A) := \sup_{n\in\mathbb{N}} \sup_{m>n} \tilde{\mu}_m(A \cap W_n) = \sup_{n\in\mathbb{N}} \tilde{\mu}_{n+1}(A \cap W_n).$$

It is easy to check that μ is a lower semicontinuous content and is hence a measure (see Theorem 1.36). By construction, for any $f \in C_c(E)$, we infer

$$\int f\, d\mu = \lim_{n\to\infty} \int f\, d\mu_{k_n^n}.$$

Concluding, we have $\mu = \text{v-lim}_{n\to\infty} \mu_{k_n^n}$. □

Remark 13.32 The implication (ii) in Theorem 13.29 is less useful but a lot simpler to prove. Here we need that E is Polish since clearly every singleton is weakly compact but is tight only under additional assumptions; for example, if E is Polish (see Lemma 13.5). ◊

Proof of Theorem 13.29(ii) We start as in the proof of Lemma 13.5. Let $\{x_1, x_2, \ldots\} \subset E$ be dense. For $n \in \mathbb{N}$, define $A_{n,N} := \bigcup_{i=1}^N B_{1/n}(x_i)$. Then $A_{n,N} \uparrow E$ for $N \to \infty$ for all $n \in \mathbb{N}$. Let

$$\delta := \sup_{n\in\mathbb{N}} \inf_{N\in\mathbb{N}} \sup_{\mu\in\mathcal{F}} \mu\big(A_{n,N}^c\big).$$

Then there is an $n \in \mathbb{N}$ such that for any $N \in \mathbb{N}$, there is a $\mu_N \in \mathcal{F}$ with $\mu_N(A_{n,N}^c) \geq \delta/2$. As $\mathcal{F}$ is weakly relatively sequentially compact, $(\mu_N)_{N\in\mathbb{N}}$ has a weakly convergent subsequence $(\mu_{N_k})_{k\in\mathbb{N}}$ whose weak limit will be denoted by $\mu \in \mathcal{M}_{\leq 1}(E)$. By the Portemanteau theorem (Theorem 13.16(iv)), for any $N \in \mathbb{N}$,

$$\mu\big(A_{n,N}^c\big) \geq \liminf_{k\to\infty} \mu_{N_k}\big(A_{n,N}^c\big) \geq \liminf_{k\to\infty} \mu_{N_k}\big(A_{n,N_k}^c\big) \geq \delta/2.$$

On the other hand, $A_{n,N}^c \downarrow \emptyset$ for $N \to \infty$; hence $\mu(A_{n,N}^c) \overset{N\to\infty}{\longrightarrow} 0$. Thus $\delta = 0$.

Now fix $\varepsilon > 0$. By the above, for any $n \in \mathbb{N}$, we can choose an $N_n' \in \mathbb{N}$ such that $\mu(A_{n,N_n'}^c) < \varepsilon/2^n$ for all $\mu \in \mathcal{F}$. By construction, the set $A := \bigcap_{n=1}^\infty A_{n,N_n'}$ is totally bounded and hence relatively compact. Further, for every $\mu \in \mathcal{F}$,

$$\mu\big((\overline{A})^c\big) \leq \mu\big(A^c\big) \leq \sum_{n=1}^\infty \mu\big(A_{n,N_n'}^c\big) \leq \varepsilon.$$

Hence $\mathcal{F}$ is tight. □

The other implication in Prohorov's theorem is more difficult to prove, especially in the case of a general metric space. For this reason, we first give a proof only for the case $E = \mathbb{R}$ and come to applications before proving the difficult implication in the general situation.

The problem consists in finding a candidate for a weak limit point. For distributions on $\mathbb{R}$, the problem is equivalent to finding a weak limit point for a sequence of distribution functions. Here Helly's theorem is the tool. It is based on a diagonal sequence argument that will be recycled later in the proof of Prohorov's theorem in the general case.

Let

$$V = \{F : \mathbb{R} \to \mathbb{R} \text{ is right continuous, monotone increasing and bounded}\}$$

be the set of distribution functions of finite measures on $\mathbb{R}$.

