3 Measure and Integration

Man is the measure of all things.
— Pythagoras

Lebesgue is the measure of almost all things.
— Anonymous

3.1 Motivation

We shall give a few reasons why it is worth bothering with measure theory and the Lebesgue integral. To this end, we stress the importance of measure theory in three different areas.

3.1.1 We want a powerful integral

At the end of the previous chapter we encountered a neat application of Banach's fixed point theorem to solve ordinary differential equations. An essential ingredient in the argument was the observation in Lemma 2.77 that the operation of differentiation could be replaced by integration. Note that differentiation is an operation that destroys regularity, while integration yields further regularity. It is a consequence of the fundamental theorem of calculus that the indefinite integral of a continuous function is a continuously differentiable function. So far we used the elementary notion of the *Riemann integral*. Let us quickly recall the definition of the Riemann integral on a bounded interval.

Definition 3.1. Let $[a,b]$ with $-\infty < a < b < \infty$ be a compact interval. A **partition** of $[a,b]$ is a finite sequence $\pi := (t_0, \dots, t_N)$ such that $a = t_0 < t_1 < \cdots < t_N = b$. The **mesh size** of π is $|\pi| := \max_{1 \leq k \leq N} |t_k - t_{k-1}|$. Given a partition π of $[a,b]$, an **associated vector of sample points** (frequently also called **tags**) is a vector $\xi = (\xi_1, \dots, \xi_N)$ such that $\xi_k \in [t_{k-1}, t_k]$. Given a function $f\colon [a,b] \to \mathbb{R}$ and a tagged partition (π, ξ) of $[a, b]$, the **Riemann sum** $S(f, \pi, \xi)$ is defined by

$$S(f, \pi, \xi) := \sum_{k=1}^{N} f(\xi_k)(t_k - t_{k-1}).$$

We note that there are many different notions of integrals (Cauchy, Riemann, Choquet, Lebesgue, Stieltjes, Daniell, Henstock–Kurzweil, Itô, to name a few). Each notion has its specific advantages and use cases. A common theme is that integration is linear and generalises summation.

A function f is called **Riemann integrable** if there exists a number $A \in \mathbb{R}$ such that for every sequence of tagged partitions (π_n, ξ_n) with $|\pi_n| \to 0$ one has $S(f, \pi_n, \xi_n) \to A$ as $n \to \infty$. In this case, A is called the **Riemann integral of f over $[a, b]$**. Notation: $\text{R-}\int_a^b f(t)\,\mathrm{d}t := A$.

Exercise 3.2. Show that a Riemann integrable function must be bounded, i.e., if $f\colon [a,b] \to \mathbb{R}$ is Riemann integrable then $f \in \mathcal{F}_{\mathrm{b}}[a,b]$. Moreover, if $\alpha \in \mathbb{R}$ and $f, g\colon [a,b] \to \mathbb{R}$ are Riemann integrable, show that $\alpha f + g$ is Riemann integrable with

$$\text{R-}\int_a^b (\alpha f(t) + g(t))\,\mathrm{d}t = \alpha\text{R-}\int_a^b f(t)\,\mathrm{d}t + \text{R-}\int_a^b g(t)\,\mathrm{d}t.$$

It is not hard to see that every continuous function $f\colon [a,b] \to \mathbb{R}$ is Riemann integrable. Moreover, one has the following result.

Theorem 3.3 (Fundamental theorem of calculus)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous. Then $F\colon [a,b] \to \mathbb{R}$ defined by*

$$F(t) := \text{R-}\int_a^t f(s)\,\mathrm{d}s$$

*is differentiable such that $F' = f$, i.e., F is an **antiderivative** of f. Moreover, if F is any antiderivative of f, then*

$$\text{R-}\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a).$$

This theorem is remarkable in that it allows to conveniently compute values of definite Riemann integrals. The following exercise is a direct application for Riemann sums.

Exercise 3.4. Calculate the value of the limit

$$\lim_{N \to \infty} \sum_{k=1}^{N} \frac{N}{k^2 + N^2}$$

by using the Riemann integral.

However, the Riemann integral has some serious shortcomings for the applications that we have in mind.

Clearly, it is inconvenient that merely bounded functions can be Riemann integrable and that the Riemann integral is only defined for compact intervals. For example, to integrate a bounded function f over the whole of $\mathbb{R}$, one needs to consider expressions like

$$\lim_{A\to-\infty}\lim_{B\to\infty} \mathsf{R}\text{-}\int_A^B f(t)\,\mathrm{d}t$$

and, provided the limit exists, call it the improper Riemann integral of f over $\mathbb{R}$. While one can similarly use improper Riemann integrals to integrate mildly unbounded functions, this is a siginificant inconvenience.

Note that $\lim_{A\to\infty} \mathsf{R}\text{-}\int_{-A}^{A} t\,\mathrm{d}t = 0$, while the doubly infinite improper Riemann integral of $f(t) = t$ does not exist.

The second shortcoming is more fundamental. Suppose we have a sequence (f_n) in $\mathrm{C}[0,1]$ that converges pointwise to a function $f \in \mathcal{F}_\mathrm{b}[0,1]$. We wish to ensure that f is Riemann integrable and that $\mathsf{R}\text{-}\int_0^1 f_n$ converges to $\mathsf{R}\text{-}\int_0^1 f$ making only weak additional assumptions. While it is sufficient to assume that (f_n) converges uniformly to f, such a result is insatisfactory as uniform convergence is a very strong form of convergence and its assumption is therefore extremely restrictive. The Riemann integral is not well-suited for obtaining better convergence results, one reason being that not enough functions are Riemann integrable.

Example 3.5 (Dirichlet function)**.** Let (q_n) be an enumeration of $\mathbb{Q}\cap[0,1]$. Let $f\colon [0,1]\to\mathbb{R}$ be defined by $f(t) := \mathbb{1}_{\mathbb{Q}\cap[0,1]}(t)$. Then f is not Riemann integrable. However, f is the pointwise limit of the Riemann integrable functions $f_n\colon [0,1]\to\mathbb{R}$ given by $f(t) := \mathbb{1}_{\{q_1,\dots,q_n\}}(t)$.

*Let M be a set and $A \subset M$. Then the function $\mathbb{1}_A\colon M\to\mathbb{R}$ defined by $\mathbb{1}_A(x) = 1$ if $x \in M$ and $\mathbb{1}_A(x) = 0$ if $x \notin M$ is called the **indicator function** of A in M.*

So we would like to be able to integrate much larger classes of functions, and obtain convergence results for the integrals under less restrictive assumptions.

Exercise 3.6. Show that the map

$$\|f\|_{\mathscr{L}^1} := \int_0^1 |f(t)|\,\mathrm{d}t$$

defines a norm on $\mathrm{C}[0,1]$. Construct a sequence (f_n) of functions in $C[0,1]$ that is bounded with respect to $\|\cdot\|_\infty$ such that (f_n) is a Cauchy sequence with respect to $\|\cdot\|_{\mathscr{L}^1}$ that converges pointwise to a non-Riemann integrable indicator function. Deduce that $(C[0,1], \|\cdot\|_{\mathscr{L}^1})$ is not complete.

So the norm $\|f\|_{\mathscr{L}^1}$ on the Riemann integrable functions does not yield a complete space. We wish to consider a notion of the integral that makes limit functions as in Exercise 3.6 integrable. The overall aim is to obtain complete normed spaces of suitably integrable functions.

Finally, we wish to be able to integrate over more general sets than compact intervals of the real line. This is both important for geometry (say we would like to integrate functions defined on a 2-dimensional surface) as it is for probabilty theory (where we wish to determine expected values of random variables on general probability spaces).

We shall see that the notion of the *Lebesgue integral* addresses all the aforementioned shortcomings of the Rieman integral. It allows us to integrate more functions, provides us with powerful and easy to check convergence results and is capable of integrating in a much more general setting.

We close this first subsection by hinting on how the Lebesgue integral, which we shall properly introduce in the following sections, compares to the Riemann integral for functions on the compact interval $[0, 1]$. While for the Riemann integral one directly partitions the inteval $[0, 1]$, the Lebesgue integral works by partitioning the range of f. One then looks at the preimages under f of these parts of the range of f, which are possibly rather irregular subsets of $[0, 1]$. For the example of the Dirichlet function, one obtains that the preimage of $\{1\}$ is $\mathbb{Q} \cap [0, 1]$, while the preimage of $\{0\}$ is $[0, 1] \setminus \mathbb{Q}$. Now suppose that $\lambda(A)$ is some hypothetical function that assigns a 'measure' to subsets $A \subset [0, 1]$. Then the integral of the Dirichlet function should equal $0 \cdot \lambda([0, 1] \setminus \mathbb{Q}) + 1 \cdot \lambda([0, 1] \cap \mathbb{Q})$. Intuitively $\lambda([0, 1])$ should be 1, and since a single point $q \in [0, 1]$ is contained in an arbitarily short interval $\lambda(\{q\})$ should be 0. As $\mathbb{Q} \cap [0, 1]$ is merely a countable union of such one point sets, it should have 'measure' 0 (also since it is much 'smaller' than $[0, 1] \setminus \mathbb{Q}$). So we expect that $\lambda(\mathbb{Q} \cap [0, 1]) = 0$ and $\lambda([0, 1] \setminus \mathbb{R}) = 1$ because $\lambda([0, 1]) = 1$. Consequently the Lebesgue integral of the Dirichlet function should (and in fact does) equal 0.

