Lecture notes
Probability and Measure

Gábor Szabó
KU Leuven – G0P63a

Academic Year 2022–2023, Semester 1

Contents

1 Lebesgue's Integration Theory

1.1 σ-Algebras and Measurable Spaces

Definition 1.1. Let Ω be a non-empty set. A σ-algebra $\mathfrak{M}$ on Ω is a set of subsets of Ω satisfying

A1. $\emptyset \in \mathfrak{M}$.

A2. If $E \in \mathfrak{M}$, then $E^c \in \mathfrak{M}$.

A3. For any sequence $E_n \in \mathfrak{M}$, one has $\bigcup_{n\in\mathbb{N}} E_n \in \mathfrak{M}$.

We will refer to the pair $(\Omega, \mathfrak{M})$ as a measurable space.

Remark 1.2. A σ-algebra $\mathfrak{M}$ is closed with respect to all countable operations on sets one can perform using complement, union, intersection, and difference. In fact, any intersection and difference can be rewritten in terms of unions and complements, namely

$$E \cap F = (E^c \cup F^c)^c, \quad E \setminus F = (E^c \cup F)^c.$$

Example 1.3. The entire power set 2^Ω of Ω is the largest possible σ-algebra on Ω, whereas $\{\emptyset, \Omega\}$ is the smallest one. We may also consider

$$\mathfrak{M} = \{E \subseteq \Omega \mid E \text{ or } E^c \text{ is countable}\},$$

which sits strictly in between if Ω is an uncountably infinite set.

Proposition 1.4. *If $\{\mathfrak{M}_i\}_{i\in I}$ is an arbitrary family of σ-algebras on a set Ω, then the intersection $\bigcap_{i\in I} \mathfrak{M}_i$ is a σ-algebra. Consequently, if $\mathcal{S} \subseteq 2^\Omega$ is a family of subsets of Ω, then there exists a smallest σ-algebra $\mathfrak{M}$ containing $\mathcal{S}$. We will refer to it as the σ-algebra generated by $\mathcal{S}$.*

Proof. It is a trivial exercise to prove the first part of the statement. For the second part, if $\mathcal{S}$ is given, consider the family

$$\left\{\mathfrak{M} \subseteq 2^\Omega \mid \mathfrak{M} \text{ is a } \sigma\text{-algebra with } \mathcal{S} \subseteq \mathfrak{M}\right\}.$$

Since this family contains 2^Ω, it is non-empty, and hence the intersection of this family is the smallest possible σ-algebra that contains $\mathcal{S}$. $\square$

Example 1.5. Recall that a topology $\mathcal{T}$ on a set Ω is a family of subsets of Ω satisfying

T1. $\emptyset, \Omega \in \mathcal{T}$.

T2. For any family of sets $\mathcal{O} \subseteq \mathcal{T}$, one has $\bigcup \mathcal{O} \in \mathcal{T}$.

T3. For $O_1, O_2 \in \mathcal{T}$, one has $O_1 \cap O_2 \in \mathcal{T}$.

If we are given a topological space $(\Omega, \mathcal{T})$, then the σ-algebra generated by $\mathcal{T}$ is called the Borel-σ-algebra of $(\Omega, \mathcal{T})$. Its elements are called the Borel subsets of Ω.

Remark 1.6. Given a topology $\mathcal{T}$ as above, a base $\mathcal{B} \subseteq \mathcal{T}$ is a subset with the property that every set $O \in \mathcal{T}$ can be expressed as a union of a family of sets in $\mathcal{B}$. If $\mathcal{T}$ has a countable base $\mathcal{B}$, then every σ-algebra $\mathfrak{M}$ containing $\mathcal{B}$ will necessarily also contain $\mathcal{T}$. It follows that the Borel-σ-algebra generated by $\mathcal{T}$ coincides with the σ-algebra generated by $\mathcal{B}$.

Proposition 1.7. *Let (Ω, d) be a separable metric space. Then Ω has a countable base consisting of open balls, i.e., sets of the form*

$$B(x, r) = \{y \in \Omega \mid d(x, y) < r\}$$

for $x \in \Omega$ and $r > 0$. In particular, the Borel-σ-algebra on Ω coincides with the σ-algebra generated by all open balls.

Proof. In light of the previous remark, the second part of the statement follows automatically if we can show that the metric topology on Ω has a countable base consisting of open balls. As we assumed Ω to be separable, choose a countable dense set $D \subseteq \Omega$, and consider

$$\mathcal{B} = \{B(x, r) \mid x \in D,\ 0 < r \in \mathbb{Q}\}.$$

This is a countable family of open balls, and we claim that it is a base for the metric topology. Indeed, let $O \subseteq \Omega$ be an open set. For $x \in D \cap O$, set $R_x = \{0 < r \in \mathbb{Q} \mid B(x, r) \subseteq O\}$. Then evidently

$$\bigcup \{B(x, r) \mid x \in D \cap O, r \in R_x\} \subseteq O.$$

We claim that this is an equality: Given $y \in O$, it follows from the definition of openness that there exists $\delta > 0$ with $B(y, \delta) \subseteq O$. By making δ smaller, if necessary, we may assume $\delta \in \mathbb{Q}$. Since D is a dense subset of Ω, there exists $x \in D$ with $d(x, y) < \delta/2$. By the triangle inequality, we have $y \in B(x, \delta/2) \subseteq B(y, \delta) \subseteq O$. In summary, we have shown that O is the countable union of sets in $\mathcal{B}$, which finishes the proof. $\square$

Example 1.8. Let us consider $\Omega = \mathbb{R}$ with the standard topology. Then half-open intervals are Borel because, for example, for $a, b \in \mathbb{R}$ with $a < b$, one has

$$[a, b) = \bigcap_{n \in \mathbb{N}} (a - \frac{1}{n}, b).$$

Singleton sets are also Borel since they are closed. To summarize, there are in general many more Borel sets than open sets. Note that in light of the previous remark, the Borel-σ-algebra on $\mathbb{R}$ is generated by all open intervals.

1.2 Measurable functions

Definition 1.9. Let $(\Omega_1, \mathfrak{M}_1)$ and $(\Omega_2, \mathfrak{M}_2)$ be two measurable spaces. A map $f : \Omega_1 \to \Omega_2$ is called measurable, if for every subset $E \subseteq \Omega_2$, $E \in \mathfrak{M}_2$ implies $f^{-1}(E) \in \mathfrak{M}_1$.

Remark 1.10. The composition of two measurable maps is always measurable.

Proposition 1.11. *Let $(\Omega_1, \mathfrak{M}_1)$ and $(\Omega_2, \mathfrak{M}_2)$ be two measurable spaces and $f : \Omega_1 \to \Omega_2$ a map. Suppose that $\mathfrak{M}_2$ is the σ-algebra generated by a set $\mathcal{S} \subseteq 2^{\Omega_2}$. Then f is measurable if and only if for all $O \in \mathcal{S}$, one has $f^{-1}(O) \in \mathfrak{M}_1$.*

Proof. The "only if" part is trivial, so let us consider the "if part". Consider the set

$$\mathfrak{M} = \left\{ O \subseteq \Omega_2 \mid f^{-1}(O) \in \mathfrak{M}_1 \right\}.$$

By assumption, we have $\mathcal{S} \subseteq \mathfrak{M}$. The claim amounts to showing that $\mathfrak{M}_2 \subseteq \mathfrak{M}$, and by assumption on $\mathcal{S}$, it therefore suffices to show that $\mathfrak{M}$ is a σ-algebra. But this will be part of the exercise sessions.

□

Proposition 1.12. *Let $(\Omega, \mathfrak{M})$ be a measurable space and $(Y, \mathcal{T})$ a topological space. For a map $f : \Omega \to Y$, the following are equivalent:*

(i) f is measurable with respect to the Borel-σ-algebra on Y.

(ii) For all $O \in \mathcal{T}$, one has $f^{-1}(O) \in \mathfrak{M}$.

If furthermore there exists a countable base $\mathcal{B} \subseteq \mathcal{T}$, then this is equivalent to

(iii) For all $O \in \mathcal{B}$, one has $f^{-1}(O) \in \mathfrak{M}$.

Proof. By definition of the Borel-σ-algebra as the one generated by $\mathcal{T}$, the equivalence (i)$\Leftrightarrow$(ii) becomes a special case of Proposition 1.11. If we furthermore assume that $\mathcal{B}$ is a countable base of $\mathcal{T}$, then the equivalence (i)$\Leftrightarrow$(iii) is also a special case of Proposition 1.11 in light of Remark 1.6. $\square$

Theorem 1.13. *Let $(\Omega, \mathfrak{M})$ be a measurable space and (Y, d) a metric space. We equip Y with the Borel-σ-algebra associated to the metric topology. Suppose that a sequence of measurable functions $f_n : \Omega \to Y$ converges to a map $f : \Omega \to Y$ pointwise. Then f is measurable.*

Proof. As a consequence of Proposition 1.12, it suffices to show that the preimages of open sets under f belong to $\mathfrak{M}$. Since preimages respect complements of sets, it suffices to show that the preimage of every closed set $C \subseteq Y$ belongs to $\mathfrak{M}$. Since we have a metric space, we have

$$C = \overline{C} = \bigcap_{k\in\mathbb{N}} \underbrace{\left\{ y \in Y \;\middle|\; \inf_{x\in C} d(x,y) < \frac{1}{k} \right\}}_{=:C_k}$$

Then each of the sets C_k is open. For every $x \in \Omega$, we use $f_n(x) \to f(x)$ and observe

$$\begin{aligned}
&x \in f^{-1}(C) \\
&\quad\Leftrightarrow \quad f(x) \in C \\
&\quad\Leftrightarrow \quad \forall\, k \in \mathbb{N} : f(x) \in C_k \\
&\overset{f_n(x)\to f(x)}{\Leftrightarrow} \quad \forall\, k \in \mathbb{N} : \exists\, n_0 \in \mathbb{N} : \forall\, n \geq n_0 : f_n(x) \in C_k \\
&\quad\Leftrightarrow \quad x \in \bigcap_{k\in\mathbb{N}} \bigcup_{n_0\in\mathbb{N}} \bigcap_{n\geq n_0} f_n^{-1}(C_k).
\end{aligned}$$

In summary, the preimage $f^{-1}(C)$ can be realized as a countable union of countable intersections of sets in $\mathfrak{M}$, and hence also belongs to $\mathfrak{M}$. $\square$

Notation 1.14. We will equip $[0,\infty] := [0,\infty) \cup \{\infty\}$ with the topology of the one point compactification, that is, we define a subset $O \subseteq [0,\infty]$ to be open when the following holds:

- if $0 \in O$, then there exists $\varepsilon > 0$ with $[0,\varepsilon) \subseteq O$.
- for every $x \in O \cap (0,\infty)$, there exists $\varepsilon > 0$ with $(x-\varepsilon, x+\varepsilon) \subseteq O$.
- if $\infty \in O$, then there is $b \geq 0$ with $(b,\infty] \subseteq O$.