Theorem 13.33 (Helly's theorem) *Let $(F_n)_{n\in\mathbb{N}}$ be a uniformly bounded sequence in V. Then there exists an $F \in V$ and a subsequence $(F_{n_k})_{k\in\mathbb{N}}$ with*

$$F_{n_k}(x) \overset{k\to\infty}{\longrightarrow} F(x) \text{ at all points of continuity of } F.$$

Proof We use a diagonal sequence argument. Choose an enumeration of the rational numbers $\mathbb{Q} = \{q_1, q_2, q_3, \ldots\}$. By the Bolzano–Weierstraß theorem, the sequence $(F_n(q_1))_{n\in\mathbb{N}}$ has a convergent subsequence $(F_{n_k^1}(q_1))_{k\in\mathbb{N}}$. Analogously, we find a subsequence $(n_k^2)_{k\in\mathbb{N}}$ of $(n_k^1)_{k\in\mathbb{N}}$ such that $(F_{n_k^2}(q_2))_{k\in\mathbb{N}}$ converges. Inductively, we obtain subsequences $(n_k^1) \supset (n_k^2) \supset (n_k^3) \supset \ldots$ such that $(F_{n_k^l}(q_l))_{k\in\mathbb{N}}$ converges for all $l \in \mathbb{N}$. Now define $n_k := n_k^k$. Then $(F_{n_k}(q))_{k\in\mathbb{N}}$ converges for all $q \in \mathbb{Q}$. Define $\widetilde{F}(q) = \lim_{k\to\infty} F_{n_k}(q)$ and

$$F(x) = \inf\big\{\widetilde{F}(q) : q \in \mathbb{Q} \text{ with } q > x\big\}.$$

As $\widetilde{F}$ is monotone increasing, F is right continuous and monotone increasing.

If F is continuous at x, then for every $\varepsilon > 0$, there exist numbers $q^-, q^+ \in \mathbb{Q}$, $q^- < x < q^+$ with $\widetilde{F}(q^-) \geq F(x) - \varepsilon$ and $\widetilde{F}(q^+) \leq F(x) + \varepsilon$. By construction,

$$\limsup_{k\to\infty} F_{n_k}(x) \leq \lim_{k\to\infty} F_{n_k}\big(q^+\big) = \widetilde{F}\big(q^+\big) \leq F(x) + \varepsilon.$$

Hence $\limsup_{k\to\infty} F_{n_k}(x) \leq F(x)$. A similar argument for q^- yields $\liminf_{k\to\infty} F_{n_k}(x) \geq F(x)$. □

Proof of Theorem 13.29(i) for the case $E = \mathbb{R}$ Assume $\mathcal{F}$ is tight and $(\mu_n)_{n\in\mathbb{N}}$ is a sequence in $\mathcal{F}$ with distribution functions $F_n : x \mapsto \mu_N((-\infty, x])$. By Helly's theorem, there is a monotone right continuous function $F : \mathbb{R} \to [0, 1]$ and a subsequence $(F_{n_k})_{k\in\mathbb{N}}$ of $(F_n)_{n\in\mathbb{N}}$ with $F_{n_k}(x) \overset{k\to\infty}{\longrightarrow} F(x)$ at all points of continuity x of F. By Theorem 13.23, it is enough to show that $F(\infty) \geq \limsup_{k\to\infty} F_{n_k}(\infty)$.

As $\mathcal{F}$ is tight, for every $\varepsilon > 0$, there is a $K < \infty$ with $F_n(\infty) - F_n(x) < \varepsilon$ for all $n \in \mathbb{N}$ and $x > K$. If $x > K$ is a point of continuity of F, then $\limsup_{k\to\infty} F_{n_k}(\infty) \leq \limsup_{k\to\infty} F_{n_k}(x) + \varepsilon = F(x) + \varepsilon \leq F(\infty) + \varepsilon$. □

We come to a first application of Prohorov's theorem. The full strength of that theorem will become manifest when suitable separating classes of functions are at our disposal. We come back to this point in more detail in Chapter 15.

Theorem 13.34 *Let E be Polish and let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_{\leq 1}(E)$. Then the following are equivalent.*

(i) $\mu = \text{w-lim}_{n\to\infty} \mu_n$.
(ii) $(\mu_n)_{n\in\mathbb{N}}$ *is tight, and there is a separating family $\mathcal{C} \subset C_b(E)$ such that*

$$\int f\, d\mu = \lim_{n\to\infty} \int f\, d\mu_n \quad \textit{for all } f \in \mathcal{C}. \tag{13.8}$$

Proof "(i)$\Longrightarrow$(ii)" By the simple implication in Prohorov's theorem (Theorem 13.29(ii)), weak convergence implies tightness.