*If $f\colon A \to B$ is a map, the **preimage** of $M \subset B$ under f is the set $f^{-1}[M] := \{x \in A : f(x) \in M\}$.*

3.1.2 We want to measure the volume of general geometric objects

The objective of measure theory is to assign to subsets A of a given basic set Ω a nonnegative number $\mu(A)$, called the measure of A, which, in one way or another, measures its 'size'. Geometrically this 'size' could be the volume.

As an instructive example, let us look at determining the area of a two dimensional object. Our basic set Ω is $\mathbb{R}^2$ and, for $A \subset \mathbb{R}^2$, the measure $\lambda(A)$ should be its area. If R is a rectangle with side lengths a and b, then $\lambda(R)$ should be $a \cdot b$. With this information alone and some geometric considerations, one can already determine the area of more complicated objects. For example, the area of a right triangle T, where the two shorter sides have length a resp. b should be $\frac{1}{2}a \cdot b$, since two of these triangles can be assembled into a rectangle of side lengths a and b.

Proceeding, we can determine the area of any object which can be decomposed into finitely many right triangles, in particular for polygons. We could then use polygons to approximate a more complex object, e.g. a circle. Already Archimedes (287–212 BC) approximated the circle with regular polygons to obtain an approximation for the area of a circle.

Analyzing the operations above, such an geometric measure λ on $\mathbb{R}^N$ should have the following properties:

1. The unit cube should have measure 1.
2. If $A_1, \dots, A_n$ are disjoint, then
$$\lambda(A_1 \cup \dots \cup A_n) = \lambda(A_1) + \dots + \lambda(A_n).$$
3. If B is obtained by translating (i.e., rigidly moving) and rotating A, then $\lambda(A) = \lambda(B)$.
4. If A_n is an increasing sequence with $\bigcup_{n=1}^{\infty} A_n = B$, then $\lambda(B) = \sup_{n\in\mathbb{N}} \lambda(A_n)$.

It is not at all clear, whether a measure λ with the above geometric properties can be defined on all subsets of $\mathbb{R}^N$. In fact, this is not possible. The following example is particularly drastic. In three dimensions, Banach and Tarski proved, making use of the set theoretic axiom of choice, that one can partition a 3-dimensional ball into finitely many (actually, five will do) pieces and then, rotating and translating these pieces, reassemble them into *two* copies of the ball (thus doubling its volume?). As you can imagine, one cannot imagine how these pieces look like, but they must be rather irregular.

To get around such problems, one generally defines measures merely on a collection Σ of 'well-behaved' subsets that can sensibly be assigned a measure. Sometimes Σ will be the whole power set of a given set Ω, but for the geometric Lebesgue measure (even in one dimension) Σ cannot contain all of $\mathcal{P}(\mathbb{R})$.

Let us point out that not all measures are of a geometric nature. On arbitrary sets Ω geometric operations like translation and rotation make no sense. Specifically in probability theory one frequently consideres measures that have no geometric intepretation attached. We want a theory that is flexible enough to accomodate for this!

3.1.3 We want notions that are suited for probability theory

In an elementary probability course, you may have encountered *Laplace experiments*, i.e. random experiments with only finitely many possible outcomes. Typical examples are the tossing of a coin or the rolling of a die. Here the basic set Ω consists of all possible outcomes, in the aforementioned examples

$$\Omega_1 = \{\mathsf{H}, \mathsf{T}\} \quad \text{for coin tossing}$$

and

$$\Omega_2 = \{1, 2, 3, 4, 5, 6\} \quad \text{for rolling a die.}$$

A subset A of Ω is then called an *event*: $A_1 = \{\mathsf{H}\}$ corresponds in the first experiment to the event 'the coin came up heads'; in the second experiment, $A_2 = \{1,3,5\}$ corresponds to the event 'an odd number was rolled'. Such an event can now be assigned a 'likelihood' or 'probability' $\mathbb{P}(A)$, for example by setting

$$\mathbb{P}(A) = \frac{\#A}{\#\Omega},$$

where $\#S$ denotes the number of elements in a set S.

We would like to have a theory that allows to deal seamlessly with different measures and probability distributions, with the associated integration of functions and random variables, and it should be suitable to describe concepts like information. The study of stochastic processes requires all of the aforementioned.

3.2 σ-Algebras and their generators

We begin with the concept of a σ-algebra. A σ-algebra has the right structure for the collection of 'well-behaved' sets. In fact, measures will be (certain) maps defined on a σ-alegbra.

Definition 3.7. Let Ω be a nonempty set. A subset Σ of the power set $\mathscr{P}(\Omega)$ is called ***σ*-algebra** (on Ω) if the following three properties hold.

This can be interpreted in the context of probability spaces. The full space should be an event by (S1) (with probability 1), and (S2) says that if A is an event, then also its negation/complement is one.

Property (S3) is the reason for the prefix σ in σ-algebra. The prefix σ means 'countable'.

(S1) $\Omega \in \Sigma$.

(S2) $A \in \Sigma$ implies $A^{\mathrm{c}} \in \Sigma$.

(S3) $A_k \in \Sigma$ for all $k \in \mathbb{N}$ implies $\bigcup_{k\in\mathbb{N}} A_k \in \Sigma$.

A **measurable space** is a pair (Ω, Σ), where Ω is a nonempty set and Σ is a σ-algebra on Ω. The elements A of Σ are called **measurable sets**.

Example 3.8. Let Ω be a nonempty set. Then $\{\emptyset, \Omega\}$ is a σ-algebra on Ω; it is the smallest σ-algebra on Ω. Moreover, $\mathscr{P}(\Omega)$ is a σ-algebra on Ω; it is the largest σ-algebra on Ω. Finally, if $A \subset \Omega$, then $\{\emptyset, A, A^{\mathrm{c}}, \Omega\}$ is a σ-algebra on Ω.

Lemma 3.9. *Let (Ω, Σ) be a measurable space. Then one has the following.*

(a) $\emptyset \in \Sigma$.

(b) *If $A_1, \ldots, A_n \in \Sigma$, then $A_1 \cup \cdots \cup A_n \in \Sigma$ and $A_1 \cap \cdots \cap A_n \in \Sigma$.*

(c) *If $A_k \in \Sigma$ for all $k \in \mathbb{N}$, then $\bigcap_{k\in\mathbb{N}} A_k \in \Sigma$.*

Proof. (a) Follows directly from (S1) and (S2).

(b) Put $B_k := A_k$ for $k = 1, \ldots, n$ and $B_k = \emptyset$ for $k \geq n+1$. Then $B_k \in \Sigma$ for all $k \in \mathbb{N}$ and thus $\bigcup_{k\in\mathbb{N}} B_k = A_1 \cup \cdots \cup A_n \in \Sigma$ by (S3). Moreover, $A_1 \cap \cdots \cap A_n = (A_1^{\mathrm{c}} \cup \cdots \cup A_n^{\mathrm{c}})^{\mathrm{c}} \in \Sigma$ by (S2) and what was just proved.

(c) By DeMorgan's law,

$$\bigcap_{k\in\mathbb{N}} A_k = \big(\bigcup_{k\in\mathbb{N}} A_k^{\mathsf{c}}\big)^{\mathsf{c}} \in \Sigma$$

by (S2) and (S3). □

Lemma 3.10. *Let Ω be a nonempty set, J a nonempty index set and Σ_j a σ-algebra on Ω for all $j \in J$. Then $\bigcap_{j\in J} \Sigma_j$ is a σ-algebra on Ω.*

In other words, an arbitrary intersection of σ-algebras on the same base set is again a σ-algebra.

Proof. By (S1), one has $\Omega \in \Sigma_j$ for all $j \in J$ and thus $\Omega \in \bigcap_{j\in J} \Sigma_j$, proving (S1).

If $A \in \bigcap_{j\in J} \Sigma_j$, then $A \in \Sigma_j$ for all $j \in J$. Hence, by (S2), $A^{\mathsf{c}} \in \Sigma_j$ for all $j \in J$ and thus $A^{\mathsf{c}} \in \bigcap_{j\in J} \Sigma_j$, proving (S2).

Finally, if $A_k \in \bigcap_{j\in J} \Sigma_j$ for all $k \in \mathbb{N}$, then $A_k \in \Sigma_j$ for all $k \in \mathbb{N}$ and $j \in J$. Thus, by (S3), $\bigcup_{k\in\mathbb{N}} A_k \in \Sigma_j$ for all $j \in J$, proving $\bigcup_{k\in\mathbb{N}} A_k \in \bigcap_{j\in J} \Sigma_j$. □

Definition 3.11. Let Ω be a nonempty set, $\mathscr{A} \subset \mathscr{P}(\Omega)$. Then

$$\sigma(\mathscr{A}) := \bigcap \{\Sigma \subset \mathscr{P}(\Omega) : \Sigma \text{ is a } \sigma\text{-algebra and contains } \mathscr{A}\}$$

is a σ-algebra by Lemma 3.10. It is called the **σ-algebra generated by $\mathscr{A}$**. If Σ is a σ-algebra on Ω, then any $\mathscr{A} \subset \mathscr{P}(\Omega)$ with $\sigma(\mathscr{A}) = \Sigma$ is called a **generator of Σ**.

Example 3.12. If $A \subset \Omega$, then $\sigma(\{A\}) = \{\emptyset, A, A^{\mathsf{c}}, \Omega\}$. Moreover, $\sigma(\emptyset) = \{\emptyset, \Omega\}$.