This topology is induced by a metric, for example by

$$d(x,y) = \left| \frac{1}{1+x} - \frac{1}{1+y} \right|,$$

where we follow the convention $\frac{1}{\infty} := 0$. We extend the usual addition and multiplication from $[0, \infty)$ to $[0, \infty]$ by defining[1]

$$x + \infty := \infty, \quad x \cdot \infty := \infty \ (x > 0), \quad 0 \cdot \infty := 0.$$

Then the addition map $+ : [0, \infty] \times [0, \infty] \to [0, \infty]$ is continuous, but this is not true for the multiplication map.[2] We also extend the usual order relation "$\leq$" of numbers to $[0, \infty]$ in the obvious way.

1.3 Measures on σ-algebras, Measure Spaces

Definition 1.15. Let $(\Omega, \mathfrak{M})$ be a measurable space. A (positive) measure μ on $(\Omega, \mathfrak{M})$ is a map $\mathfrak{M} \to [0, \infty]$ satisfying:

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

M2. μ is σ-additive, i.e., for every sequence $E_n \in \mathfrak{M}$ consisting of pairwise disjoint sets, one has

$$\mu\Big(\bigcup_{n \in \mathbb{N}} E_n\Big) = \sum_{n \in \mathbb{N}} \mu(E_n).$$

The triple $(\Omega, \mathfrak{M}, \mu)$ is called a measure space. If $\mu(\Omega) < \infty$, we call μ a finite measure, and if more specifically $\mu(\Omega) = 1$, we call it a probability measure and the triple $(\Omega, \mathfrak{M}, \mu)$ a probability space. If there exists a sequence $E_n \in \mathfrak{M}$ with $\mu(E_n) < \infty$ and $\Omega = \bigcup_{n \in \mathbb{N}} E_n$, then we say that μ is σ-finite.

Remark 1.16. Note that

- the series $\sum_{n \in \mathbb{N}} \mu(E_n)$ is a series in $[0, \infty]$, which we may define as the supremum $\sup_{N \geq 1} \sum_{n=1}^{N} \mu(E_n)$.
- σ-additivity implies finite additivity: If $E_1, E_2 \in \mathfrak{M}$ are two disjoint sets, then

$$\begin{aligned} \mu(E_1 \cup E_2) &= \mu(E_1 \cup E_2 \cup \emptyset \cup \emptyset \cup \dots) \\ &= \mu(E_1) + \mu(E_2) + \underbrace{\mu(\emptyset) + \mu(\emptyset) + \dots}_{=0} \\ &= \mu(E_1) + \mu(E_2). \end{aligned}$$

[1] Keep in mind that we implicitly force everything to be commutative, so the order of addition and multiplication does not matter here by default.

[2] Convince yourself why this is not the case!

- If there is at least some $E \in \mathfrak{M}$ with $\mu(E) < \infty$, then σ-additivity already implies $\mu(\emptyset) = 0$:

$$\begin{aligned} \mu(E) &= \mu(E \cup \emptyset \cup \emptyset \cup \dots) \\ &= \mu(E) + \infty \cdot \mu(\emptyset). \end{aligned}$$

  If $\mu(E) < \infty$, then the above can only happen when $\mu(\emptyset) = 0$.

Example 1.17 (Counting measure)**.** For any non-empty set Ω, we can consider the σ-algebra 2^Ω and define a measure μ via

$$\mu(E) = \begin{cases} \#E & , \quad E \text{ is finite} \\ \infty & , \quad E \text{ is infinite.} \end{cases}$$

Example 1.18 (Dirac measure)**.** If $(\Omega, \mathfrak{M})$ is any measurable space with a distinguished point $x \in \Omega$, one can define the measure δ_x via

$$\delta_x(E) = \begin{cases} 1 & , \quad x \in E \\ 0 & , \quad x \notin E. \end{cases}$$

Example 1.19. Consider $\mathbb{R}$ with its Borel-σ-algebra $\mathfrak{B}$. We will later construct the Lebesgue measure on $\mathfrak{B}$, which is a unique measure $\mu : \mathfrak{B} \to [0, \infty]$ with the property that $\mu([a, b]) = b - a$ for all $a, b \in \mathbb{R}$ with $a \leq b$. An analogous unique measure exists on $\mathbb{R}^n$ as well.

Proposition 1.20. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. Then μ has the following extra properties:*

(i) *μ is monotone, i.e., for $E, F \in \mathfrak{M}$ with $E \subseteq F$ one has $\mu(E) \leq \mu(F)$. If $\mu(F) < \infty$, then $\mu(E) = \mu(F) - \mu(F \setminus E)$.*

(ii) *μ is subadditive, i.e., for $E, F \in \mathfrak{M}$ one has $\mu(E \cup F) \leq \mu(E) + \mu(F)$.*

(iii) *μ is σ-subadditive, i.e., for a sequence $E_n \in \mathfrak{M}$ one has $\mu(\bigcup_{n \in \mathbb{N}} E_n) \leq \sum_{n \in \mathbb{N}} \mu(E_n)$.*

(iv) *If $E_n \in \mathfrak{M}$ is an increasing sequence (w.r.t. "$\subseteq$"), then $\mu(\bigcup_{n \in \mathbb{N}} E_n) = \sup_{n \in \mathbb{N}} \mu(E_n)$.*

(v) *If $E_n \in \mathfrak{M}$ is a decreasing sequence and $\mu(E_1) < \infty$, then $\mu(\bigcap_{n \in \mathbb{N}} E_n) = \inf_{n \in \mathbb{N}} \mu(E_n)$.*

Proof. (i): If $E \subseteq F$, then we may write $F = E \cup (F \setminus E)$ as a disjoint union, so it follows from additivity that $\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E)$. If $\mu(F) < \infty$, we also obtain the last part of the claim by substracting $\mu(F \setminus E)$.

(ii): One has $E \cup F = (E \setminus F) \cup F$, which is a disjoint union. Thus $\mu(E \cup F) = \mu(E \setminus F) + \mu(F) \leq \mu(E) + \mu(F)$.

(iii)+(iv): We construct a sequence $F_n \in \mathfrak{M}$ as follows. We set $F_1 = E_1$ and $F_n = E_n \setminus (\bigcup_{k<n} E_k)$ for $n \geq 2$. Then the sequence F_n consists of pairwise disjoint sets, but at the same time one has that $\bigcup_{k \leq n} E_k = \bigcup_{k \leq n} F_k$ for all $n \geq 1$. Hence $\mu(\bigcup_{n \in \mathbb{N}} E_n) = \mu(\bigcup_{n \in \mathbb{N}} F_n) = \sum_{n \in \mathbb{N}} \mu(F_n) \leq \sum_{n \in \mathbb{N}} \mu(E_n)$, which proves (iii). For (iv), additionally assume that E_n was increasing. Then

$$\begin{aligned} \mu\Big(\bigcup_{n \in \mathbb{N}} E_n \Big) &= \sum_{n \in \mathbb{N}} \mu(F_n) \\ &= \sup_{n \in \mathbb{N}} \sum_{k \leq n} \mu(F_k) \\ &= \sup_{n \in \mathbb{N}} \mu\Big(\bigcup_{k \leq n} F_k \Big) \\ &= \sup_{n \in \mathbb{N}} \mu(E_n). \end{aligned}$$

(v): Consider $F_n = E_1 \setminus E_n$ for all $n \geq 1$. As E_n was assumed to be decreasing, the sets F_n will be increasing, and moreover $\bigcup_{n \in \mathbb{N}} F_n = E_1 \setminus \Big(\bigcap_{n \in \mathbb{N}} E_n \Big)$. By using (i) and (iv), we see

$$\begin{aligned} \mu(E_1) - \inf_{n \in \mathbb{N}} \mu(E_n) &= \sup_{n \in \mathbb{N}} \mu(F_n) \\ &= \mu\Big(\bigcup_{n \in \mathbb{N}} F_n \Big) \\ &= \mu(E_1) - \mu\Big(\bigcap_{n \in \mathbb{N}} E_n \Big) \end{aligned}$$

This implies the claim. □

Remark 1.21. In condition (v) above, it is really necessary to assume that at least one of the sets E_n has finite measure. For example, we may consider $\mathbb{R}$ with the counting measure μ, and the sets $E_n = (0, \frac{1}{n})$. Then $\mu(E_n) = \infty$ for all n, the sequence is decreasing, and $\bigcap_{n \in \mathbb{N}} E_n = \emptyset$.

We omit the proof of the following statement, which is a very simple exercise:

Proposition 1.22. *Let* $(\Omega, \mathfrak{M})$ *be a measurable space with two measures* μ_1, μ_2. *For any numbers* $\alpha_1, \alpha_2 \geq 0$, *the map*

$$\alpha_1 \mu_1 + \alpha_2 \mu_2 : \mathfrak{M} \to [0, \infty], \quad E \mapsto \alpha_1 \mu_1(E) + \alpha_2 \mu_2(E)$$

is a measure.

Definition 1.23. Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. A subset $N \subseteq \Omega$ is called a null set if there is some $E \in \mathfrak{M}$ with $N \subseteq E$ and $\mu(E) = 0$.

If $P(x)$ is a statement about elements $x \in \Omega$ that can either be true or false, we say that P holds almost everywhere (w.r.t. μ) if the set of elements $x \in \Omega$ for which $P(x)$ is false is a null set.

Notation 1.24. For brevity, we may also say "P holds a.e." (a.e.=almost everywhere) or "$P(x)$ holds for (μ-)almost all x".