"(ii)$\Longrightarrow$(i)" Let $(\mu_n)_{n\in\mathbb{N}}$ be tight and let $\mathcal{C} \subset C_b(E)$ be a separating class with (13.8). Assume that $(\mu_n)_{n\in\mathbb{N}}$ does not converge weakly to μ. Then there are $\varepsilon > 0$, $f \in C_b(E)$ and $(n_k)_{k\in\mathbb{N}}$ with $n_k \uparrow \infty$ and such that

$$\left| \int f\, d\mu_{n_k} - \int f\, d\mu \right| > \varepsilon \quad \text{for all } k \in \mathbb{N}. \tag{13.9}$$

By Prohorov's theorem, there exists a $\nu \in \mathcal{M}_{\leq 1}(E)$ and a subsequence $(n'_k)_{k\in\mathbb{N}}$ of $(n_k)_{k\in\mathbb{N}}$ with $\mu_{n'_k} \to \nu$ weakly. Due to (13.9), we have $|\int f\, d\mu - \int f\, d\nu| \geq \varepsilon$; hence $\mu \neq \nu$. On the other hand,

$$\int h\, d\mu = \lim_{k\to\infty} \int h\, d\mu_{n'_k} = \int h\, d\nu \quad \text{for all } h \in \mathcal{C};$$

hence $\mu = \nu$. This contradicts the assumption and thus (i) holds. □

We want to shed some more light on the connection between weak and vague convergence.

Theorem 13.35 *Let E be a locally compact Polish space and let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_f(E)$. Then the following are equivalent.*

(i) $\mu = \text{w-lim}_{n\to\infty} \mu_n$.
(ii) $\mu = \text{v-lim}_{n\to\infty} \mu_n$ *and* $\mu(E) = \lim_{n\to\infty} \mu_n(E)$.
(iii) $\mu = \text{v-lim}_{n\to\infty} \mu_n$ *and* $\mu(E) \geq \limsup_{n\to\infty} \mu_n(E)$.
(iv) $\mu = \text{v-lim}_{n\to\infty} \mu_n$ *and* $\{\mu_n, n \in \mathbb{N}\}$ *is tight.*

Proof "(i) $\Longleftrightarrow$ (ii) $\Longleftrightarrow$ (iii)" This follows by the Portemanteau theorem.

"(ii) $\Longrightarrow$(iv)" It is enough to show that for any $\varepsilon > 0$, there is a compact set $K \subset E$ with $\limsup_{n\to\infty} \mu_n(E \setminus K) \leq \varepsilon$. As μ is regular (Theorem 13.6), there is a compact set $L \subset E$ with $\mu(E \setminus L) < \varepsilon$. Since E is locally compact, there exists a compact set $K \subset E$ with $K^\circ \supset L$ and a $\rho_{L,K} \in C_c(E)$ with $\mathbb{1}_L \le \rho_{L,K}(x) \le \mathbb{1}_K$. Therefore,

$$\limsup_{n\to\infty} \mu_n(E \setminus K) \le \limsup_{n\to\infty}\Big(\mu_n(E) - \int \rho_{L,K}\, d\mu_n\Big)$$
$$= \mu(E) - \int \rho_{L,K}\, d\mu \le \mu(E \setminus L) < \varepsilon.$$

"(iv) $\Longrightarrow$(i)" Let $L \subset E$ be compact with $\mu_n(E \setminus L) \le 1$ for all $n \in \mathbb{N}$. Let $\rho \in C_c(E)$ with $\rho \ge \mathbb{1}_L$. Since $\int \rho\, d\mu_n$ converges by assumption, we thus have

$$\sup_{n\in\mathbb{N}} \mu_n(E) \le 1 + \sup_{n\in\mathbb{N}} \mu_n(L) \le 1 + \sup_{n\in\mathbb{N}} \int \rho\, d\mu_n < \infty.$$

Hence also

$$C := \max\big(\mu(E), \sup\{\mu_n(E) : n \in \mathbb{N}\}\big) < \infty,$$

and we can pass to μ/C and μ_n/C. Thus, without loss of generality assume that all measures are in $\mathcal{M}_{\le 1}(E)$. As $C_c(E)$ is a separating class for $\mathcal{M}_{\le 1}(E)$ (see Theorem 13.11), (i) follows by Theorem 13.34. □