If we have a topological structure on our base set, i.e., if we can speak about open subsets of the base set, we usually consider a special σ-algebra connected to this structure.

Definition 3.13. Let $(X, \|\cdot\|)$ be a normed space and $M \subset X$. Let τ be the collection of all relatively open subsets of M, i.e.

$$\tau := \{U \cap M : U \text{ is open in } (X, \|\cdot\|)\}.$$

Then $\sigma(\tau)$ is called the **Borel σ-algebra** on M and denoted by $\mathscr{B}(M)$ understanding $(X, \|\cdot\|)$ as a given.

The Borel σ-algebra can more generally be defined for metric and topological spaces. The important point is that the Borel σ-algebra connects measure theory with topology. In particular, continuous functions behave well with respect to the Borel σ-algebra.

Example 3.14. Let X, Y be normed spaces and $f\colon X \to Y$ a map. We shall show that f is continuous if and only if every preimage of an open set is open, i.e. if for all V open in Y one has that $f^{-1}[V]$ is open in X.

It will be a consequence of the above and Lemma 3.18 that $f^{-1}[V] \in \mathscr{B}(X)$ for all $V \in \mathscr{B}(Y)$.

Proof. Suppose $f\colon X \to Y$ is continuous. Let V be open in Y. Define $U := f^{-1}[V]$. If $U = \emptyset$, then U is trivially open. So suppose $x \in U$ and set $y := f(x)$. Then $y \in V$. Let $\varepsilon > 0$ be such that $B(y, \varepsilon) \subset V$. Assume for contradiction that there exists no $\delta > 0$ such that $B(x, \delta) \subset U$. In other words, assume

$$\forall \delta > 0 \; \exists z \in B(x, \delta) : f(z) \notin B(y, \varepsilon).$$

Then $\forall n \in \mathbb{N} \; \exists x_n \in B(x, \frac{1}{n}) : f(x_n) \notin B(y, \varepsilon)$. It follows that $x_n \to x$ and by continuity of f that $f(x_n) \to f(x) = y$. This contradicts $\|y - f(x_n)\|_Y \geq \varepsilon$ for all $n \in \mathbb{N}$. So there exists a $\delta > 0$ such that $B(x, \delta) \subset U$. As x was an arbitrary element of U, we obtain that U is open.

Conversely, suppose that preimages under f of open sets are open. Let (x_n) be a sequence in X such that $x_n \to x$ in X. Define $y := f(x)$. Let $\varepsilon > 0$. Then by assumption $U := f^{-1}[B(y, \varepsilon)]$ is open. As $x \in U$, there exists a $\delta > 0$ such that $B(x, \delta) \subset U$. Since $x_n \to x$, there exists an $n_0 \in \mathbb{N}$ such that $x_n \in B(x, \delta)$ for all $n \geq n_0$. But then $x_n \in U$ and $\|y - f(x_n)\|_Y < \varepsilon$ for all $n \geq n_0$. Hence $f(x_n) \to f(x)$. As (x_n) was an arbitrary convergent sequence, this shows that f is continuous. $\square$

Proposition 3.15. *Let $(X, \|\cdot\|)$ be a normed space and $M \subset X$. Denote by $\mathcal{C}$ the collection of relatively closed subsets of M, i.e.*

$$\mathcal{C} := \{F \cap M : F \text{ is closed in } (X, \|\cdot\|)\}.$$

Moreover, let $\mathcal{K}$ be the collection of all subsets of M that are compact in X.

(a) *$\mathscr{B}(M) = \sigma(\mathcal{C})$.*

(b) *If M is the countable union of compact sets, then $\mathscr{B}(M) = \sigma(\mathcal{K})$.*

In this case M is called σ-compact.

(c) *If M has a countable subset D such that $M \subset \overline{D}$, then $\mathscr{B}(M)$ is generated by a countable collection of sets of the type $B(x, \varepsilon) \cap M$ with $x \in M$ and $\varepsilon > 0$.*

Both (b) and (c) in Proposition 3.15 are false without the additional assumptions.

Proof. For the proof we shall assume that $M = X$. The general case follows in the same way by intersecting with M and using realtively open sets instead of open sets, for example. This is best understood in the context of metric spaces which we will briefly encounter later.

(a) Let Σ be a σ-algebra containing all closed sets. By (S2), Σ contains all open sets and thus $\mathscr{B}(X)$. Hence $\mathscr{B}(X) \subset \sigma(\mathcal{C})$. Conversely, if Σ be a σ-algebra containing all open sets, Σ contains all closed sets and thus $\sigma(\mathcal{C})$ by (S2). Hence $\sigma(\mathcal{C}) \subset \mathscr{B}(X)$.

(b) Since $\mathcal{K} \subset \mathcal{C}$ we clearly have $\sigma(\mathcal{K}) \subset \sigma(\mathcal{C}) \subset \mathscr{B}(X)$. Now assume that $X = \bigcup_{k \in \mathbb{N}} K_n$ where K_n is compact. If F is closed then $F \cap K_n$ is compact (Exercise!) and hence an element of $\sigma(\mathcal{K})$ for all $n \in \mathbb{N}$. Since $\bigcup_{n \in \mathbb{N}} K_n = \Omega$, we have $\bigcup_{n \in \mathbb{N}} F \cap K_n = F$. Hence, by (S3), $F \in \sigma(\mathcal{K})$. Thus $\sigma(\mathcal{K})$ contains all closed sets and thus the σ-algebra generated by them, i.e. $\mathscr{B}(X)$.

(c) We only have to show that there is a countable union of open balls such that any open set is contained in the σ-algebra generated by these balls. To that end, let $\{x_n : n \in \mathbb{N}\}$ be a countable dense subset of X. Consider the open balls $B(x_n, \frac{1}{k})$ for all $n \in \mathbb{N}$ and $k \in \mathbb{N}$. This is a countable collection of open balls and every open subset U of M is a countable union of such balls. Indeed, if U is empty, there is nothing to prove. If $x \in U$ then $B(x, \frac{2}{k_0}) \subset U$ for some $k_0 \in \mathbb{N}$. Since $\{x_n : n \in \mathbb{N}\}$ is dense, there exists an $n_0 \in \mathbb{N}$ with $x_{n_0} \in B(x, \frac{1}{k_0})$. Hence $x \in B(x_{n_0}, \frac{1}{k_0})$. Moreover, $B(x_{n_0}, \frac{1}{k_0}) \subset B(x, \frac{2}{k_0}) \subset U$.

It now follows by (S3) that U belongs to the σ-algebra generated by these open balls. □

3.3 Measurable maps

We introduce the concept of a *measurable map*. Loosely speaking, measurable maps are for measurable spaces what continuous maps are for normed (or more generally metric or topological) spaces.

Definition 3.16. Let (Ω_1, Σ_1) and (Ω_2, Σ_2) be measurable spaces. A map $f\colon \Omega_1 \to \Omega_2$ is called **measurable**, more precisely Σ_1/Σ_2-measurable, if $f^{-1}[A] \in \Sigma_1$ for all $A \in \Sigma_2$.

Exercise 3.17. Show that the composition of measurable maps is measurable.

Lemma 3.18. *Let (Ω_1, Σ_1) and (Ω_2, Σ_2) be a measurable spaces and $f\colon \Omega_1 \to \Omega_2$ be a map. Moreover, let $\mathscr{A} \subset \mathscr{P}(\Omega_2)$ be a generator of Σ_2, i.e. $\sigma(\mathscr{A}) = \Sigma_2$. Then f is Σ_1/Σ_2-measurable if and only if $f^{-1}[A] \in \Sigma_1$ for all $A \in \mathscr{A}$.*

Proof. We use the *principle of good sets*:

Let $\mathscr{G} := \{A \in \Sigma_2 : f^{-1}[A] \in \Sigma_1\}$. Then $\mathscr{G}$ is a σ-algebra. Indeed, $f^{-1}[\Omega_2] = \Omega_1 \in \Sigma_1$. Moreover, it follows from the properties of the preimage that $f^{-1}[A^c] = (f^{-1}[A])^c$ and

$$f^{-1}\Big[\bigcup_{k\in\mathbb{N}} A_k\Big] = \bigcup_{k\in\mathbb{N}} f^{-1}[A_k].$$

These equalities and the axioms of a σ-algebra for Σ_1 imply that $\mathscr{G}$ contains A^c if it contains A and it contains $\bigcup_{k\in\mathbb{N}} A_k$ provided it contains A_k for every $k \in \mathbb{N}$.

Thus, if $\mathscr{G}$ contains $\mathscr{A}$, then it contains $\sigma(\mathscr{A}) = \Sigma_2$, i.e. if $f^{-1}[A] \in \Sigma_1$ for all $A \in \mathscr{A}$, then f is Σ_1/Σ_2-measurable. The converse is obvious. □

The principle of good sets *is a standard method of proof in measure theory. In the proof to the left, $\mathscr{G}$ is defined as the collection of 'good' sets in Σ_2. Then it is shown that $\mathscr{G}$ is as big as it can be, i.e., $\mathscr{G} = \Sigma_2$.*

Example 3.19. Let X, Y be normed spaces, $M \subset X$ and $f\colon M \to Y$ be continuous. Then f is $\mathscr{B}(M)/\mathscr{B}(Y)$-measurable by Exercise 3.14 and Lemma 3.18.

We next prove that, in a certain sense, the Borel σ-algebra is the smallest σ-algebra such that all continuous functions are measureable. To that end, we first introduce the following concept.