1.4 Integration theory

Notation 1.25. For a set Ω with a subset $E \subseteq \Omega$, its indicator function (or characteristic function) is defined via

$$\chi_E : \Omega \to \{0, 1\}, \quad \chi_E(x) = \begin{cases} 1 & , \quad x \in E \\ 0 & , \quad x \notin E. \end{cases}$$

Definition 1.26. Let Ω be a set. A simple function (or step function) on Ω is a function with finite range. In particular, a simple function $s : \Omega \to Y$ with $Y \in \{[0, \infty), [0, \infty], \mathbb{R}, \mathbb{C}\}$ can always be written as

$$s = \sum_{\lambda \in s(\Omega)} \lambda \cdot \chi_{s^{-1}(\lambda)}.$$

We will refer to this as the canonical form for s.

Proposition 1.27. *Let $(\Omega, \mathfrak{M})$ be a measurable space. Let $s : \Omega \to Y$ be a simple function, and assume that Y is equipped with a σ-algebra that contains all singleton sets. (In particular this is the case for $Y \in \{[0, \infty), [0, \infty], \mathbb{R}, \mathbb{C}\}$.) Then s is measurable if and only if $s^{-1}(\lambda) \in \mathfrak{M}$ for all $\lambda \in s(\Omega)$.*

Proof. The "only if" part is clear. Since we assumed the σ-algebra on Y to contain all singleton sets, it follows in particular that $s^{-1}(y) \in \mathfrak{M}$ for all $y \in Y$.

For the "if" part, let A be a measurable subset of Y. Then $s^{-1}(A) = \bigcup_{\lambda \in s(\Omega)} s^{-1}(A \cap \{\lambda\})$. By assumption this is a finite union of sets of the form $s^{-1}(\lambda)$ for $\lambda \in s(\Omega)$, and hence if these are measurable, then so is $s^{-1}(A)$. $\square$

Definition 1.28. Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. Let $s : \Omega \to [0, \infty)$ be a measurable simple function with canonical form $s = \sum_{\lambda \in s(\Omega)} \lambda \cdot \chi_{s^{-1}(\lambda)}$. We define the integral of s over Ω with respect to μ as

$$\int_\Omega s \, d\mu := \sum_{\lambda \in s(\Omega)} \lambda \cdot \mu(s^{-1}(\lambda)).$$

(Keep in mind the convention $0 \cdot \infty = 0$.)

Lemma 1.29. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. For all $k \geq 0$, numbers $\alpha_1, \ldots, \alpha_k \geq 0$, and sets $A_1, \ldots, A_k \in \mathfrak{M}$, the function $s = \sum_{j=1}^{k} \alpha_j \cdot \chi_{A_j}$ is a positive measurable simple function with $\int_\Omega s \, d\mu = \sum_{j=1}^{k} \alpha_j \mu(A_j)$.*

Proof. It is clear that these are positive measurable simple functions, so the relevant part of the claim is the last equation. To prove it, we proceed by induction on k. Let us first consider $k = 1$, so $s = \alpha_1 \chi_{A_1}$. If $\alpha_1 = 0$, there is nothing to prove, so let us assume $\alpha_1 \neq 0$. Then the canonical form of s is $s = \alpha_1 \chi_{A_1} + 0 \cdot \chi_{A_1^c}$, and hence $\int_\Omega s \, d\mu = \alpha_1 \mu(A_1) + 0 \cdot \mu(A_1^c) = \alpha_1 \mu(A_1)$.

Now suppose that $\ell \geq 1$ is any natural number and that the statement of the Lemma is true for all $k \leq \ell$. Suppose that $s = \sum_{j=1}^{\ell+1} \alpha_j \chi_{A_j}$ is a step function with $\ell + 1$ non-zero summands. Write $s_0 = \sum_{j=1}^{\ell} \alpha_j \chi_{A_j}$. We will try to find the canonical form for the sum $s = s_0 + \alpha_{\ell+1} \chi_{A_{\ell+1}}$ in terms of the canonical form of s_0. Let $\lambda \in [0, \infty)$. Then

$$s^{-1}(\lambda) = \begin{cases} \emptyset & , \quad \lambda \notin s_0(\Omega) \cup (s_0(\Omega) + \alpha_{\ell+1}) \\ s_0^{-1}(\lambda - \alpha_{\ell+1}) \cap A_{\ell+1} & , \quad \lambda \in (s_0(\Omega) + \alpha_{\ell+1}) \setminus s_0(\Omega) \\ s_0^{-1}(\lambda) \setminus A_{\ell+1} & , \quad \lambda \in s_0(\Omega) \setminus (s_0(\Omega) + \alpha_{\ell+1}) \\ [s_0^{-1}(\lambda - \alpha_{\ell+1}) \cap A_{\ell+1}] \cup [s_0^{-1}(\lambda) \setminus A_{\ell+1}] & , \quad \lambda \in s_0(\Omega) \cap (s_0(\Omega) + \alpha_{\ell+1}) \end{cases}$$

This allows us to write the canonical form of s as

$$\begin{aligned} s \quad = \quad & \sum_{\lambda \in (s_0(\Omega) + \alpha_{\ell+1}) \setminus s_0(\Omega)} \lambda \cdot \chi_{s_0^{-1}(\lambda - \alpha_{\ell+1}) \cap A_{\ell+1}} \\ & + \sum_{\lambda \in s_0(\Omega) \setminus (s_0(\Omega) + \alpha_{\ell+1})} \lambda \cdot \chi_{s_0^{-1}(\lambda) \setminus A_{\ell+1}} \\ & + \sum_{\lambda \in s_0(\Omega) \cap (s_0(\Omega) + \alpha_{\ell+1})} \lambda \cdot \chi_{[s_0^{-1}(\lambda - \alpha_{\ell+1}) \cap A_{\ell+1}] \cup [s_0^{-1}(\lambda) \setminus A_{\ell+1}]} \end{aligned}$$

We compute its integral as follows by using our computation rules for measures and the fact that $\Omega = \bigsqcup_{\lambda \in s_0(\Omega)} s_0^{-1}(\lambda)$:

$$\begin{aligned} \int_\Omega s \, d\mu \quad = \quad & \sum_{\lambda \in (s_0(\Omega) + \alpha_{\ell+1}) \setminus s_0(\Omega)} \lambda \cdot \mu(s_0^{-1}(\lambda - \alpha_{\ell+1}) \cap A_{\ell+1}) \\ & + \sum_{\lambda \in s_0(\Omega) \setminus (s_0(\Omega) + \alpha_{\ell+1})} \lambda \cdot \mu(s_0^{-1}(\lambda) \setminus A_{\ell+1}) \\ & + \sum_{\lambda \in s_0(\Omega) \cap (s_0(\Omega) + \alpha_{\ell+1})} \lambda \cdot \mu\Big([s_0^{-1}(\lambda - \alpha_{\ell+1}) \cap A_{\ell+1}] \cup [s_0^{-1}(\lambda) \setminus A_{\ell+1}]\Big) \\ = \quad & \sum_{\lambda \in s_0(\Omega)} \lambda \cdot \mu(s_0^{-1}(\lambda) \setminus A_{\ell+1}) + \sum_{\lambda \in s_0(\Omega)} (\lambda + \alpha_{\ell+1}) \cdot \mu(s_0^{-1}(\lambda) \cap A_{\ell+1}) \\ = \quad & \sum_{\lambda \in s_0(\Omega)} \lambda \cdot \mu(s_0^{-1}(\lambda)) + \alpha_{\ell+1} \cdot \mu(A_{\ell+1}) \\ = \quad & \int_\Omega s_0 \, d\mu + \alpha_{\ell+1} \mu(A_{\ell+1}). \end{aligned}$$

By induction hypothesis, the latter sum is equal to

$$\sum_{j=1}^{\ell} \alpha_j \mu(A_j) + \alpha_{\ell+1}\mu(A_{\ell+1}) = \sum_{j=1}^{\ell+1} \alpha_j \mu(A_j),$$

which shows our claim. $\square$

Proposition 1.30. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. Then the assignment $s \mapsto \int_\Omega s \, d\mu$, which assigns to every measurable step function $\Omega \to [0, \infty)$ its integral, satisfies the following properties:*

(i) $\displaystyle\int_\Omega (c \cdot s) \, d\mu = c \cdot \int_\Omega s \, d\mu$ *for all constants* $c \geq 0$.

(ii) $\displaystyle\int_\Omega (s + t) \, d\mu = \int_\Omega s \, d\mu + \int_\Omega t \, d\mu.$

(iii) *If* $s \leq t$, *then* $\displaystyle\int_\Omega s \, d\mu \leq \int_\Omega t \, d\mu.$

Proof. Both (i) and (ii) are straightforward consequences of Lemma 1.29.

For (iii), suppose that s and t are given as

$$s = \sum_{j=1}^{k} \alpha_j \chi_{A_j}, \quad t = \sum_{i=1}^{\ell} \beta_i \chi_{B_i}$$

with $\Omega = \bigsqcup_{j=1}^{k} A_j = \bigsqcup_{i=1}^{\ell} B_i$, for example via their canonical forms. Then we can also write

$$s = \sum_{j=1}^{k} \alpha_j \cdot \sum_{i=1}^{\ell} \chi_{A_j \cap B_i}, \quad t = \sum_{i=1}^{\ell} \beta_i \cdot \sum_{j=1}^{k} \chi_{A_j \cap B_i}.$$

Now clearly $s \leq t$ implies that $\alpha_j \leq \beta_i$ whenever $A_j \cap B_i \neq \emptyset$. Hence it follows from Lemma 1.29 applied to the above equalities that $\int_\Omega s \, d\mu \leq \int_\Omega t \, d\mu$. $\square$

Proposition 1.31. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space and $s : \Omega \to [0, \infty)$ a measurable simple function. Then the map $\nu : \mathfrak{M} \to [0, \infty], E \mapsto \int_\Omega s\chi_E \, d\mu$ defines a measure.*

Proof. Since positive linear combinations of measures are again measures, it suffices by Proposition 1.30 to consider the case where s is of the form $s = \chi_A$ for $A \in \mathfrak{M}$. In this case $\nu(E) = \int_\Omega \chi_A \chi_E \, d\mu = \int_\Omega \chi_{A \cap E} \, d\mu = \mu(A \cap E)$.