Proof of Prohorov's Theorem, Part (i), General Case There are two main routes for proving Prohorov's theorem in the general situation. One possibility is to show the claim first for measures on $\mathbb{R}^d$. (We have done this already for $d = 1$, see Exercise 13.3.4 for $d \ge 2$.) In a second step, the statement is lifted to sequence spaces $\mathbb{R}^{\mathbb{N}}$. Finally, in the third step, an embedding of E into $\mathbb{R}^{\mathbb{N}}$ is constructed. For a detailed description, see [12] or [83].

Here we follow the alternative route as described in [13] (and later [14]) or [44]. The main point of this proof consists in finding a candidate for a weak limit point for the family $\mathcal{F}$. This candidate will be constructed first as a content on a countable class of sets. From this an outer measure will be derived. Finally, we show that closed sets are measurable with respect to this outer measure. As you see, the argument follows a pattern similar to the proof of Carathéodory's theorem.

Let (E, d) be a metric space and let $\mathcal{F} \subset \mathcal{M}_{\le 1}(E)$ be tight. Then there exists an increasing sequence $K_1 \subset K_2 \subset K_3 \subset \ldots$ of compact sets in E such that $\mu(K_n^c) < \frac{1}{n}$ for all $\mu \in \mathcal{F}$ and all $n \in \mathbb{N}$. Define $E' := \bigcup_{n=1}^{\infty} K_n$. Then E' is a σ-compact metric space and therefore in particular, separable. By construction, $\mu(E \setminus E') = 0$ for all $\mu \in \mathcal{F}$. Thus, any μ can be regarded as a measure on E'. Without loss of generality, we may hence assume that E is σ-compact and thus separable. Hence there exists a countable base $\mathcal{U}$ of the topology $\tau|_E$ on E; that is, a countable set $\mathcal{U}$ of open sets such that $A = \bigcup_{U\in\mathcal{U}, U\subset A} U$ for any open $A \subset E$. Define

$$\mathcal{C}' := \{\overline{U} \cap K_n : U \in \mathcal{U}, n \in \mathbb{N}\}$$

and

$$\mathcal{C} := \left\{ \bigcup_{n=1}^{N} C_n : N \in \mathbb{N} \text{ and } C_1, \ldots, C_N \in \mathcal{C}' \right\}.$$

Clearly, $\mathcal{C}$ is a countable set of compact sets in E, and $\mathcal{C}$ is stable under formation of unions. Any K_n possesses a finite covering with sets from $\mathcal{U}$; hence $K_n \in \mathcal{C}$.

Now let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{F}$. By virtue of the diagonal sequence argument (see the proof of Helly's theorem, Theorem 13.33), we can find a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ such that for all $C \in \mathcal{C}$, there exists the limit

$$\alpha(C) := \lim_{k\to\infty} \mu_{n_k}(C). \tag{13.10}$$

Assume that we can show that there is a measure μ on the Borel σ-algebra $\mathcal{E}$ of E such that

$$\mu(A) = \sup\big\{\alpha(C) : C \in \mathcal{C} \text{ with } C \subset A\big\} \quad \text{for all } A \subset E \text{ open.} \tag{13.11}$$

Then

$$\begin{aligned}\mu(E) &\ge \sup_{n\in\mathbb{N}} \alpha(K_n) = \sup_{n\in\mathbb{N}} \lim_{k\to\infty} \mu_{n_k}(K_n)\\ &\ge \sup_{n\in\mathbb{N}} \limsup_{k\to\infty} \left(\mu_{n_k}(E) - \frac{1}{n}\right)\\ &= \limsup_{k\to\infty} \mu_{n_k}(E).\end{aligned}$$

Furthermore, for open A and for $C \in \mathcal{C}$ with $C \subset A$,

$$\alpha(C) = \lim_{k\to\infty} \mu_{n_k}(C) \le \liminf_{k\to\infty} \mu_{n_k}(A),$$

hence $\mu(A) \le \liminf_{k\to\infty} \mu_{n_k}(A)$. By the Portemanteau theorem (Theorem 13.16), $\mu = \text{w-lim}_{k\to\infty} \mu_{n_k}$; hence $\mathcal{F}$ is recognized as weakly relatively sequentially compact. It remains to show that there exists a measure μ on $(E, \mathcal{E})$ that satisfies (13.11).