Definition 3.20. Let Ω_1 be a nonempty set and (Ω_2, Σ_2) be a measurable space. Moreover, let $\mathscr{F}$ be a collection of maps from Ω_1 to Ω_2.

Then $\sigma(\mathscr{F}) := \sigma(\{f^{-1}[A] : f \in \mathscr{F},\ A \in \Sigma_2\})$ is called the ***σ*-algebra generated by** $\mathscr{F}$ (on Ω_1).

Remark 3.21. In the situation of Definition 3.20, $\sigma(\mathscr{F})$ is the smallest σ-algebra Σ_1 such that every $f \in \mathscr{F}$ is Σ_1/Σ_2-measurable. Using the principle of good sets, one sees that if $\mathscr{A}$ is a generator of Σ_2, then $\sigma(\mathscr{F})$ is generated by $\{f^{-1}[A] : f \in \mathscr{F},\ A \in \mathscr{A}\}$.

Exercise 3.22. Let $\Omega := \{0,1\}^2$ and $f\colon \Omega \to \mathbb{R}$ be given by $f(x_1, x_2) = x_1 + x_2$. Endow $\mathbb{R}$ with its Borel σ-algebra and determine $\sigma(\{f\})$. Decide whether the map $g\colon \Omega \to \mathbb{R}$ given by $g(x_1, x_2) = x_1$ is $\sigma(\{f\})/\mathscr{B}(\mathbb{R})$-measurable.

We now prove that we can actually *characterize* the Borel σ-algebra as the one generated by the continuous functions.

We only assume $\mathbb{K} = \mathbb{R}$ for convenience. The result also holds in the context of metric spaces.

Proposition 3.23. *Let $\mathbb{K} = \mathbb{R}$ and X a normed space. Then $\mathscr{B}(X) = \sigma(\mathrm{C}(X;\mathbb{R}))$.*

Proof. Since $f^{-1}[U]$ is open in X for all open subsets U of $\mathbb{R}$ and $f \in \mathrm{C}(X;\mathbb{R})$, we have

$$\sigma(\{f^{-1}[U] : f \in \mathrm{C}(X;\mathbb{R}), U \subset \mathbb{R} \text{ open}\}) \subset \mathscr{B}(X).$$

By Lemma 3.18, $\sigma(\mathrm{C}(X;\mathbb{R})) \subset \mathscr{B}(X)$.

Conversely, let F be a closed subset of X. Then

$$F_n := \left\{x \in X : \inf_{y \in F} \|x - y\| \geq \frac{1}{n}\right\}$$

is closed and disjoint from F for all $n \in \mathbb{N}$. By Urysohn's lemma 2.64 there exist continuous functions $f_n\colon X \to [0,1]$ such that $f_n(x) = 1$ for $x \in F$ and $f_n(x) = 0$ for $x \in F_n$. But then $F = \bigcap_{n \in \mathbb{N}} f_n^{-1}[\{1\}]$. It follows that $\sigma(\mathrm{C}(X;\mathbb{R}))$ contains all closed sets and hence $\mathscr{B}(X)$. □

Observe that the somewhat ambiguous notation $\sigma(\{\pi_t : t \in [0,1]\})$ is supposed to yield a σ-algebra on $\mathrm{C}[0,1]$.

Proposition 3.24. *Define the **point evaluation maps** $\pi_t\colon \mathrm{C}[0,1] \to \mathbb{R}$ by $\pi_t(f) = f(t)$ for all $t \in [0,1]$. Then one has*

$$\mathscr{B}(\mathrm{C}[0,1]) = \Sigma := \sigma(\{\pi_t : t \in [0,1]\}).$$

In the lecture we skipped the proof of Proposition 3.24 and the following example.

Proof. Note that π_t is a continuous function from $C[0,1]$ to $\mathbb{R}$ for all $t \in [0,1]$. Thus, by Proposition 3.23, $\Sigma \subset \mathscr{B}(C[0,1])$.

To prove the converse inclusion, fix $g \in C[0,1]$ and $\varepsilon > 0$ and consider

$$C(g,\varepsilon) := \{f \in C[0,1] : \|f-g\|_\infty \le \varepsilon\}.$$

Let $\{t_k : k \in \mathbb{N}\} = [0,1] \cap \mathbb{Q}$. Clearly,

$$F := \bigcap_{k\in\mathbb{N}} \pi_{t_k}^{-1}\Big[[g(t_k)-\varepsilon, g(t_k)+\varepsilon]\Big]$$

is a set in Σ. We claim that $C(g,\varepsilon) = F$. Indeed, the inclusion '$\subset$' is clear. To see the converse, let $f \in F$. Then $|f(t)-g(t)| \le \varepsilon$ for all $t \in [0,1] \cap \mathbb{Q}$. Now let $t \in [0,1]$. By density, there exists a sequence $t_k \in [0,1] \cap \mathbb{Q}$ converging to t. Since f and g are continuous, $|f(t)-g(t)| = \lim_{k\to\infty}|f(t_k)-g(t_k)| \le \varepsilon$, proving that $f \in C(g,\varepsilon)$.

Hence Σ contains all closed balls. But then also

$$B(g,\varepsilon) = \bigcup_{n\in\mathbb{N}} C(g,(1-\tfrac{1}{n})\varepsilon) \in \Sigma.$$

Since $C[0,1]$ is separable by Corollary 2.69, it follows from Proposition 3.15 that $\mathscr{B}(C[0,1]) \subset \Sigma$. □

Example 3.25. A set $A \subset C[0,1]$ is called a **cylinder set** if there exists $n \in \mathbb{N}$, $t_1,\dots,t_n \in [0,1]$ and $A_1,\dots,A_n \in \mathscr{B}(\mathbb{R})$ such that

$$A = \{f \in C[0,1] : f(t_k) \in A_k \text{ for all } k = 1,\dots,n\}.$$

It follows from Proposition 3.24 that the Borel σ-algebra $\mathscr{B}(C[0,1])$ is generated by the cylinder sets.

Exercise 3.26. We define $\mathscr{B}_t := \sigma(\{\pi_s : s \in [0,t]\})$ and $\mathscr{B}_{t+} := \bigcap_{s>t}\mathscr{B}_s$. Prove that $\mathscr{B}_t \neq \mathscr{B}_{t+}$ for all $t \in [0,1)$.

Hint: If $g,h\colon \Omega \to \mathbb{R}$ and $\mathbb{R}$ is endowed with the Borel sigma algebra, show that the set $\{x \in \Omega : g(x) \le h(x)\}$ is contained in $\sigma(\{g,h\})$. The same holds if $\le$ is replaced with $<$ or $=$. Deduce that $\{f \in C[0,1] : \pi_q(f) \le \pi_r(f)\} \in \mathscr{B}_t$ if $0 \le q,r \le t$.

Now consider the set $A_{\max}(t) := \{f \in C[0,1] : f \text{ has a local maximum at } t\}$. Prove that $A_{\max}(t) \in \mathscr{B}_{t+}$. Show then that if $B \in \mathscr{B}_t$, then $f \in B$ and $f(s) = g(s)$ for all $s \in [0,t]$ implies that $g \in B$. Use this to prove that $A_{\max}(t) \notin \mathscr{B}_t$.

The exercise to the side has an interesting interpretation in the context of stock values. In case of the event $A_{\max}(t)$, it would be ideal to sell stock at time t. However, normally the knowledge/information at time t is equal to $\mathscr{B}_t$, but not to $\mathscr{B}_{t+}$. So it is not possible to decide at time t, whether one is in a local maximum (and whether it is an ideal time to sell).

3.4 Measures

In this section we study measures. A measure is a map that assigns something like a 'volume' to sets in a σ-algebra.

Definition 3.27. Let (Ω,Σ) be a measurable space. A **measure** on (Ω,Σ) is a map $\mu\colon \Sigma \to [0,\infty]$ such that

(M1) $\mu(\emptyset) = 0$.

*The property (M2) is called **σ-additivity**.*

(M2) If (A_k) is a sequence of pairwise disjoint sets in Σ, then

$$\mu\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) = \sum_{k\in\mathbb{N}} \mu(A_k).$$

A **measure space** is a triple (Ω, Σ, μ) where (Ω, Σ) is a measurable space and μ is a measure on (Ω, Σ).

A measure space (Ω, Σ, μ) (or sometimes the measure μ) is called **finite** if $\mu(\Omega) < \infty$. If $\mu(\Omega) = 1$, then (Ω, Σ, μ) is called **probability space** and μ is called **probability measure**. The measure space is called **σ-finite** if there exists a sequence (Ω_n) in Σ with $\mu(\Omega_n) < \infty$ and $\bigcup_{n\in\mathbb{N}} \Omega_n = \Omega$.

Remark 3.28. The σ-additivity (M2) ensures that the measure is well-behaved with respect to taking countable unions. As we will see below in Proposition 3.30, (M2) implies items 2. and 4. from Subsection 3.1.2 in the introduction. In fact, it is equivalent to them.

Frequently, statements about σ-finite measure spaces can be reduced to the corresponding statements about finite measure spaces. So σ-finiteness is a convenient assumption that is often satisfied and easily checked.

Example 3.29. We give some preliminary examples:

(a) Let (Ω, Σ) be a measurable space. Then $\mathbf{0}\colon \Sigma \to [0, \infty)$ given by $\mathbf{0}(A) = 0$ is a measure on (Ω, Σ).