Indeed it is clear that $\nu(\emptyset) = 0$. For σ-additivity, let $E_n \in \mathfrak{M}$ be a sequence of pairwise disjoint sets, and observe

$$\begin{aligned} \nu\Big(\bigcup_{n\in\mathbb{N}} E_n\Big) &= \mu\Big(A \cap \bigcup_{n\in\mathbb{N}} E_n\Big) \\ &= \mu\Big(\bigcup_{n\in\mathbb{N}} (A \cap E_n)\Big) \\ &= \sum_{n\in\mathbb{N}} \mu(A \cap E_n) \\ &= \sum_{n\in\mathbb{N}} \nu(E_n). \end{aligned}$$

□

Definition 1.32. Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. For a measurable function $f : \Omega \to [0, \infty]$, we define its integral as

$$\int_\Omega f \, d\mu = \sup \left\{ \int_\Omega s \, d\mu \;\middle|\; s \text{ is a positive measurable step function with } s \leq f \right\}.$$

We call f integrable when $\int_\Omega f \, d\mu < \infty$.

Remark 1.33. We should first convince ourselves that the new definition of the integral does not contradict the old one in the case where f is assumed to be a step function. But indeed, $s = f$ is the largest step function with the property that $s \leq f$, so by Proposition 1.30(iii) the above supremum is indeed attained at the value $\int_\Omega s \, d\mu$ in the sense of Definition 1.28.

Secondly, we remark that the above definition a priori makes sense even when f is not assumed to be measurable. However, we will see in the exercises that the resulting notion of integral will have various undesirable properties when evaluated on non-measurable functions, for example not being additive.

Example 1.34. If δ_x is the Dirac measure associated to a point $x \in \Omega$ in a measurable space, then $\int_\Omega f \, d\delta_x = f(x)$.

If μ is the counting measure on an infinite set Ω, then $\int_\Omega f \, d\mu = \sum_{x\in\Omega} f(x)$.[3]

If μ is the (yet to be constructed) Lebesgue measure and $f : [a, b] \to [0, \infty)$ is a continuous function, then f is measurable and integrable, and its integral in the sense of Definition 1.32 will coincide with its Riemann integral.

Proposition 1.35. *Let $(\Omega, \mathfrak{M})$ be a measureable space and $f : \Omega \to [0, \infty]$ a positive function. Then f is measurable if and only if there exists a (pointwise) increasing sequence of positive measurable step functions $s_n : \Omega \to [0, \infty)$ such that $f = \sup_{n\in\mathbb{N}} s_n = \lim_{n\to\infty} s_n$.*

[3]Note that this sum might be over an uncountable index set!

Proof. We already know that the "if" part is true. For the "only if" part, assume that f is measurable. We claim that it suffices to consider the special case $f = \mathrm{id}$ as a function $[0,\infty] \to [0,\infty]$. Indeed, if we can realize $\mathrm{id} = \sup_{n\in\mathbb{N}} t_n$ for an increasing sequence of positive measurable step functions t_n, then

$$f = \mathrm{id} \circ f = \lim_{n\to\infty} t_n \circ f = \sup_{n\in\mathbb{N}} (t_n \circ f),$$

where $s_n = t_n \circ f$ is an increasing sequence of positive measurable step functions. But we can come up with the following sequence t_n, which is easily seen to do the trick:

$$t_n = n\chi_{[n,\infty]} + \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \chi_{[\frac{k-1}{2^n}, \frac{k}{2^n})}$$

□

Proposition 1.36. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space and $f : \Omega \to [0,\infty]$ a positive measurable function. Suppose that s_n is an increasing sequence of positive measurable step functions such that $f = \sup_{n\in\mathbb{N}} s_n$. Then $\int_\Omega f\ d\mu = \sup_{n\in\mathbb{N}} \int_\Omega s_n\ d\mu$.*

Proof. Since we have by assumption $s_n \leq f$ for all n, it follows from the definition of the integral that $\sup_{n\in\mathbb{N}} \int_\Omega s_n\ d\mu \leq \int_\Omega f\ d\mu$. For the reverse inequality, it suffices to prove that $c \cdot \int_\Omega s\ d\mu \leq \sup_{n\in\mathbb{N}} \int_\Omega s_n\ d\mu$ for every positive measurable step function $s \leq f$ and all constants $0 < c < 1$. For a fixed such function s and constant c, we consider the sequence of sets

$$E_n = \{x \in \Omega \mid s_n(x) \geq cs(x)\}.$$

Since s_n and s are measurable, it follows that $E_n \in \mathfrak{M}$. Since s_n converges to f pointwise and $s \leq f$, we can see that $\Omega = \bigcup_{n\in\mathbb{N}} E_n$. As in Proposition 1.31, we define a new measure ν via $\nu(E) = \int_\Omega s\chi_E\ d\mu$. Then it follows from Proposition 1.20(iv) that

$$\begin{aligned} c\int_\Omega s\ d\mu &= c\nu(\Omega) \\ &= c \cdot \sup\nolimits_{n\in\mathbb{N}} \nu(E_n) \\ &= \sup_{n\in\mathbb{N}} \int_\Omega cs\chi_{E_n}\ d\mu \\ &\leq \sup_{n\in\mathbb{N}} \int_\Omega s_n\chi_{E_n}\ d\mu \\ &\leq \sup_{n\in\mathbb{N}} \int_\Omega s_n\ d\mu. \end{aligned}$$

□

Proposition 1.37. *Let* $(\Omega, \mathfrak{M}, \mu)$ *be a measure space. Let* $f, g : \Omega \to [0, \infty]$ *be measurable functions. Then*

(i) $\displaystyle\int_\Omega (c \cdot f)\ d\mu = c \cdot \int_\Omega f\ d\mu$ *for all constants* $c \geq 0$.

(ii) $\displaystyle\int_\Omega (f + g)\ d\mu = \int_\Omega f\ d\mu + \int_\Omega g\ d\mu$.

(iii) *If* $f \leq g$, *then* $\displaystyle\int_\Omega f\ d\mu \leq \int_\Omega g\ d\mu$.

Proof. Both (i) and (iii) are trivial consequences of the definition of the integral. For (ii), take sequences s_n and t_n of positive measurable step functions that increase pointwise to f and g, respectively. Then $s_n + t_n$ increases pointwise to $f + g$. Hence it follows from Proposition 1.36 that

$$\begin{aligned} \int_\Omega f + g\ d\mu &= \sup_{n \in \mathbb{N}} \int_\Omega s_n + t_n\ d\mu \\ &= (\sup \int_\Omega s_n\ d\mu) + (\sup \int_\Omega t_n\ d\mu) \\ &= \int_\Omega f\ d\mu + \int_\Omega g\ d\mu. \end{aligned}$$

Here we have used that these suprema are in fact limits. □

Proposition 1.38. *Let* $(\Omega, \mathfrak{M}, \mu)$ *be a measure space and* $f : \Omega \to [0, \infty]$ *a measurable function. Then one has* $\int_\Omega f\ d\mu = 0$ *if and only if* $f(x) = 0$ *for* μ*-almost all* $x \in \Omega$. *Furthermore, if* f *is integrable, then* $f(x) < \infty$ *for* μ*-almost all* $x \in \Omega$.

Proof. Suppose $\int_\Omega f\ d\mu = 0$. Consider the measurable set $E_n = f^{-1}((\frac{1}{n}, \infty])$, and observe that by definition $\frac{1}{n}\chi_{E_n} \leq f$. Hence $\frac{1}{n}\mu(E_n) \leq \int_\Omega f\ d\mu = 0$, so $\mu(E_n) = 0$ for all n. The union of E_n is equal to $E = f^{-1}((0, \infty])$. By continuity of measures, we have $\mu(E) = \sup_{n \in \mathbb{N}} \mu(E_n) = 0$, so indeed $f(x) = 0$ for μ-almost all $x \in \Omega$.

For the converse, assume that $f(x) = 0$ for μ-almost all $x \in \Omega$, which means that the set E above is a null set. If $s : \Omega \to [0, \infty)$ is a simple measurable function with $s \leq f$, then it means in particular that $s^{-1}(\lambda) \subseteq E$ for all $\lambda \neq 0$. This immediately implies $\int_\Omega s\ d\mu = 0$, but hence also $\int_\Omega f\ d\mu = 0$ by definition.

Now assume that f is integrable. Consider $E = f^{-1}(\infty) \in \mathfrak{M}$ and $s_n = n\chi_E$. Then $s_n \leq f$ for all $n \in \mathbb{N}$, and hence $n\mu(E) = \int_\Omega s_n\ d\mu \leq \int_\Omega f\ d\mu$. Since this holds for all $n \geq 1$ and we assume the right side to be finite, this is only possible for $\mu(E) = 0$. But this is what it means that $f(x) < \infty$ for μ-almost all $x \in \Omega$. This finishes the proof. □

Remark 1.39. Let Ω be a set and $f : \Omega \to \mathbb{C}$ a function. Then it is always possible to decompose f into $f = u + i \cdot v$ for real-valued functions $u, v : \Omega \to \mathbb{R}$, the real and imaginary parts of f. Furthermore we can always uniquely write

$$u = u^+ - u^-, \; v = v^+ - v^-, \; u^+ u^- = 0, \; v^+ v^- = 0,$$

where u^+, u^-, v^+, v^- take values in $[0, \infty)$.[4] We then have $u^\pm, v^\pm \leq |f|$, but also $|f| \leq u^+ + u^- + v^+ + v^-$ by triangle inequality.

Furthermore, if Ω carries a σ-algebra for which f becomes measurable, then the functions u^+, u^-, v^+, v^- are also measurable. In what follows we wish to exploit such decompositions to linearly extend the integral construction.

Definition 1.40. Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. We say that a measurable function $f : \Omega \to \mathbb{C}$ is integrable, if $|f|$ is integrable in the sense of Definition 1.32. In this case we define the integral as

$$\int_\Omega f \; d\mu = \int_\Omega u^+ \; d\mu - \int_\Omega u^- \; d\mu + i\Big(\int_\Omega v^+ \; d\mu - \int_\Omega v^- \; d\mu \Big),$$

where we use the decomposition from Remark 1.39. The set of all such functions is denoted $\mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ or $\mathcal{L}^1(\mu)$.