Clearly, the set function α on $\mathcal{C}$ is monotone, additive and subadditive:

$$\begin{aligned}\alpha(C_1) &\le \alpha(C_2), && \text{if } C_1 \subset C_2,\\ \alpha(C_1 \cup C_2) &= \alpha(C_1) + \alpha(C_2), && \text{if } C_1 \cap C_2 = \emptyset,\\ \alpha(C_1 \cup C_2) &\le \alpha(C_1) + \alpha(C_2).\end{aligned} \tag{13.12}$$

We define

$$\beta(A) := \sup\big\{\alpha(C) : C \in \mathcal{C} \text{ with } C \subset A\big\} \quad \text{for } A \subset E \text{ open}$$

and

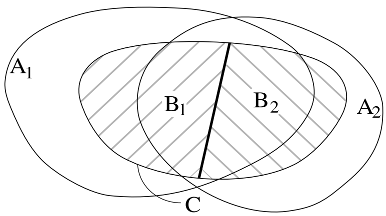
$$\mu^*(G) := \inf\big\{\beta(A) : A \supset G \text{ is open}\big\} \quad \text{for } G \in 2^E.$$

Manifestly, $\beta(A) = \mu^*(A)$ for any open A. It is enough to show (Steps 1–3 below) that μ^* is an outer measure (see Definition 1.46) and that (Step 4) the σ-algebra of μ^*-measurable sets (see Definition 1.48 and Lemma 1.52) contains the closed sets and thus $\mathcal{E}$. Indeed, Lemma 1.52 would then imply that μ^* is a measure on the σ-algebra of μ^*-measurable sets and the restricted measure $\mu := \mu^*|_{\mathcal{E}}$ fulfills $\mu(A) = \mu^*(A) = \beta(A)$ for all open A. Hence Eq. (13.11) holds.

Evidently, $\mu^*(\emptyset) = 0$ and μ^* is monotone. In order to show that μ^* is an outer measure, it only remains to check that μ^* is σ-subadditive.

Step 1 (Finite subadditivity of β) Let $A_1, A_2 \subset E$ be open and let $C \in \mathcal{C}$ with $C \subset A_1 \cup A_2$. Let $n \in N$ with $C \subset K_n$. Define two sets

$$B_1 := \{x \in C : d(x, A_1^c) \ge d(x, A_2^c)\},$$
$$B_2 := \{x \in C : d(x, A_1^c) \le d(x, A_2^c)\}.$$



Evidently, $B_1 \subset A_1$ and $B_2 \subset A_2$. As $x \mapsto d(x, A_i^c)$ is continuous for $i = 1, 2$, the closed subsets B_1 and B_2 of C are compact. Hence $d(B_1, A_1^c) > 0$. Thus there exists an open set D_1 with $B_1 \subset D_1 \subset \overline{D}_1 \subset A_1$. (One could choose D_1 as the union of the sets of a finite covering of B_1 with balls of radius $d(B_1, A_1^c)/2$. These balls, as well as their closures, are subsets of A_1.) Let $\mathcal{U}_{D_1} := \{U \in \mathcal{U} : U \subset D_1\}$. Then $B_1 \subset D_1 = \bigcup_{U \in \mathcal{U}_{D_1}} U$. Now choose a finite subcovering $\{U_1, \ldots, U_N\} \subset \mathcal{U}_{D_1}$ of B_1 and define $C_1 := \bigcup_{i=1}^N \overline{U}_i \cap K_n$. Then $B_1 \subset C_1 \subset A_1$ and $C_1 \in \mathcal{C}$. Similarly, choose $C_2 \in \mathcal{C}$ with $B_2 \subset C_2 \subset A_2$. Thus

$$\alpha(C) \le \alpha(C_1 \cup C_2) \le \alpha(C_1) + \alpha(C_2) \le \beta(A_1) + \beta(A_2).$$

Hence also

$$\begin{aligned}\beta(A_1 \cup A_2) &= \sup\{\alpha(C) : C \in \mathcal{C} \text{ with } C \subset A_1 \cup A_2\} \\ &\le \beta(A_1) + \beta(A_2).\end{aligned}$$