(b) Let (Ω, Σ) be a measurable space and $x \in \Omega$. Then $\delta_x\colon \Sigma \to \{0, 1\}$, defined by $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \notin A$, is a probability measure on (Ω, Σ), the so-called **Dirac measure** in x.

(c) Consider $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. Then

$$\zeta(A) := \begin{cases} \infty, & \text{if } A \text{ is infinite,} \\ \#A, & \text{if } A \text{ is finite,} \end{cases}$$

where $\#A$ is the number of elements in A, defines a σ-finite measure on $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$, the so-called **counting measure** on $\mathbb{N}$.

(d) On $M := \mathbb{R}$, consider $\Sigma = \{A \subset M : A \text{ is countable or } A^{\mathsf{c}} \text{ is countable}\}$. Define

$$\zeta(A) := \begin{cases} \infty, & \text{if } A \text{ is infinite,} \\ \#A, & \text{if } A \text{ is finite.} \end{cases}$$

Then ζ defines a measure on (M, Σ), which is not σ-finite.

We next collect basic properties of measures.

Proposition 3.30. *Let (Ω, Σ, μ) be a measure space.*

(a) *If $A, B \in \Sigma$ with $A \subset B$, then $\mu(A) \le \mu(B)$.*

(b) *If $A, B \in \Sigma$ with $A \subset B$ and $\mu(B) < \infty$, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.*

(c) *If (A_k) is a sequence in Σ (not necessarily disjoint), then*

σ-subadditivity

$$\mu\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) \le \sum_{k=1}^{\infty} \mu(A_k).$$

(d) *If (A_k) is an increasing sequence in Σ, i.e. $A_k \subset A_{k+1}$ for all $k \in \mathbb{N}$, and $A = \bigcup_{k\in\mathbb{N}} A_k$ (we write $A_k \uparrow A$), then $\mu(A_k) \uparrow \mu(A)$.*

Continuity from below

(e) *If $A_k \downarrow A$, i.e. $A_{k+1} \subset A_k$ for all $k \in \mathbb{N}$ and $A = \bigcap_{k\in\mathbb{N}} A_k$, and $\mu(A_1) < \infty$, then $\mu(A_k) \downarrow \mu(A)$.*

Continuity from above. Give an example which shows that the assumption $\mu(A_1) < \infty$ is needed in (e).

Proof. (a) B is the disjoint union of B and $B \setminus A$. Hence, by (M2), $\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A)$, since $\mu(A \setminus B) \ge 0$. If $\mu(B) < \infty$ then also $\mu(A) < \infty$ and subtracting $\mu(A)$, also (b) follows.

(c) Define $B_1 := A_1$, $B_2 := A_2 \setminus A_1$, $B_3 = A_3 \setminus (A_1 \cup A_2)$,..., $B_k := A_k \setminus (A_1 \cup \cdots \cup A_{k-1})$. By the properties of a σ-algebra, $B_k \in \Sigma$ for all $k \in \mathbb{N}$. Moreover, $B_k \subset A_k$ for all $k \in \mathbb{N}$, the sets B_k are disjoint and $\bigcup_{k\in\mathbb{N}} A_k = \bigcup_{k\in\mathbb{N}} B_k$. Hence, by (M2)

$$\mu\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \mu\Big(\bigcup_{k=1}^{\infty} B_k\Big) = \sum_{k=1}^{\infty} \mu(B_k) \le \sum_{k=1}^{\infty} \mu(A_k),$$

where we have used (a) in the last estimate.

(d) By (a) $\mu(A_k)$ increases in k. Now put $A_0 = \varnothing$ and then $B_k := A_k \setminus A_{k-1}$. Then the B_k are pairwise disjoint, belong to Σ and $\bigcup_{k\in\mathbb{N}} B_k = A$. Hence, by (M2)

$$\mu(A) = \sum_{k=1}^{\infty} \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) = \lim_{n\to\infty} \mu\Big(\bigcup_{k=1}^{n} B_k\Big) = \lim_{n\to\infty} \mu(A_n).$$

(e) This is immediate from (b) and (d), since $(A_1 \setminus A_n) \uparrow (A_1 \setminus A)$. Hence

$$\mu(A_1) - \mu(A) = \mu(A_1 \setminus A) = \lim \mu(A_1 \setminus A_n) = \mu(A_1) - \lim \mu(A_n). \quad \square$$

A convenient way to obtain new measures from known ones is through measurable maps:

Lemma 3.31. *Let (Ω, Σ, μ) be a measure space and (Ω_2, Σ_2) be a measurable space. Let $f\colon \Omega \to \Omega_2$ be measurable. Then $\mu_f\colon \Sigma_2 \to [0, \infty]$, defined by $\mu_f(A) := \mu(f^{-1}[A])$ is a measure. It is called **the push-forward of μ under f**.*

*If (Ω, Σ, μ) is a probability space and $(\Omega_2, \Sigma_2) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, then the push-forward μ_f is the **distribution** of the random variable f. In particular, in this case μ_f is a probability measure on $\mathbb{R}$.*

Proof. This follows from the fact that $f^{-1}[\varnothing] = \varnothing$ and $f^{-1}[\bigcup_{k\in\mathbb{N}} A_k] = \bigcup_{k\in\mathbb{N}} f^{-1}[A_k]$, where the latter sets are disjoint if the A_k are. $\square$

We now study the question, whether a measure is already uniquely determined by its values on a smaller set than Σ. Of particular importance is the question whether it is uniquely determined by its values on a generator of Σ.

Let us first look at an example.

Example 3.32. Let $\Omega = \{0,1,2,3\}$ and $\Sigma = \sigma(\{0,1\},\{1,2\})$. Then $\mu = \delta_0 + \delta_1 + \delta_2$ and $\nu = 2\delta_1 + \delta_3$ satisfy $\mu(\{0,1\}) = \nu(\{0,1\}) = 2$ and $\mu(\{1,2\}) = \nu(\{1,2\}) = 2$, but $\mu \neq \nu$ since $\mu(\{1\}) = 1 \neq 2 = \nu(\{1\})$.

The key in studying uniqueness of measures lies in considering so-called *Dynkin systems*.

Definition 3.33. Let Ω be a nonempty set. A **Dynkin system** is a collection $\mathscr{D} \subset \mathscr{P}(\Omega)$ such that the following three properties hold.

(D1) $\Omega \in \mathscr{D}$.

(D2) If $A \in \mathscr{D}$ then $A^{\mathsf{c}} \in \mathscr{D}$.

(D3) If (A_k) is a sequence of *pairwise disjoint* subsets of $\mathscr{D}$, then

$$\bigcup_{k\in\mathbb{N}} A_k \in \mathscr{D}.$$

Obviously, every σ-algebra is a Dynkin system. In fact, the only difference between a Dynkin system and a σ-algebra is that in (D3), different from (S3), the sequence (A_k) is required to consist of disjoint subsets. Similarly to $\sigma(\mathscr{A})$ there is a smallest Dynkin system containing a given $\mathscr{A}$, denoted by $\operatorname{dyn}(\mathscr{A})$, the **Dynkin system generated** by $\mathscr{A}$.

Exercise 3.34. Show that the arbitrary intersection of Dynkin systems on Ω is again a Dynkin system.

Lemma 3.35. *A Dynkin system $\mathscr{D}$ is a σ-algebra if and only if whenever A and B belong to $\mathscr{D}$ then also $A \cap B \in \mathscr{D}$. For the latter we say that $\mathscr{D}$ is **stable under intersections**.*

Proof. Since every σ-algebra is stable under intersections (see Lemma 3.9), we only need to prove that if a Dynkin system is stable under intersection, then it satisfies (S3) (since (S1) and (S2) clearly hold).

So let a sequence (A_k) of (not necessarily disjoint) sets in $\mathscr{D}$ be given. We put $B_1 := A_1$ and then inductively $B_{k+1} := A_k \cap (B_1 \cup \cdots \cup B_k)^{\mathsf{c}}$. One then proves that the sequence B_k consists of disjoint sets, which then in turn proves that $B_k \in \mathscr{D}$, since the latter was assumed to be stable under intersections. Moreover, $\bigcup_{k\in\mathbb{N}} A_k = \bigcup_{k\in\mathbb{N}} B_k$ and the latter belongs to $\mathscr{D}$ by the pairwise disjointness and (D3). This proves (S3). $\square$

Lemma 3.36 (Dynkin's π–λ theorem)**.** *Let Ω be a nonempty set and $\mathscr{A} \subset \mathscr{P}(\Omega)$ be stable under intersections. Then $\operatorname{dyn}(\mathscr{A}) = \sigma(\mathscr{A})$.*

Proof. Clearly, $\operatorname{dyn}(\mathscr{A}) \subset \sigma(\mathscr{A})$ since $\sigma(\mathscr{A})$ is a Dynkin system containing $\mathscr{A}$. To prove the converse inclusion, by Lemma 3.35 it suffices to prove that $\operatorname{dyn}(\mathscr{A})$ is stable under intersections.