Theorem 1.41. $\mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ *is a complex vector space with the usual operations. Moreover the integral*

$$\mathcal{L}^1(\Omega, \mathfrak{M}, \mu) \to \mathbb{C}, \quad f \mapsto \int_\Omega f \; d\mu$$

is a linear map.

Proof. The fact that $\mathcal{L}^1$ is a vector space is left as an exercise. Let us proceed to prove linearity of the integral.

Let us first assume that $f, g \in \mathcal{L}^1$ are real-valued. Set $h = f + g$ and observe that hence $h^+ - h^- = f^+ - f^- + g^+ - g^-$, or equivalently $h^+ + f^- + g^- = h^- + f^+ + g^+$. All of these summands are positive integrable functions, so it follows by Proposition 1.37 that

$$\int_\Omega h^+ \; d\mu + \int_\Omega f^- \; d\mu + \int_\Omega g^- \; d\mu = \int_\Omega h^- \; d\mu + \int_\Omega f^+ \; d\mu + \int_\Omega g^+ \; d\mu.$$

All of these summands are finite, and hence we can rearrange this equation to

$$\int_\Omega h^+ \; d\mu - \int_\Omega h^- \; d\mu = \int_\Omega f^+ \; d\mu - \int_\Omega f^- \; d\mu + \int_\Omega g^+ \; d\mu - \int_\Omega g^- \; d\mu,$$

[4] For example, u^+ is the composition of u with the function $\mathbb{R} \to [0, \infty)$ that is the identity on $[0, \infty)$ and sends every negative number to zero.

and hence $\int_\Omega h\ d\mu = \int_\Omega f\ d\mu + \int_\Omega g\ d\mu$. It is also clear from the definition that $\int_\Omega f + ig\ d\mu = \int_\Omega f\ d\mu + i\int_\Omega g\ d\mu$. These two equations imply together that the integral is indeed additive.

Consider a scalar $\alpha > 0$. Then $(\alpha f)^\pm = \alpha \cdot f^\pm$, so from Proposition 1.37 it follows that

$$\int_\Omega \alpha f\ d\mu = \int_\Omega \alpha f^+\ d\mu - \int_\Omega \alpha f^-\ d\mu = \alpha \int_\Omega f\ d\mu.$$

If $\alpha < 0$, then $(\alpha f)^\pm = (-\alpha) f^\mp$, so the similar calculation shows

$$\int_\Omega \alpha f\ d\mu = \int_\Omega (-\alpha f^-)\ d\mu - \int_\Omega (-\alpha f^+)\ d\mu = \alpha \int_\Omega f\ d\mu.$$

If more generally $\alpha = \alpha_1 + i\alpha_2 \in \mathbb{C}$ for $\alpha_1, \alpha_2 \in \mathbb{R}$, then we use the above to calculate

$$\begin{aligned}
\int_\Omega \alpha(f+ig)\ d\mu &= \int_\Omega \alpha_1 f - \alpha_2 g + i(\alpha_1 g + \alpha_2 f)\ d\mu \\
&= \int_\Omega \alpha_1 f - \alpha_2 g\ d\mu + i\int_\Omega \alpha_1 g + \alpha_2 f\ d\mu \\
&= \alpha_1 \int_\Omega f\ d\mu - \alpha_2 \int_\Omega g\ d\mu + i\Big(\alpha_1 \int_\Omega g\ d\mu + \alpha_2 \int_\Omega f\ d\mu\Big) \\
&= \alpha\Big(\int_\Omega f\ d\mu + i\int_\Omega g\ d\mu\Big) \\
&= \alpha \int_\Omega f + ig\ d\mu.
\end{aligned}$$

This finishes the proof. □

Proposition 1.42. *For all $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$, one has*

$$\Big|\int_\Omega f\ d\mu\Big| \leq \int_\Omega |f|\ d\mu.$$

Proof. By considering the polar decomposition of $\int_\Omega f\ d\mu$ and multiplying f with a scalar of modulus one, we may assume without loss of generality that $\int_\Omega f\ d\mu \geq 0$. Write $f = u + iv$ for real-valued functions u, v. Then

$$0 \leq \int_\Omega f\ d\mu = \int_\Omega u\ d\mu \leq \int_\Omega u^+\ d\mu \overset{u^+ \leq |f|}{\leq} \int_\Omega |f|\ d\mu.$$

□

Proposition 1.43. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space and $f : \Omega \to \mathbb{C}$ a measurable function.*

(i) *If $f = 0$ a.e., then f is integrable and $\int_\Omega f\ d\mu = 0$.*

(ii) *If f is integrable and $g : \Omega \to \mathbb{C}$ is another measurable function such that $f = g$ a.e., then g is integrable and $\int_\Omega f \; d\mu = \int_\Omega g \; d\mu$.*

(iii) *If f is integrable and $\int_\Omega f \cdot \chi_E \; d\mu = 0$ for all $E \in \mathfrak{M}$, then $f = 0$ a.e.*

Proof. (i): By the previous proposition, it suffices to show $\int_\Omega |f| \; d\mu = 0$, but this is a consequence of Proposition 1.38.

(ii): The assumption means that $g - f = 0$ a.e., so $g - f$ is integrable and has vanishing integral. Hence $g = f + (g - f)$ is integrable, and according to additivity of the integral it has the same integral as f.

(iii): Consider $E = \{x \in \Omega \mid \mathrm{Re} f(x) \geq 0\}$. Decompose $f = u^+ - u^- + i(v^+ - v^-)$ as before, and observe

$$0 = \mathrm{Re}\Big(\int_\Omega f\chi_E \; d\mu \Big) = \int_\Omega \mathrm{Re}(f\chi_E) \; d\mu = \int_\Omega u^+ \; d\mu.$$

By Proposition 1.38 it follows that $u^+ = 0$ a.e.. Repeating this argument with different choices for E, one can also see that u^-, v^+, v^- vanish almost everywhere. But this clearly implies that $f = 0$ a.e.. □

Proposition 1.44. *Let $(\Omega, \mathfrak{M}, \mu)$ be a σ-finite measure space and $f : \Omega \to \mathbb{C}$ an integrable function. Let $F \subseteq \mathbb{C}$ be a closed set with the property that for all $E \in \mathfrak{M}$ with $0 < \mu(E) < \infty$, one has*

$$\frac{1}{\mu(E)} \int_\Omega f\chi_E \; d\mu \;\in\; F.$$

Then $f(x) \in F$ for μ-almost all $x \in \Omega$.

Proof. As μ is σ-finite, we can write $\Omega = \bigcup E_n$ as an increasing union of sets $E_n \in \mathfrak{M}$ with $\mu(E_n) < \infty$. If the set of all $x \in E_n$ with $f(x) \notin F$ is a null set for all $n \geq 1$, then by σ-subadditivity the same is true for the set of all $x \in \Omega$ with $f(x) \notin F$. So by restricting the problem to each E_n in place of Ω, we may assume without loss of generality that μ is a finite measure.

If $F = \mathbb{C}$, there is nothing to prove, so let us assume that F has non-empty complement. Let $B \subset \mathbb{C} \setminus F$ be a closed ball around some point z with radius $r > 0$. Set $g = f\chi_{f^{-1}(B)}$ and write $g = z\chi_{f^{-1}(B)} + g_0$, where $|g_0| \leq r\chi_{f^{-1}(B)}$. Then either $\mu(f^{-1}(B)) = 0$, or

$$\Big| z - \frac{1}{\mu(f^{-1}(B))} \int_\Omega g \; d\mu \Big| = \Big| \frac{1}{\mu(f^{-1}(B))} \int_\Omega g_0 \; d\mu \Big| \leq r.$$

The latter would imply that the number $\frac{1}{\mu(f^{-1}(B))} \int_\Omega g \; d\mu$ is in B, which is disjoint from F, a contradiction. Hence we conclude $\mu(f^{-1}(B)) = 0$. Since we can write $\mathbb{C} \setminus F$ as a countable union of closed balls, this implies $\mu(f^{-1}(\mathbb{C} \setminus F)) = 0$, which confirms the claim. □

Proposition 1.45. *Let* $(\Omega, \mathfrak{M}, \mu)$ *be a measure space. Consider*

$$N = \{f : \Omega \to \mathbb{C} \mid f \text{ is measurable and } f(x) = 0 \text{ for } \mu\text{-almost all } x\}.$$

Then N *is a linear subspace of* $\mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$*, and* $f \in N$ *holds if and only if* f *is measurable and* $|f| \in N$*.*

Proof. From the previous proposition we get that indeed $N \subseteq \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$, and the second equivalence of the statement is clear.

Let $f, g \in N$ be given. First of all, it is clear that $\alpha \cdot f \in N$ for all $\alpha \in \mathbb{C}$. So we shall show $f + g \in N$. We do have that this function is measurable, so it remains to show that it vanishes almost everywhere. Indeed, we have $(f+g)^{-1}(\mathbb{C} \setminus \{0\}) \subseteq f^{-1}(\mathbb{C} \setminus \{0\}) \cup g^{-1}(\mathbb{C} \setminus \{0\})$, which by assumption implies $\mu((f+g)^{-1}(\mathbb{C} \setminus \{0\})) = 0$, or that indeed $f(x) + g(x) = 0$ holds for μ-almost all x. □

Definition 1.46. Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. Then in light of the above, we define the quotient vector space

$$L^1(\mu) = L^1(\Omega, \mathfrak{M}, \mu) = \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)/N.$$

For a function $f \in \mathcal{L}^1$, we may for the moment denote its coset by $[f] = f + N$, but in time we will transition to denote it simply by f, even though this is some abuse of notation.[5] From the previous results above it follows that the assignment

$$L^1(\Omega, \mathfrak{M}, \mu) \to \mathbb{C}, \quad [f] \mapsto \int_\Omega f \, d\mu$$

is a well-defined linear map, which we will call the integral. Since $f \in N$ happens precisely when $|f| \in N$, it makes sense to define $|[f]| = [|f|]$. Consequently, the following semi-norm $\|\cdot\|_1$ on $\mathcal{L}^1$ becomes a norm (cf. Proposition 1.38) on the L^1-space via

$$\|f\|_1 = \int_\Omega |f| \, d\mu, \quad \|[f]\|_1 := \|f\|_1.$$

1.5 Convergence Theorems

Theorem 1.47 (Monotone Convergence Theorem)**.** *Let* $(\Omega, \mathfrak{M}, \mu)$ *be a measure space and* $f_n : \Omega \to [0, \infty]$ *a (pointwise) increasing sequence of measurable functions. Then* $f = \sup_{n \geq 1} f_n$ *is a also measurable, and we have*

$$\int_\Omega f \, d\mu = \sup_{n \geq 1} \int_\Omega f_n \, d\mu.$$

[5] But this is justified by the fact that we will mostly form integrals, which do not depend on the representative for such a coset.