Step 2 (σ-subadditivity of β) Let $A_1, A_2, \ldots$ be open sets and let $C \in \mathcal{C}$ with $C \subset \bigcup_{i=1}^\infty A_i$. As C is compact, there exists an $n \in \mathbb{N}$ with $C \subset \bigcup_{i=1}^n A_i$. As shown above, β is subadditive; thus

$$\alpha(C) \le \beta\left(\bigcup_{i=1}^n A_i\right) \le \sum_{i=1}^\infty \beta(A_i).$$

Taking the supremum over such C yields

$$\beta\left(\bigcup_{i=1}^\infty A_i\right) = \sup\left\{\alpha(C) : C \in \mathcal{C} \text{ with } C \subset \bigcup_{i=1}^\infty A_i\right\} \le \sum_{i=1}^\infty \beta(A_i).$$

Step 3 (σ-subadditivity of μ^*) Let $G_1, G_2, \ldots \in 2^E$. Let $\varepsilon > 0$. For any $n \in \mathbb{N}$ choose an open set $A_n \supset G_n$ with $\beta(A_n) < \mu^*(G_n) + \varepsilon/2^n$. By the σ-subadditivity of β,

$$\mu^*\left(\bigcup_{n=1}^{\infty} G_n\right) \le \beta\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \beta(A_n) \le \varepsilon + \sum_{n=1}^{\infty} \mu^*(G_n).$$

Letting $\varepsilon \downarrow 0$ yields $\mu^*(\bigcup_{n=1}^{\infty} G_n) \le \sum_{n=1}^{\infty} \mu^*(G_n)$. Hence μ^* is an outer measure.

Step 4 (Closed sets are μ^*-measurable) By Lemma 1.49, a set $B \subset E$ is μ^*-measurable if and only if

$$\mu^*(B \cap G) + \mu^*\big(B^c \cap G\big) \le \mu^*(G) \quad \text{for all } G \in 2^E.$$

Taking the infimum over all open sets $A \supset G$, it is enough to show that for every open B and every open $A \subset E$,

$$\mu^*(B \cap A) + \mu^*\big(B^c \cap A\big) \le \beta(A). \tag{13.13}$$

Let $\varepsilon > 0$. Choose $C_1 \in \mathcal{C}$ with $C_1 \subset A \cap B^c$ and $\alpha(C_1) > \beta(A \cap B^c) - \varepsilon$. Further, let $C_2 \in \mathcal{C}$ with $C_2 \subset A \cap C_1^c$ and $\alpha(C_2) > \beta(A \cap C_1^c) - \varepsilon$. Since $C_1 \cap C_2 = \emptyset$ and $C_1 \cup C_2 \subset A$, we get

$$\begin{aligned}\beta(A) &\ge \alpha(C_1 \cup C_2)\\ &= \alpha(C_1) + \alpha(C_2) \ge \beta\big(A \cap B^c\big) + \beta\big(A \cap C_1^c\big) - 2\varepsilon\\ &\ge \mu^*\big(A \cap B^c\big) + \mu^*(A \cap B) - 2\varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$, we get (13.13). This completes the proof of Prohorov's theorem. □

Exercise 13.3.1 Show that a family $\mathcal{F} \subset \mathcal{M}_f(\mathbb{R})$ is tight if and only if there exists a measurable map $f : \mathbb{R} \to [0, \infty)$ such that $f(x) \to \infty$ for $|x| \to \infty$ and $\sup_{\mu \in \mathcal{F}} \int f \, d\mu < \infty$.

Exercise 13.3.2 Let $L \subset \mathbb{R} \times (0, \infty)$ and let $\mathcal{F} = \{\mathcal{N}_{\mu,\sigma^2} : (\mu, \sigma^2) \in L\}$ be a family of normal distributions with parameters in L. Show that $\mathcal{F}$ is tight if and only if L is bounded.

Exercise 13.3.3 If P is a probability measure on $[0, \infty)$ with $m_P := \int x \, P(dx) \in (0, \infty)$, then we define the *size-biased distribution* $\widehat{P}$ on $[0, \infty)$ by

$$\widehat{P}(A) = m_P^{-1} \int_A x \, P(dx). \tag{13.14}$$

Now let $(X_i)_{i \in I}$ be a family of random variables on $[0, \infty)$ with $\mathbf{E}[X_i] = 1$. Show that $(\widehat{\mathbf{P}_{X_i}})_{i \in I}$ is tight if and only if $(X_i)_{i \in I}$ is uniformly integrable.