To that end, for $B \in \operatorname{dyn}(\mathscr{A})$ define

$$\mathscr{G}_B := \{A \in \operatorname{dyn}(\mathscr{A}) : A \cap B \in \operatorname{dyn}(\mathscr{A})\}.$$

Fix $B \in \mathscr{A}$. Then $\mathscr{G}_B$ is a Dynkin system. Indeed, $\Omega \cap B = B \in \operatorname{dyn}(\mathscr{A})$ proving (D1). Also (D3) easily follows since if the sets $A_k \in \mathscr{G}_B$ are disjoint, then $(\bigcup_{k\in\mathbb{N}} A_k) \cap B = \bigcup_{k\in\mathbb{N}}(A_k \cap B)$ and the latter union is also disjoint. For (D2), let $A \in \mathscr{G}_B$ and observe that $A^{\mathrm{c}} \cap B = B \cap (A \cap B)^{\mathrm{c}} = (B^{\mathrm{c}} \cup (A \cap B))^{\mathrm{c}}$. By assumption, $A \cap B \in \mathscr{G}_B$. Since $B^{\mathrm{c}} \cap (A \cap B) = \emptyset$, it follows that $A^{\mathrm{c}} \cap B \in \mathscr{G}_B$.

By assumption $\mathscr{A} \subset \mathscr{G}_B$ for every $B \in \mathscr{A}$ and hence $\operatorname{dyn}(\mathscr{A}) \subset \mathscr{G}_B \subset \operatorname{dyn}(\mathscr{A})$ for all $B \in \mathscr{A}$. Now set

$$\begin{aligned}\mathscr{G} &:= \{B \in \operatorname{dyn}(\mathscr{A}) : A \cap B \in \operatorname{dyn}(\mathscr{A}) \text{ for all } A \in \operatorname{dyn}(\mathscr{A})\} \\ &= \{B \in \operatorname{dyn}(\mathscr{A}) : \mathscr{G}_B = \operatorname{dyn}(\mathscr{A})\}.\end{aligned}$$

By what was done so far, $\mathscr{A} \subset \mathscr{G}$. Similarly as above (Exercise!), one sees that $\mathscr{G}$ is a Dynkin system. It now follows that $\mathscr{G} = \operatorname{dyn}(\mathscr{A}) = \sigma(\mathscr{A}) = \Sigma$, as a consequence of Lemma 3.35. ☐

We can now prove the following result on uniqueness of measures.

Theorem 3.37. *Let (Ω, Σ) be a measurable space and $\mathscr{A}$ be a generator of Σ which is stable under intersections. If μ and ν are two finite measures on (Ω, Σ) with $\mu(\Omega) = \nu(\Omega)$ such that $\mu(A) = \nu(A)$ for all $A \in \mathscr{A}$, then $\mu = \nu$.*

Proof. Let $\mathscr{G} = \{A \in \Sigma : \mu(A) = \nu(A)\}$. Then $\Omega \in \mathscr{G}$. If $A \in \mathscr{G}$, then $A^{\mathrm{c}} \in \mathscr{G}$ since $\mu(A^{\mathrm{c}}) = \mu(\Omega) - \mu(A) = \nu(\Omega) - \nu(A) = \nu(A^{\mathrm{c}})$. Moreover, if (A_k) is a sequence of disjoint sets in $\mathscr{G}$, then

$$\mu\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k) = \sum_{k=1}^{\infty} \nu(A_k) = \nu\Big(\bigcup_{k\in\mathbb{N}} A_k\Big)$$

since μ and ν are measures. Thus $\mathscr{G}$ is a Dynkin system. Since $\mathscr{A} \subset \mathscr{G}$ by assumption, it follows from Lemma 3.36 that $\Sigma = \sigma(\mathscr{A}) = \operatorname{dyn}(\mathscr{A}) \subset \mathscr{G}$, proving that $\mu(A) = \nu(A)$ for all $A \in \Sigma$. ☐

Corollary 3.38. *Let (Ω, Σ) be a measurable space and $\mathscr{A}$ be a generator of Σ that is stable under intersections. Suppose that there exists an increasing sequence (Ω_n) in $\mathscr{A}$ with $\bigcup \Omega_n = \Omega$. If μ and ν are two (σ-finite) measures on (Ω, Σ) such that $\mu(A) = \nu(A) < \infty$ for all $A \in \mathscr{A}$, then $\mu = \nu$.*

Proof. For fixed n, consider $\Sigma_n := \sigma(\{A \cap \Omega_n : A \in \mathscr{A}\})$. Then it is easily checked that $\Sigma_n = \Omega_n \cap \Sigma$ (Exercise!). By Theorem 3.37, $\mu(A) = \nu(A)$ for all $A \in \Sigma_n$. Now if $A \in \Sigma$, then

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \nu(A \cap \Omega_n) = \nu(A),$$

where we have used continuity from below for μ and ν, and the fact that $A \cap \Omega_n \in \Sigma_n$ for all $A \in \Sigma$, which follows from the assumption that $\mathscr{A}$ generates Σ. □

3.5 Construction of measures

In order to construct a measure, one first defines it on a system of sets much smaller than a σ-algebra.

Definition 3.39. Let Ω be a nonempty set. A **ring** on Ω is a subset $\mathscr{R}$ of $\mathscr{P}(\Omega)$ such that the following two properties hold.

(R1) $\emptyset \in \mathscr{R}$.

(R2) If $A, B \in \mathscr{R}$, then $A \cup B \in \mathscr{R}$ and $B \setminus A \in \mathscr{R}$.

Note that $A \cap B \in \mathscr{R}$ as $A \cap B = B \setminus (B \setminus A)$.

A **pre-measure** is a function $\mu\colon \mathscr{R} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and if (A_k) is a sequence of disjoint sets in $\mathscr{R}$ with $\bigcup_{k\in\mathbb{N}} A_k \in \mathscr{R}$, then

$$\mu\Big(\bigcup_{k\in\mathbb{N}} A_k\Big) = \sum_{k=1}^{\infty} \mu(A_k).$$

Remark 3.40. The properties in Proposition 3.30 remain valid for pre-measures, if one additionally requires that the countable unions or intersections appearing belong to $\mathscr{R}$. In particular, pre-measures are monotone, finitely additive and, provided that the respective countable union belongs to $\mathscr{R}$, also σ-subadditive.

Example 3.41. Let $\Omega = \mathbb{R}$ and $\mathscr{R}$ be the collection of all finite unions of bounded (possibly empty), right-open intervals. A typical element of $\mathscr{R}$ is of the form

$$[a_1, b_1) \cup [a_2, b_2) \cup \cdots \cup [a_n, b_n)$$

with $-\infty < a_1 \le b_1 < a_2 \le b_2 < \cdots < a_n \le b_n < \infty$. Then $\mathscr{R}$ is a ring. The map $\lambda\colon \mathscr{R} \to [0, \infty)$, given by

$$\lambda([a_1, b_1) \cup [a_2, b_2) \cup \cdots \cup [a_n, b_n)) = \sum_{j=1}^{n} (b_j - a_j),$$

defines a pre-measure on $\mathscr{R}$, called the **Lebesgue pre-measure**. We leave the verification that λ is a pre-measure as an (important!) exercise.

We should note that in the above example, $\sigma(\mathscr{R}) = \mathscr{B}(\mathbb{R})$, as is easy to see (Exercise!). The question arises, whether λ can be extended to a measure on $\mathscr{B}(\mathbb{R})$. It follows from the following theorem that this is the case.

Theorem 3.42 (Carathéodory's extension theorem). *Let $\mathscr{R}$ be a ring on the nonempty set Ω and μ be a pre-measure on $\mathscr{R}$. Then μ extends to a measure on $\sigma(\mathscr{R})$.*

This result is also called Hahn–Kolmogorov theorem.

Proof. For any $B \subset E$, define

In the lecture we have only given a short outline of this proof.

$$\mu^*(B) := \inf \sum_{n=1}^{\infty} \mu(A_n),$$

where the infimum is taken over all sequences (A_k) in $\mathscr{R}$ such that $B \subset \bigcup_{k\in\mathbb{N}} A_k$. If no such sequence exists, we put $\mu^*(B) = \infty$. We now proceed in several steps:

Step1: We prove that μ^* is σ-subadditive, i.e. if (B_n) is a sequence of subsets of Ω, then $\mu^*(\bigcup_{n\in\mathbb{N}} B_n) \le \sum_{n=1}^{\infty} \mu^*(B_n)$.
If $\mu^*(B_n) = \infty$ for some n, then there is nothing to prove. So let us assume that $\mu^*(B_n) < \infty$ for all $n \in \mathbb{N}$. By definition, for all $\varepsilon > 0$ there exist sequences $(A^{\varepsilon}_{n,m})_{m\in\mathbb{N}}$ in $\mathscr{R}$ such that

(i) $B_n \subset \bigcup_{m\in\mathbb{N}} A^{\varepsilon}_{n,m}$, and
(ii) $\mu^*(B_n) \ge \sum_{m=1}^{\infty} \mu(A^{\varepsilon}_{n,m}) - \varepsilon 2^{-n}$.

In this situation, $B := \bigcup_{n\in\mathbb{N}} B_n \subset \bigcup_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}} A^{\varepsilon}_{n,m}$ and hence, by the definition of μ^*,

$$\mu^*(B) \le \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \mu(A^{\varepsilon}_{n,m}) \le \sum_{n=1}^{\infty} (\mu^*(B_n) + \varepsilon 2^{-n}) \le \varepsilon + \sum_{n=1}^{\infty} \mu^*(B_n).$$

Since $\varepsilon > 0$ was arbitrary, the σ-subadditivity follows.