Proof. By Proposition 1.35, for every $n \geq 1$, we find an increasing sequence of measurable step functions $s_{n,k} : \Omega \to [0,\infty]$ such that $f_n = \sup_{k\geq 1} s_{n,k}$. Not only that, but the construction of these functions in the proof allows us to assume that $|f_n(x) - s_{n,k}(x)| \leq 2^{-k}$ whenever $f_n(x) \leq k$, and $s_{n,k} = k$ whenever $f_n(x) \geq k$.

Since f_n is increasing, we may consider the positive measurable step functions $t_n = \max_{j\leq n} s_{j,n}$, and ensure that t_n is increasing and $t_n \leq f_n$ for all n. We claim $f = \sup_{n\geq 1} t_n$. Indeed, let $x \in \Omega$ be given. If $f(x) = \infty$, then $f_n(x) \to \infty$, so $t_n(x) \geq s_{n,n}(x) \geq \min(n, f_n(x) - 2^{-n}) \to \infty$. If $f(x) < \infty$, then in particular $f(x) \leq n$ for all large enough n, and hence

$$\begin{aligned} |f(x) - t_n(x)| &\leq |f(x) - s_{n,n}(x)| \\ &\leq |f_n(x) - f(x)| + |f_n(x) - s_{n,n}(x)| \\ &\leq |f_n(x) - f(x)| + 2^{-n} \ \to \ 0. \end{aligned}$$

We appeal to Proposition 1.36 and see that

$$\int_\Omega f \ d\mu = \sup_{n\geq 1} \int_\Omega t_n \ d\mu \overset{t_n\leq f_n}{\leq} \sup_{n\geq 1} \int_\Omega f_n \ d\mu.$$

The "$\geq$" relation is on the other hand clear from the fact that the integral is monotone. This finishes the proof. $\square$

Lemma 1.48 (Fatou)**.** *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space and $f_n : \Omega \to [0,\infty]$ be a sequence of measurable functions. Then $\liminf_{n\to\infty} f_n$ is also measurable, and we have*

$$\int_\Omega (\liminf_{n\to\infty} f_n) \ d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \ d\mu.$$

Proof. Denote $g_k = \inf_{n\geq k} f_n$, and recall that $\liminf_{n\to\infty} f_n = \sup_{k\geq 1} g_k$. We thus see that $\liminf_{n\to\infty} f_n$ is a measurable function. If $n \geq k$, then evidently $g_k \leq f_n$, so $\int_\Omega g_k \ d\mu \leq \int_\Omega f_n \ d\mu$. In particular, this is true for arbitrarily large n, so $\int_\Omega g_k \ d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \ d\mu$. Using the Monotone Convergence Theorem, we thus see

$$\int_\Omega \liminf_{n\to\infty} f_n \ d\mu = \sup_{k\geq 1} \int_\Omega g_k \ d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \ d\mu.$$

$\square$

Theorem 1.49 (Dominated Convergence Theorem)**.** *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space and $f_n \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ a sequence converging to a function f pointwise. Suppose that there exists a positive integrable function $g : \Omega \to [0,\infty)$ such that $|f_n| \leq g$ for all n. Then $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$, and*

$$\lim_{n\to\infty} \int_\Omega |f - f_n| \ d\mu = 0, \quad \int_\Omega f \ d\mu = \lim_{n\to\infty} \int_\Omega f_n \ d\mu.$$

Proof. We have that f is measurable (Theorem 1.13) and $|f| \leq g$, hence $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$. Furthermore, we have $|f - f_n| \leq 2g$, and so we may apply Fatou's Lemma to the sequence $2g - |f - f_n|$ to get

$$\int_\Omega 2g \, d\mu \leq \liminf_{n\to\infty} \int_\Omega 2g - |f - f_n| \, d\mu = \int_\Omega 2g \, d\mu - \limsup_{n\to\infty} \int_\Omega |f - f_n| \, d\mu.$$

Since g was assumed to be integrable, this is equivalent to

$$0 = \limsup_{n\to\infty} \int_\Omega |f - f_n| \, d\mu = \lim_{n\to\infty} \int_\Omega |f - f_n| \, d\mu.$$

As a consequence we obtain

$$\Big| \int_\Omega f \, d\mu - \int_\Omega f_n \, d\mu \Big| = \Big| \int_\Omega f - f_n \, d\mu \Big| \leq \int_\Omega |f - f_n| \, d\mu \stackrel{n\to\infty}{\longrightarrow} 0.$$

This finishes the proof. □

Remark 1.50. For practical applications of the Dominated Convergence Theorem, it is useful to observe that it is only necessary to assume that the pointwise convergence $f_n(x) \to f(x)$ and the inequality $|f_n(x)| \leq g(x)$ holds for μ-almost all $x \in \Omega$. This is a consequence of Proposition 1.43, as we may just re-define all functions f_n on a common measurable null set to have value zero and thus enforce these statements to hold on all points, yet the integrals in the conclusion of the statement remain the same.

Remark 1.51. The Dominated Convergence Theorem really only holds for sequences of functions, and its analogous generalizations for more general families of functions (such as nets) are false. A counterexample is discussed in the exercise sessions.

Theorem 1.52. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. Suppose that a sequence $f_n \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ satisfies the Cauchy criterion in the semi-norm $\|\cdot\|_1$. Then there exists a subsequence $(f_{n_k})_k$ and a function $f \in \mathcal{L}^1(\Omega, \mathfrak{M}, \mu)$ such that*

$$f_{n_k}(x) \stackrel{k\to\infty}{\longrightarrow} f(x) \quad \text{for } \mu\text{-almost all } x \in \Omega,$$

and moreover $\|f - f_n\|_1 \stackrel{n\to\infty}{\longrightarrow} 0$. In particular, it follows that $L^1(\Omega, \mathfrak{M}, \mu)$ is a Banach space with respect to the norm $\|\cdot\|_1$.

Proof. By applying the Cauchy criterion inductively, we can find a subsequence $(f_{n_k})_k$ such that $\|f_{n_k} - f_{n_{k+1}}\| \leq 2^{-k}$. We define the measurable function $g : \Omega \to [0, \infty]$ as $g(x) = \sum_{k=1}^\infty |f_{n_k}(x) - f_{n_{k+1}}(x)|$. Then it follows from the Monotone Convergence Theorem that

$$\int_\Omega g \, d\mu = \sum_{k=1}^\infty \int_\Omega |f_{n_k} - f_{n_{k+1}}| \, d\mu = \sum_{k=1}^\infty \|f_{n_k} - f_{n_{k+1}}\|_1 \leq \sum_{k=1}^\infty 2^{-k} = 1.$$

In particular g is integrable, and by Proposition 1.38, the set $E = g^{-1}([0,\infty))$ has a complement of zero measure. For all $x \in E$, we have by definition that the series $\sum_{k=1}^{\infty}[f_{n_{k+1}}(x) - f_{n_k}(x)]$ converges absolutely, and hence the function

$$f(x) := f_{n_1}(x) + \sum_{k=1}^{\infty}[f_{n_{k+1}}(x) - f_{n_k}(x)] = \lim_{k\to\infty} f_{n_k}(x), \quad x \in E$$

is well-defined and measurable on E. We extend f to a measurable function on Ω by defining it to be zero on the complement of E. We get by the triangle inequality that for all $k \geq 1$, the function f_{n_k} is dominated (on E) by the integrable function $|f_{n_1}| + g$. Therefore it follows from the Dominated Convergence Theorem that $\|f - f_{n_k}\|_1 \stackrel{k\to\infty}{\longrightarrow} 0$. Since the sequence $(f_n)_n$ was assumed to satisfy the Cauchy criterion in $\|\cdot\|_1$ and we just showed that a subsequence converges to f in this norm, it follows that also $\|f - f_n\| \to 0$.[6] This finishes the proof. □

Proposition 1.53. *Let $(\Omega_1, \mathfrak{M}_1, \mu_1)$ be a measure space, Ω_2 a non-empty set, and $\varphi : \Omega_1 \to \Omega_2$ a map. Then*

$$\varphi_*\mathfrak{M}_1 := \left\{E \subseteq \Omega_2 \mid \varphi^{-1}(E) \in \mathfrak{M}_1\right\}$$

is a σ-algebra and

$$\varphi_*\mu_1 : \varphi_*\mathfrak{M}_1 \to [0,\infty], \quad \varphi_*\mu_1(E) = \mu_1(\varphi^{-1}(E))$$

is a measure. These are called the push-forward σ-algebra and the push-forward measure with respect to φ. Then with respect to this measure space structure on Ω_2, φ becomes a measurable map.

Proof. We have already seen in the exercise sessions that $\varphi_*\mathfrak{M}_1$ is a σ-algebra. For the measure condition, we observe that $\varphi_*\mu_1(\emptyset) = 0$ is trivial. Moreover if $E_n \in f_*\mathfrak{M}_1$ is a sequence of disjoint sets, then $\varphi^{-1}(E_n) \in \mathfrak{M}_1$ is a sequence of disjoint sets, and hence

$$\varphi_*\mu_1(\bigcup E_n) = \mu_1(\bigcup \varphi^{-1}(E_n)) = \sum \mu_1(\varphi^{-1}(E_n)) = \sum \varphi_*\mu_1(E_n).$$

The last part of the statement is trivial. □

Theorem 1.54. *Let $(\Omega_1, \mathfrak{M}_1, \mu_1)$ be a measure space, Ω_2 a non-empty set, and $\varphi : \Omega_1 \to \Omega_2$ a map. We define $\mathfrak{M}_2 = \varphi_*\mathfrak{M}_1$ and $\mu_2 = \varphi_*\mu_1$ in the above sense.*

[6] As an exercise, fill in this detail yourself! This is a standard $\varepsilon/2$-argument, and it is exactly how one proves the completeness of $\mathbb{R}$ out of the axiom that bounded sets have suprema.