Exercise 13.3.4 (Helly's theorem in $\mathbb{R}^d$) Let $x = (x^1, \ldots, x^d) \in \mathbb{R}^d$ and $y = (y^1, \ldots, y^d) \in \mathbb{R}^d$. Recall the notation $x \le y$ if $x^i \le y^i$ for all $i = 1, \ldots, d$. A map $F : \mathbb{R}^d \to \mathbb{R}$ is called monotone increasing if $F(x) \le F(y)$ whenever $x \le y$. F is called right continuous if $F(x) = \lim_{n\to\infty} F(x_n)$ for all $x \in \mathbb{R}^d$ and every sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^d$ with $x_1 \ge x_2 \ge x_3 \ge \ldots$ and $x = \lim_{n\to\infty} x_n$. By V_d denote the set of monotone increasing, bounded right continuous functions on $\mathbb{R}^d$.

(i) Show the validity of Helly's theorem with V replaced by V_d.
(ii) Conclude that Prohorov's theorem holds for $E = \mathbb{R}^d$.

13.4 Application: A Fresh Look at de Finetti's Theorem

(After an idea of Götz Kersting.) Let E be a Polish space and let $X_1, X_2, \ldots$ be an exchangeable sequence of random variables with values in E. As an alternative to the backwards martingale argument of Section 12.3, here we give a different proof of de Finetti's theorem (Theorem 12.26). Recall that de Finetti's theorem states that there exists a random probability measure Ξ on E such that, given Ξ, the random variables $X_1, X_2, \ldots$ are independent and Ξ-distributed. For $x = (x_1, x_2, \ldots) \in E^{\mathbb{N}}$, let $\xi_n(x) := \frac{1}{n}\sum_{l=1}^n \delta_{x_l}$ be the empirical distribution of $x_1, \ldots, x_n$. Let

$$\mu_{n,k}(x) := \xi_n(x)^{\otimes k} = n^{-k} \sum_{i_1,\ldots,i_k=1}^{n} \delta_{(x_{i_1},\ldots,x_{i_k})}$$

be the distribution on E^k that describes k-fold independent sampling *with replacement* (respecting the order) from $(x_1, \ldots, x_n)$. Let

$$\nu_{n,k}(x) := \frac{(n-k)!}{n!} \sum_{\substack{i_1,\ldots,i_k=1\\ \#\{i_1,\ldots,i_k\}=k}}^{n} \delta_{(x_{i_1},\ldots,x_{i_k})}$$

be the distribution on E^k that describes k-fold independent sampling *without replacement* (respecting the order) from $(x_1, \ldots, x_n)$. For all $x \in E^{\mathbb{N}}$,

$$\left\| \mu_{n,k}(x) - \nu_{n,k}(x) \right\|_{TV} \le R_{n,k} := \frac{k(k-1)}{n}.$$

Indeed, the probability $p_{n,k}$ that we do not see any ball twice when drawing k balls (with replacement) from n different balls is

$$p_{n,k} = \prod_{l=1}^{k-1} (1 - l/n)$$

and thus $R_{n,k} \ge 2(1 - p_{n,k})$. We therefore obtain the rather intuitive statement that as $n \to \infty$ the distributions of k-samples with replacement and without replacement, respectively, become the same:

$$\lim_{n\to\infty}\sup_{x\in E^{\mathbb{N}}}\big\|\mu_{n,k}(x)-\nu_{n,k}(x)\big\|_{TV}=0.$$

Now let $f_1,\ldots,f_k\in C_b(E)$ and $F(x_1,\ldots,x_k):=f_1(x_1)\ldots f_k(x_k)$. As the sequence $X_1,X_2,\ldots$ is exchangeable, for any choice of pairwise distinct numbers $1\le i_1,\ldots,i_k\le n$,

$$\mathbf{E}\big[F(X_1,\ldots,X_k)\big]=\mathbf{E}\big[F(X_{i_1},\ldots,X_{i_k})\big].$$

Averaging over all choices $i_1,\ldots,i_k$, we get

$$\mathbf{E}\big[f_1(X_1)\ldots f_k(X_k)\big]=\mathbf{E}\big[F(X_1,\ldots,X_k)\big]=\mathbf{E}\bigg[\int F\,d\nu_{n,k}(X)\bigg].$$