Step 2: We prove that $\mu^*(A) = \mu(A)$ for all $A \in \mathscr{R}$. Clearly, $\mu^*(A) \le \mu(A)$ for $A \in \mathscr{R}$, since $A \subset A \cup \emptyset \cup \emptyset \cup \ldots$. On the other hand, if $A \in \mathscr{R}$ and $A \subset \bigcup_{n\in\mathbb{N}} A_n$ for some sequence (A_n) in $\mathscr{R}$, then

$$\mu(A) \le \sum_{k=1}^{\infty} \mu(A \cap A_k) \le \sum_{k=1}^{\infty} \mu(A_k)$$

since μ is countably subadditive and monotone. Hence $\mu(A) \le \mu^*(A)$.
Define $\mathscr{M} := \{A \subset \Omega : \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A) \text{ for all } B \subset \Omega\}$.

Step 3: We prove that $\mathscr{M}$ is a σ-algebra and μ^* is a measure on $(\Omega, \mathscr{M})$. By Lemma 3.35, for the first claim it suffices to prove that $\mathscr{M}$ is a Dynkin system that is stable under intersections. Clearly, (D1) and (D2) hold. Now let (A_k) be a sequence of disjoint sets in $\mathscr{M}$. Let $B \subset \Omega$ be fixed. Then

$$\begin{aligned}\mu(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) \\ &\quad + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\ &= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c),\end{aligned}$$

since A_1 and A_2 are disjoint, and therefore $A_1 \cap A_2 = \emptyset$, $A_1 \cap A_2^c = A_1$ and $A_1^c \cap A_2 = A_2$.
Proceeding in a similar way, we obtain for all $n \in \mathbb{N}$ that

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A_1^c \cap \cdots \cap A_n^c) + \sum_{k=1}^{n} \mu^*(B \cap A_k) \\ &\geq \mu^*\Big(B \cap \Big(\bigcup_{k=1}^{\infty} A_k\Big)^c\Big) + \sum_{k=1}^{n} \mu^*(B \cap A_k).\end{aligned}$$

Hence, upon $n \to \infty$, we obtain

$$\begin{aligned}\mu^*(B) &\geq \sum_{k=1}^{\infty} \mu^*(B \cap A_k) + \mu^*\Big(B \cap \Big(\bigcup_{k\in\mathbb{N}} A_k\Big)^c\Big) \\ &\geq \mu^*\Big(B \cap \bigcup_{k\in\mathbb{N}} A_k\Big) + \mu^*\Big(B \cap \Big(\bigcup_{k\in\mathbb{N}} A_k\Big)^c\Big),\end{aligned} \tag{3.1}$$

where we have used Step 1 in the last estimate. Since the reverse inequality holds by subadditivity, $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{M}$, i.e. (D3) holds. Moreover, taking $B = \bigcup_{k\in\mathbb{N}} A_k$ in (3.1), the σ-additivity of μ^* on $\mathscr{M}$ follows.
It thus remains to prove that $\mathscr{M}$ is stable under intersections. To that end, let $A_1, A_2 \in \mathscr{M}$. Then, for $B \subset \Omega$, we have

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\ &\qquad \text{(as } A_1 \in \mathscr{M}) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \\ &\qquad \text{(as } A_2 \in \mathscr{M}) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\ &\qquad \text{(as } A_1^c \subset (A_1 \cap A_2)^c) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\ &\qquad \text{(as } (A_1 \cap A_2)^c \cap A_1 = A_1 \cap A_2^c) \\ &= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c). \\ &\qquad \text{(as } A_1 \in \mathscr{M})\end{aligned}$$

This proves that $A_1 \cap A_2 \in \mathscr{M}$.

Step 4: We finish the proof.

It remains to prove that $\mathscr{R} \subset \mathscr{M}$, for this implies $\sigma(\mathscr{R}) \subset \mathscr{M}$ and thus, since μ^* is a measure on $\mathscr{M}$ by Step 3, it is a measure on $\sigma(\mathscr{R})$ which, by Step 2, extends μ.
Thus, let $A \in \mathscr{R}$ be given. We need to prove that $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A)$ for all $B \subset \Omega$. In fact, by subadditivity, it suffices to prove that $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \setminus A)$ for all $B \subset \Omega$. If $\mu^*(B) = \infty$, there is nothing to prove. So assume that $\mu^*(B) < \infty$. Given $\varepsilon > 0$, we find a sequence (A_k) in $\mathscr{R}$ such that $B \subset \bigcup_{k \in \mathbb{N}} A_k$ and $\mu^*(B) \geq \sum_{k=1}^{\infty} \mu(A_k) - \varepsilon$. Note that $B \cap A \subset \bigcup_{k \in \mathbb{N}} (A_k \cap A)$ and $B \setminus A \subset \bigcup_{k \in \mathbb{N}} (A_k \setminus A)$, and that $A_k \cap A$ and $A_k \setminus A$ belong to $\mathscr{R}$ for all $k \in \mathbb{N}$. It follows that

$$\begin{aligned} \mu^*(B \cap A) + \mu^*(B \setminus A) &\leq \sum_{k=1}^{\infty} \mu(A_k \cap A) + \sum_{k=1}^{\infty} \mu(A_k \setminus A) \\ &= \sum_{k=1}^{\infty} \mu(A_k) \leq \mu^*(B) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, we are done. □

Corollary 3.43. *There exists a unique measure λ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with $\lambda([a,b)) = b - a$ for all $b > a$. Moreover, λ is translation invariant, i.e., $\lambda(A) = \lambda(x + A)$ for all $x \in \mathbb{R}$ and $A \in \mathscr{B}(\mathbb{R})$.*

Proof. Apply Theorem 3.42 in Example 3.41 to obtain the existence of such a measure. Uniqueness follows from Corollary 3.38. The translation invariance follows from the definition of λ^* in the proof of Theorem 3.42. □

Definition 3.44. The measure λ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with $\lambda([a,b)) = b - a$ for all $b > a$ is called the **Lebesgue measure**.

Remark 3.45. Note that Step 3 in the proof of Theorem 3.42 actually provides a translation invariant measure λ^* on a σ-algebra $\mathscr{M}$ that contains $\mathscr{B}(\mathbb{R})$. The elements of $\mathscr{M}$ are called the **Lebesgue measurable** sets. It is a natural to ask whether $\mathscr{M} = \mathscr{P}(\mathbb{R})$, i.e., whether every subset of $\mathbb{R}$ is Lebesgue measurable. The answer to this question is negative. Using the axiom of choice, one can construct (highly pathological) subsets of $\mathbb{R}$ that are not Lebesgue measurable. In particular, not all subsets of $\mathbb{R}$ are Borel measurable.

The Banach–Tarski paradox, where the closed unit ball in three dimensions is partitioned into 5 pieces that are rearranged by translation and rotation into two copies of the closed unit ball, is connected to the existence of non-Lebesgue measurable sets in $\mathbb{R}^3$.

It is also possible to define a d-dimensional version of the Lebesgue measure.

Example 3.46. Similarly as in the one-dimensional situation, one can prove that there exists a unique measure λ_d on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that

$$\lambda_d([a_1, b_1) \times \cdots \times [a_d, b_d)) = (b_1 - a_1) \cdot \cdots \cdot (b_d - a_d).$$

λ_d is called ***d*-dimensional Lebesgue measure**. Again, the measure λ_d is translation invariant.

The Lebesgue–Stieltjes measure is used when integrating with respect to a distribution function F in probability theory. It is instructional to consider a few examples of measures associated with distribution functions, including distribution functions with jump discontinuities.

Exercise 3.47 (Lebesgue–Stieltjes measures)**.** Let $F\colon \mathbb{R} \to \mathbb{R}$ be a monotonically increasing function, i.e., $F(t) \leq F(s)$ if $t \leq s$. Define $F_+(t) := \inf\{F(s) : s > t\}$. Show that there exists a unique measure μ on $\mathscr{B}(\mathbb{R})$ such that $\mu((a,b]) = F_+(b) - F_+(a)$ for all $a \leq b$.

Hint: The uniqueness of such a measure follows from Corollary 3.38 as the collection $\{(a,b] : a \leq b\}$ is a generator of $\mathscr{B}(\mathbb{R})$ that is stable under intersection. The existence follows from Caratheodory's theorem after establishing that μ extended to finite unions of left half-open intervals is a pre-measure.

3.6 Measurable functions

We have already defined the notion of a measurable map from one measurable space (Ω_1, Σ_1) to a second measurable space (Ω_2, Σ_2). Of particular importance is the situation where $(\Omega_2, \Sigma_2) = (\mathbb{K}, \mathscr{B}(\mathbb{K}))$. Here, $\mathbb{K}$ is as before either $\mathbb{R}$ or $\mathbb{C}$ and $\mathscr{B}(\mathbb{K})$ is the Borel σ-algebra generated by the topology associated with $|\cdot|$.

Definition 3.48. Let (Ω, Σ) be a measurable space. A **measurable function** is a measurable map from (Ω, Σ) to $(\mathbb{K}, \mathscr{B}(\mathbb{K}))$.

Example 3.49. Let Ω be a set and $A \subset \Omega$. The **indicator function of A** is the function $\mathbb{1}_A\colon \Omega \to \mathbb{R}$ defined by $\mathbb{1}_A(x) = 1$ iff $x \in A$ and $\mathbb{1}_A(x) = 0$ iff $x \notin A$.