(i) *If* $f : \Omega_2 \to [0, \infty]$ *is a measurable function, then*

$$\int_{\Omega_2} f \, d\mu_2 = \int_{\Omega_1} f \circ \varphi \, d\mu_1.$$

(ii) *If* $f : \Omega_2 \to \mathbb{C}$ *is a* μ_2*-integrable function, then* $f \circ \varphi$ *is* μ_1*-integrable, and we have*

$$\int_{\Omega_2} f \, d\mu_2 = \int_{\Omega_1} f \circ \varphi \, d\mu_1.$$

Proof. (i): By definition, a set $E \subseteq \Omega_2$ belongs to $\mathfrak{M}_2$ precisely when $\varphi^{-1}(E) \in \mathfrak{M}_1$. Since $\chi_E \circ \varphi = \chi_{\varphi^{-1}(E)}$, we have

$$\int_{\Omega_2} \chi_E \, d\mu_2 = \mu_2(E) = \mu_1(\varphi^{-1}(E)) = \int_{\Omega_1} \chi_E \circ \varphi \, d\mu_1.$$

In other words, the claim holds for functions of the form $f = \chi_E$. By linearity of the integral, the desired equation holds for all positive measurable step functions in place of f. Now let f be as general as in the statement, and write $f = \sup_{n \in \mathbb{N}} s_n$ for an increasing sequence of positive measurable step functions $s_n : \Omega_2 \to [0, \infty)$, using Proposition 1.35. Clearly we also have $f \circ \varphi = \sup_{n \in \mathbb{N}} s_n \circ \varphi$. Then by Proposition 1.36, we see that

$$\int_{\Omega_2} f \, d\mu_2 = \sup_{n \in \mathbb{N}} \int_{\Omega_2} s_n \, d\mu_2 = \sup_{n \in \mathbb{N}} \int_{\Omega_1} s_n \circ \varphi \, d\mu_1 = \int_{\Omega_1} f \circ \varphi \, d\mu_1.$$

(ii): By definition, f being μ_2-integrable means that $\int_{\Omega_2} |f| \, d\mu_2 < \infty$, which by part (i) implies $\int_{\Omega_1} |f \circ \varphi| \, d\mu_1 < \infty$. In other words, $f \circ \varphi$ is μ_1-integrable. The desired equality thus follows directly from part (i), the linearity of the integral, and the fact that we may write f as a linear combination of integrable positive functions. □

Remark 1.55. In the above theorem, we pushed forward the measure space structure from Ω_1 to get one on Ω_2 which makes the statement of the theorem true. It may of course happen that we have an a priori given measure space $(\Omega_2, \mathfrak{M}_2, \mu_2)$ and a measurable map $\varphi : \Omega_1 \to \Omega_2$ with the property that $\mu_1(\varphi^{-1}(E)) = \mu_2(E)$ for all $E \in \mathfrak{M}_2$. Convince yourself that the statement of the theorem will still be true!

2 Carathéodory's Construction of Measures

2.1 Measures on Semi-Rings and Rings

Definition 2.1. Let Ω be a non-empty set. A semi-ring on Ω is a set of subsets $\mathfrak{A} \subseteq 2^\Omega$ such that

(1) $\emptyset \in \mathfrak{A}$.

(2) If $E, F \in \mathfrak{A}$, then $E \cap F \in \mathfrak{A}$.

(3) If $E, F \in \mathfrak{A}$, then $E \setminus F$ is a finite disjoint union of sets in $\mathfrak{A}$.

Example 2.2. The set $\mathfrak{A}$ of all subsets of Ω with at most one element form a semi-ring. More interestingly, if $n \geq 2$, then the set of half-open cubes

$$\mathfrak{A} = \{(a_1, b_1] \times \cdots \times (a_n, b_n] \mid a_j, b_j \in \mathbb{R},\ a_j \leq b_j\}$$

is also a semi-ring on $\mathbb{R}^n$. (This will be justified later.)

Definition 2.3. Let Ω be a non-empty set. A ring over Ω is a set of subsets $\mathfrak{R} \subseteq 2^\Omega$ such that

(1) $\emptyset \in \mathfrak{R}$.

(2) If $E, F \in \mathfrak{R}$, then $E \cup F \in \mathfrak{R}$.

(3) If $E, F \in \mathfrak{R}$, then $E \setminus F \in \mathfrak{R}$.

Remark 2.4. One always has $E \cap F = E \setminus (E \setminus F)$, so every ring is closed under forming intersections. In particular, it follows that every ring is also a semi-ring. The converse is of course not true, e.g., the set of (left) half open intervals in $\mathbb{R}$ is a semi-ring, but not a ring.

Lemma 2.5. *Let $\mathfrak{A}$ be a semi-ring over Ω. Then the smallest ring $\mathfrak{R}$ over Ω containing $\mathfrak{A}$ is given as*

$$\mathfrak{R} = \{E_1 \cup \cdots \cup E_n \mid E_j \in \mathfrak{A} \text{ are pairwise disjoint}\}.$$

We will refer to it as the ring generated by $\mathfrak{A}$.

Proof. If $\mathfrak{R}$ is given as above, then evidently every ring containing $\mathfrak{A}$ will also contain $\mathfrak{R}$. Hence the claim will follow once we show that $\mathfrak{R}$ is indeed a ring. It is first of all clear that $\emptyset \in \mathfrak{R}$. Let $E, F \in \mathfrak{R}$. We shall first show $E \setminus F \in \mathfrak{R}$.

By definition, we find pairwise disjoints sets $E_1, \dots, E_n \in \mathfrak{A}$ and pairwise disjoint sets $F_1, \dots, F_m \in \mathfrak{A}$ such that $E = \bigcup_{j=1}^n E_j$ and $F = \bigcup_{\ell=1}^m F_\ell$. Then

$$E \setminus F = \bigcup_{j=1}^n E_j \setminus \Big(\bigcup_{\ell=1}^m F_\ell \Big) = \bigcup_{j=1}^n \Big[\Big(\Big((E_j \setminus F_1) \setminus F_2 \Big) \setminus \dots \Big) \setminus F_m \Big]$$

Since $\mathfrak{A}$ was assumed to be a semi-ring, the set $E_j \setminus F_1$ is a finite union of pairwise disjoint sets in $\mathfrak{A}$. Once we know that it follows again by the semi-ring property that $(E_j \setminus F_1) \setminus F_2$ is a finite union of pairwise disjoint sets in

$\mathfrak{A}$. By induction, it follows that each expression of the form $\Big(\big((E_j \setminus F_1) \setminus F_2\big) \setminus \dots\Big) \setminus F_m$ is a finite union of disjoint sets in $\mathfrak{A}$. These unions are in turn pairwise disjoint in j, and hence it follows that $E \setminus F \in \mathfrak{R}$.

Now from this we also get $E \cup F \in \mathfrak{R}$ because one can write it as a disjoint union $E \cup F = E \cup (F \setminus E)$, and it is clear that $\mathfrak{R}$ is closed under disjoint unions of its elements. □

Definition 2.6. Let $\mathfrak{A}$ be a semi-ring on Ω. A measure on $\mathfrak{A}$ is a map $\mu : \mathfrak{A} \to [0, \infty]$ such that

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

(ii) If $E_n \in \mathfrak{A}$ is a sequence of pairwise disjoint sets with $\bigcup_{n \geq 1} E_n \in \mathfrak{A}$, then $\mu(\bigcup_{n \geq 1} E_n) = \sum_{n \geq 1} \mu(E_n)$. ($\sigma$-additivity)

In particular, if $\mathfrak{R}$ is a ring on Ω, then a measure on the ring $\mathfrak{R}$ is defined as a measure on $\mathfrak{R}$ viewed as a semi-ring.

Proposition 2.7. *Let $\mathfrak{A}$ be a semi-ring over Ω, and let $\mathfrak{R}$ be the ring generated by $\mathfrak{A}$. Then every measure μ_0 on $\mathfrak{A}$ extends uniquely to a measure μ on $\mathfrak{R}$.*

Proof. By the definition of $\mathfrak{R}$, it is clear that if μ exists, then the σ-additivity of μ implies that μ is uniquely determined by μ_0. So let us argue why μ exists in the first place. Let $E \in \mathfrak{R}$, and write it as $E = E_1 \cup \dots \cup E_n$ with pairwise disjoint sets $E_j \in \mathfrak{A}$. Define

$$\mu(E) := \mu_0(E_1) + \dots + \mu_0(E_n).$$

We have to show that μ is well-defined, and that it is σ-additive. Suppose that $F_1, \dots, F_m \in \mathfrak{A}$ is another collection of pairwise disjoint sets with $E = F_1 \cup \dots \cup F_m$. Then one has for all $j \in \{1, \dots, n\}$ that

$$E_j = E_j \cap E = \bigcup_{\ell=1}^{m} (E_j \cap F_\ell),$$

and this union is a disjoint union. Analogously we have $F_\ell = \bigcup_{j=1}^{n} (E_j \cap F_\ell)$ for all $\ell \in \{1, \dots, m\}$. By using the additivity of μ_0, it follows that

$$\sum_{j=1}^{n} \mu_0(E_j) = \sum_{j=1}^{n} \sum_{\ell=1}^{m} \mu_0(E_j \cap F_\ell) = \sum_{\ell=1}^{m} \sum_{j=1}^{n} \mu_0(E_j \cap F_\ell) = \sum_{\ell=1}^{m} \mu_0(F_\ell).$$

So we see that μ is a well-defined map.