Hence

$$\begin{aligned}&\bigg|\mathbf{E}\big[f_1(X_1)\ldots f_k(X_k)\big]-\mathbf{E}\bigg[\int f_1\,d\xi_n(X)\ldots\int f_k\,d\xi_n(X)\bigg]\bigg|\\&=\bigg|\mathbf{E}\bigg[\int F\,d\nu_{n,k}(X)\bigg]-\mathbf{E}\bigg[\int F\,d\mu_{n,k}(X)\bigg]\bigg|\\&\le\|F\|_\infty R_{n,k}\overset{n\to\infty}{\longrightarrow}0.\end{aligned}$$

We will exploit the following criterion for tightness of subsets of $\mathcal{M}_1(\mathcal{M}_1(E))$.

Exercise 13.4.1 Show that a subset $\mathcal{K}\subset\mathcal{M}_1(\mathcal{M}_1(E))$ is tight if and only if, for any $\varepsilon>0$, there exists a compact set $K\subset E$ with the property

$$\widetilde{\mu}\big(\big\{\mu\in\mathcal{M}_1(E):\mu\big(K^c\big)>\varepsilon\big\}\big)<\varepsilon\quad\text{for all }\widetilde{\mu}\in\mathcal{K}.$$

Since E is Polish, $\mathbf{P}_{X_1}$ is tight. Hence, for any $\varepsilon>0$, there exists a compact set $K\subset E$ with $\mathbf{P}[X_1\in K^c]<\varepsilon^2$. Therefore,

$$\mathbf{P}\big[\xi_n(X)\big(K^c\big)>\varepsilon\big]\le\varepsilon^{-1}\mathbf{E}\big[\xi_n(X)\big(K^c\big)\big]=\varepsilon^{-1}\mathbf{P}\big[X_1\in K^c\big]\le\varepsilon.$$

Hence the family $(\mathbf{P}_{\xi_n(X)})_{n\in\mathbb{N}}$ is tight. Let Ξ_∞ be a random variable (with values in $\mathcal{M}_1(E)$) such that $\mathbf{P}_{\Xi_\infty}=\text{w-lim}_{l\to\infty}\mathbf{P}_{\xi_{n_l}(X)}$ for a suitable subsequence $(n_l)_{l\in\mathbb{N}}$. The map $\xi\mapsto\int F\,d\xi=\int f_1\,d\xi\ldots\int f_k\,d\xi$ is bounded and (as a product of continuous maps) is continuous with respect to the topology of weak convergence on $\mathcal{M}_1(E)$; hence it is in $C_b(\mathcal{M}_1(E))$. Thus

$$\begin{aligned}\mathbf{E}\bigg[\int F\,d\Xi_\infty^{\otimes k}\bigg]&=\lim_{l\to\infty}\mathbf{E}\bigg[\int f_1\,d\xi_{n_l}(X)\ldots\int f_k\,d\xi_{n_l}(X)\bigg]\\&=\mathbf{E}\big[f_1(X_1)\ldots f_k(X_k)\big].\end{aligned}$$

Note that the limit does not depend on the choice of the subsequence and is thus unique. Summarising, we have

$$\mathbf{E}\big[f_1(X_1)\dots f_k(X_k)\big]=\mathbf{E}\Big[\int f_1\,d\Xi_\infty\dots\int f_k\,d\Xi_\infty\Big].$$

Since the distribution of $(X_1,\dots,X_k)$ is uniquely determined by integrals of the above type, we conclude that $\mathbf{P}_{(X_1,\dots,X_k)}=\mathbf{P}_{\Xi_\infty^{\otimes k}}$. In other words, $(X_1,\dots,X_k)\stackrel{\mathcal{D}}{=}(Y_1,\dots,Y_k)$, where, given Ξ_∞, the random variables $Y_1,\dots,Y_k$ are independent with distribution Ξ_∞.

Exercise 13.4.2 Show that a family $(X_n)_{n\in\mathbb{N}}$ of random variables is exchangeable if and only if, for every choice of natural numbers $1\le n_1<n_2<n_3<\dots$, we have

$$(X_1,X_2,\dots)\stackrel{\mathcal{D}}{=}(X_{n_1},X_{n_2},\dots).$$

Warning: One of the implications is rather difficult to show.