If (Ω, Σ) is a measurable space, then $\mathbb{1}_A$ is a measurable function if and only if $A \in \Sigma$. Indeed, if $S \in \mathscr{B}(\mathbb{R})$, then

$$\mathbb{1}_A^{-1}[S] = \begin{cases} \emptyset, & \text{if } S \cap \{0,1\} = \emptyset, \\ A, & \text{if } 1 \in S \text{ and } 0 \notin S, \\ A^{\mathrm{c}}, & \text{if } 1 \notin S \text{ and } 0 \in S, \\ \Omega, & \text{if } 1 \in S \text{ and } 0 \in S. \end{cases}$$

Proposition 3.50. *Let (Ω, Σ) be a measurable space, $f, g\colon \Omega \to \mathbb{K}$ be measurable and $\lambda \in \mathbb{K}$. Then λf, $f \cdot g$, $f + g$ and, if $\mathbb{K} = \mathbb{R}$, $f \vee g := \max\{f, g\}$ and $f \wedge g := \min\{f, g\}$ are measurable. Moreover, if (f_n) is a sequence of measurable functions from Ω to $\mathbb{K}$ which converges pointwise to a function $f\colon \Omega \to \mathbb{K}$, i.e. $f_n(x) \to f(x)$ for all $x \in \Omega$, then f is measurable.*

Proof. We define $\Phi\colon \Omega \to \mathbb{K}^2$ by $\Phi(x) = (f(x), g(x))$. As with $\mathbb{K}$, we also equip $\mathbb{K}^2$ with the Borel σ-algebra. Using Proposition 3.15, it is easily checked that Φ is measurable by looking at the preimages of balls in the $\|\cdot\|_\infty$ norm on $\mathbb{K}^2$ (Exercise!). Moreover, addition, multiplication and, if $\mathbb{K} = \mathbb{R}$, taking maximum and minimum of two numbers in $\mathbb{R}^2$, are continuous maps from $\mathbb{K}^2$ to $\mathbb{K}$ and hence measurable. Thus the first assertion follows since the composition of measurable maps is measurable.

For the second assertion, let C be a closed subset of $\mathbb{K}$. We set $C_k := \{x \in \mathbb{K} : d(x, C) \le \frac{1}{k}\}$. Note that C_k is a closed set for all $k \in \mathbb{N}$. We claim that

$$f^{-1}[C] = \bigcap_{k\in\mathbb{N}} \bigcup_{n_0\in\mathbb{N}} \bigcap_{n\ge n_0} f_n^{-1}[C_k].$$

Note that this finishes the proof since the set on the right-hand side belongs to Σ.

It remains to prove the claim. First assume $x \in f^{-1}[C]$. Then $f(x) \in C$. Since $f_n(x) \to f(x)$, given $k \in \mathbb{N}$ there exists an $n_0 \in \mathbb{N}$ such that $|f_n(x) - f(x)| \le k^{-1}$ for all $n \ge n_0$. Thus, $f_n(x) \in C_k$ for all $n \ge n_0$, proving that x belongs to the set on the right-hand side.

Conversely, assume that x belongs to the set on the right-hand side. This means that for all $k \in \mathbb{N}$ there exists an $n_0 \in \mathbb{N}$ such that $f_n(x) \in C_k$ for all $n \ge n_0$. Since C_k is closed and $f_n(x) \to f(x)$, it follows that $f(x) \in C_k$ for all $k \in \mathbb{N}$. But then $f(x) \in \bigcap_{k\in\mathbb{N}} C_k = C$. □

Definition 3.51. Let (Ω, Σ) be a measurable space. A **simple function** is a measurable function $f\colon \Omega \to \mathbb{K}$ taking only finitely many values.

The following lemma gives a description of simple functions.

Lemma 3.52. *Let (Ω, Σ) be a measurable space and $f\colon \Omega \to \mathbb{K}$ be a simple function. Suppose that $a_1, \dots, a_n$ are the finitely many values that f takes. Then*

$$f(x) = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k}$$

*where $A_k = f^{-1}[\{a_k\}] \in \Sigma$. Note that in this case the sets A_k are disjoint and satisfy $\Omega = \bigcup_{k=1}^{n} A_k$. We call this the **standard representation of f**.*

It is clear that the standard representation of a simple function is unique up to reordering.

Proposition 3.53. *Let (Ω, Σ) be a measurable space and $f\colon \Omega \to \mathbb{K}$ be a measurable function. Then there exists a sequence of simple functions $f_n\colon \Omega \to \mathbb{K}$ with $|f_n(x)| \le 2|f(x)|$ for all $n \in \mathbb{N}$ which converges pointwise to f. Moreover, if $\mathbb{K} = \mathbb{R}$, then f_n can be chosen to be real functions. If $f \ge 0$, then the sequence can be chosen to consist of positive functions and to be increasing, i.e. $f_n(x) \le f_{n+1}(x) \uparrow f(x)$.*

Proof. Let us first consider the case where $\mathbb{K} = \mathbb{R}$. For $n \in \mathbb{N}$ and $k \in \mathbb{N}_0$, define

$$A_{n,k} := [k2^{-n}, (k+1)2^{-n}) \quad \text{and} \quad B_{n,k} := [-(k+1)2^{-n}, -k2^{-n}).$$

Note that for all $n \in \mathbb{N}$ one has $\mathbb{R} = \bigcup_{k\in\mathbb{N}_0} (A_{n,k} \cup B_{n,k})$.

Now set

$$f_n := \sum_{k=0}^{2^{2n}} \frac{k}{2^n} \mathbb{1}_{f^{-1}[A_{n,k}]} - \sum_{k=0}^{2^{2n}} \frac{k}{2^n} \mathbb{1}_{f^{-1}[B_{n,k}]}$$

Then f_n is a simple function and positive whenever f is positive (the latter follows from the fact that in this case $f^{-1}[B_{n,k}] = \emptyset$ for all k, n).

Moreover, $f_n(x) \to f(x)$. Indeed, if $x \in \Omega$, then there exists an $n_0 \in \mathbb{N}$ such that $|f(x)| \leq 2^{n_0}$. Then $|f_n(x) - f(x)| \leq 2^{-n}$ for all $n \geq n_0$.

If $f \geq 0$ then $f_n \leq f_{n+1}$. Indeed, if $f_n(x) = k2^{-n}$ then $f(x) \in [k2^{-n}, (k+1)2^{-n})$. But then either

$$f(x) \in [(2k)2^{-(n+1)}, (2k+1)2^{-(n+1)}),$$

in which case $f_{n+1}(x) = (2k)2^{-(n+1)} = f_n(x)$, or

$$f(x) \in [(2k+1)2^{-(n+1)}, (2k+2)2^{-(n+1)}),$$

in which case $f_{n+1}(x) = (2k+1)2^{-(n+1)} > f_n(x)$.

In the case where $\mathbb{K} = \mathbb{C}$, we find sequences of simple functions (g_n) and (h_n) that converge to $\operatorname{Re} f$ and $\operatorname{Im} f$, respectively. Then we set $f_n := g_n + ih_n$ for all $n \in \mathbb{N}$. □

We sometimes prefer to work with a slightly more general notion of measurable functions. It has technical advantages to allow functions to take the values ∞ or $-\infty$.

For example, this is natural for stopping times of random processes.

Remark 3.54 (Extended real line)**.** We put $\overline{\mathbb{R}} := \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ which we endow with the σ-algebra $\mathscr{B}(\overline{\mathbb{R}})$, defined as $\sigma(\mathscr{B}(\mathbb{R}) \cup \{\{-\infty\}, \{\infty\}\})$. It follows that a function $f\colon (\Omega, \Sigma) \to \overline{\mathbb{R}}$ is measurable if and only if $f^{-1}[\{\infty\}], f^{-1}[\{-\infty\}] \in \Sigma$ and $f^{-1}[A] \in \Sigma$ for all $A \in \mathscr{B}(\mathbb{R})$.

Similarly, $\mathscr{B}([0,\infty])$ is defined as $\sigma(\mathscr{B}([0,\infty)) \cup \{\infty\})$. We also remark that Proposition 3.53 generalizes to this situation.

There exist metrics on $\overline{\mathbb{R}}$ resp. $[0,\infty]$ such that $\mathscr{B}(\overline{\mathbb{R}})$ resp. $\mathscr{B}([0,\infty])$ is the Borel σ-algebra for this metric.

Exercise 3.55. Let (Ω, Σ) be a measurable space and $f_n\colon \Omega \to \overline{\mathbb{R}}$ be measurable for all $n \in \mathbb{N}$. Then the functions $\liminf_{n\to\infty} f_n$, $\limsup_{n\to\infty} f_n$, $\inf_{n\in\mathbb{N}} f_n$ and $\sup_{n\in\mathbb{N}} f_n$ are measurable.

Hint: Observe that $\{x \in \Omega : \sup\{f_n(x) : n \in \mathbb{N}\} > a\} = \bigcup_{n\in\mathbb{N}}\{x \in \Omega : f_n(x) > a\}$.

3.7 The Lebesgue integral

Given a measure space (Ω, Σ, μ), we now introduce the Lebesgue integral $\int_\Omega f \,\mathrm{d}\mu$ for suitable complex-valued, measurable functions that we call 'integrable'. We proceed in several steps and first define the integral for (real-valued) measurable functions f taking values in $[0,\infty]$.

Definition 3.56. Let (Ω, Σ, μ) be a measure space. If $f\colon \Omega \to [0,\infty]$ is a simple function, with standard representation $f = \sum_{k=1}^n a_k \mathbb{1}_{A_k}$, one defines the **Lebesgue integral** of f by

$$\int_\Omega f \,\mathrm{d}\mu := \sum_{k=1}^n a_k \mu(A_k)$$