Now let us see why μ is σ-additive. Suppose that $E_n \in \mathfrak{R}$ is a sequence of pairwise disjoint sets with $E = \bigcup_{n\geq 1} E_n \in \mathfrak{R}$. Write $E = A_1 \cup \cdots \cup A_m$ for pairwise disjoint sets $A_1, \dots, A_m \in \mathfrak{A}$, and moreover write $E_n = A_{n,1} \cup \cdots \cup A_{n,m_n} \in \mathfrak{A}$ for all $n \geq 1$ some $m_n \geq 1$. Then applying σ-additivity of μ_0 several times, we can see

$$
\begin{aligned}
\mu(E) &= \sum_{j=1}^{m} \mu_0(A_j) \\
&= \sum_{j=1}^{m} \mu_0\Big(\bigcup_{n\geq 1} (A_j \cap E_n) \Big) \\
&= \sum_{j=1}^{m} \mu_0\Big(\bigcup_{n\geq 1} \bigcup_{k\leq m_n} (A_j \cap A_{n,k}) \Big) \\
&= \sum_{j=1}^{m} \sum_{n=1}^{\infty} \sum_{k=1}^{m_n} \mu_0(A_j \cap A_{n,k}) \\
&= \sum_{n=1}^{\infty} \sum_{j=1}^{m} \sum_{k=1}^{m_n} \mu_0(A_j \cap A_{n,k}) \; = \; \sum_{n=1}^{\infty} \mu(E_n).
\end{aligned}
$$

□

2.2 Outer measures

Definition 2.8. Let Ω be a non-empty set. An outer measure on Ω is a map $\nu : 2^\Omega \to [0, \infty]$ satisfying:

OM1. $\nu(\emptyset) = 0$.

OM2. If $E \subseteq F \subseteq \Omega$, then $\nu(E) \leq \nu(F)$.

OM3. If $E_n \subseteq \Omega$ is any sequence of pairwise disjoint sets, then $\nu\Big(\bigcup_{n\geq 1} E_n\Big) \leq \sum_{n=1}^{\infty} \nu(E_n)$. ($\sigma$-subadditivity)

Remark 2.9. WARNING! The terminology is a bit confusing. Contrary to the usual rules of the English language, an "outer measure" in the above sense does not describe a measure that has an additional "outerness" property. Instead, it describes a weaker concept than that of a measure.

Proposition 2.10. *Let Ω be a non-empty set, $\mathfrak{R}$ a ring on Ω, and μ a measure on $\mathfrak{R}$. Define for every subset $E \subset \Omega$ the value*

$$
\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} \mu(E_n) \;\middle|\; E_n \in \mathfrak{R} \textit{ is a sequence with } E \subseteq \bigcup_{n\geq 1} E_n \right\}.^{7}
$$

[7] Keep in mind that by convention, $\inf \emptyset := \infty$. For certain sets E there may not exist any sequence E_n with these properties.

Then μ^* defines an outer measure on Ω.

Proof. One gets $\mu^*(\emptyset)$ by choosing $E_n = \emptyset$. If $E \subseteq F \subseteq \Omega$ are two subsets, then evidently there are at least as many ways to cover E by sequences in $\mathfrak{R}$ as for F, which leads directly to $\mu^*(E) \leq \mu^*(F)$.

Now let $A_n \subseteq \Omega$ be a sequence of (pairwise disjoint) sets. We shall show that $\mu^*(\bigcup_{n\geq 1} A_n) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$. If the right side is infinite, there is nothing to show, so let us assume that it is finite. Let $\varepsilon > 0$. For every $n \geq 1$, we may by definition find a sequence $E_{n,k} \in \mathfrak{R}$ with $A_n \subseteq \bigcup_{k\geq 1} E_{n,k}$ and $\sum_{k=1}^{\infty} \mu(E_{n,k}) \leq \mu^*(A_n) + 2^{-n}\varepsilon$. Then the set $\bigcup_{n\geq 1} A_n$ is of course covered by the countably many sets $E_{n,k} \in \mathfrak{R}$ for $n, k \geq 1$, and hence

$$\mu^*(\bigcup_{n\geq 1} A_n) \leq \sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mu(E_{n,k}) \leq \sum_{n=1}^{\infty}(2^{-n}\varepsilon + \mu^*(A_n)) = \varepsilon + \sum_{n=1}^{\infty} \mu^*(A_n).$$

Since $\varepsilon > 0$ was arbitrary, this shows the claim. □

Definition 2.11. Let ν be an outer measure on a set Ω. We say that a set $E \subseteq \Omega$ is ν-measurable, if for every subset $A \subseteq \Omega$, we have $\nu(A) = \nu(A \cap E) + \nu(A \setminus E)$.

Theorem 2.12 (Carathéodory)**.** *Let ν be an outer measure on a set Ω. Then the ν-measurable subsets of Ω form a σ-algebra $\mathfrak{M}$, and the restriction of ν to $\mathfrak{M}$ defines a measure.*

Proof. Evidently $\emptyset \in \mathfrak{M}$ and if $E \in \mathfrak{M}$, then also $E^c \in \mathfrak{M}$. Let us consider finite unions. Let $E, F \in \mathfrak{M}$, and $A \subseteq \Omega$ an arbitrary subset. Then

$$\begin{array}{rcl}
\nu(A) & = & \nu\Big(A \cap (E \cup F) \cup A \setminus (E \cup F)\Big) \\
& \leq & \nu(A \cap (E \cup F)) + \nu(A \setminus (E \cup F)) \\
& = & \nu(A \cap E \cup (A \cap F \setminus E)) + \nu(A \setminus (E \cup F)) \\
& \leq & \nu(A \cap E) + \nu((A \setminus E) \cap F) + \nu((A \setminus E) \setminus F)) \\
& = & \nu(A \cap E) + \nu(A \setminus E) \ = \ \nu(A).
\end{array}$$

But from this we get equality everywhere, which implies $E \cup F \in \mathfrak{M}$ because A was arbitrary. If additionally $E \cap F = \emptyset$, then we have $\nu(A \cap (E \cup F)) = \nu(A \cap (E \cup F) \cap E) + \nu(A \cap (E \cup F) \setminus E) = \nu(A \cap E) + \nu(A \cap F)$. So inserting $A = E \cup F$ yields that ν is finitely additive on $\mathfrak{M}$.

From $E \cup F \in \mathfrak{M}$ it immediately follows that $\mathfrak{M}$ is closed under intersections and differences. Hence $\mathfrak{M}$ becomes a σ-algebra if we can show that for all sequences of pairwise disjoint sets $E_n \in \mathfrak{M}$, we have $E = \bigcup_{n\geq 1} E_n \in \mathfrak{M}$.

For $m \geq 1$ and all sets $A \subseteq \Omega$, we have

$$\begin{aligned} \nu(A) &= \nu\Big(A \cap \Big(\bigcup_{n\leq m} E_n\Big)\Big) + \nu\Big(A \setminus \Big(\bigcup_{n\leq m} E_n\Big)\Big) \\ &= \sum_{n=1}^{m} \nu(A \cap E_n) + \nu\Big(A \setminus \Big(\bigcup_{n\leq m} E_n\Big)\Big) \\ &\geq \sum_{n=1}^{m} \nu(A \cap E_n) + \nu(A \setminus E). \end{aligned}$$

Since m was arbitrary, we get $\nu(A) \geq \sum_{n=1}^{\infty} \nu(A \cap E_n) + \nu(A \setminus E)$. On the other hand, the equality $A = (A \setminus E) \cup (A \cap E) = (A \setminus E) \cup \bigcup_{n\geq 1} A \cap E_n$ together with σ-subadditivity yields

$$\begin{aligned} \nu(A) &\leq \nu(A \cap E) + \nu(A \setminus E) \\ &\leq \sum_{n=1}^{\infty} \nu(A \cap E_n) + \nu(A \setminus E) \ \leq \ \nu(A). \end{aligned}$$

In particular we get the equality $\nu(A) = \nu(A \cap E) + \nu(A \setminus E)$, which yields $E \in \mathfrak{M}$ as A was arbitrary. Furthermore, if we insert $A = E$, then we also have $\nu(E) = \sum_{n=1}^{\infty} \nu(E_n)$, which shows that ν is σ-additive on $\mathfrak{M}$. In particular it is indeed a measure when restricted to $\mathfrak{M}$. $\square$

Definition 2.13. A measure space $(\Omega, \mathfrak{M}, \mu)$ is called complete, if for all sets $E \subseteq F \subseteq \Omega$, one has that $F \in \mathfrak{M}$ and $\mu(F) = 0$ implies $E \in \mathfrak{M}$.

Proposition 2.14. *Let Ω be a non-empty set and ν an outer measure on Ω. Let $\mathfrak{M}$ be the σ-algebra of ν-measurable sets, and define the measure $\mu = \nu|_{\mathfrak{M}}$. Then $(\Omega, \mathfrak{M}, \mu)$ is a complete measure space.*

Proof. Suppose $E \subseteq F \in \mathfrak{M}$ are given with $\mu(F) = 0$. Then we observe for all $A \subseteq \Omega$ that

$$\begin{aligned} \nu(A) &\leq \nu(A \cap E) + \nu(A \setminus E) \\ &\leq \nu(A \cap F) + \nu(A \setminus E) \\ &\leq 0 + \nu(A). \end{aligned}$$

So we see that these are all equalities. Since A was arbitrary, this implies that $E \in \mathfrak{M}$. $\square$

Theorem 2.15. *Let $\mathfrak{R}$ be a ring on a set Ω, and μ a measure on $\mathfrak{R}$. Then the σ-algebra $\mathfrak{M}$ of μ^*-measurable sets contains $\mathfrak{R}$, and we have $\mu^*|_{\mathfrak{R}} = \mu$.*

Proof. We see right away for all $E \in \mathfrak{R}$ that $\mu^*(E) \leq \mu(E)$ since we can write $E = E \cup \emptyset \cup \emptyset \cup \dots$. On the other hand, if $E_n \in \mathfrak{R}$ is any sequence of sets with $E \subseteq \bigcup_{n\geq 1} E_n$, then it follows from σ-subadditivity and monotonicity of μ that

$$\mu(E) = \mu\Big(\bigcup_{n\geq 1}(E \cap E_n)\Big) \leq \sum_{n=1}^{\infty} \mu(E \cap E_n) \leq \sum_{n=1}^{\infty} \mu(E_n).$$

By taking the infimum over all possible choices of such sequences, we arrive at $\mu(E) = \mu^*(E)$. Since $E \in \mathfrak{R}$ was arbitrary, we have just shown $\mu^*|_{\mathfrak{R}} = \mu$.

Now we need to show that every set $E \in \mathfrak{R}$ is μ^*-measurable. Let $A \subseteq \Omega$ be any set. Since we always have $\mu^*(A) \leq \mu^*(A \cap E) + \mu^*(A \setminus E)$, we may assume without loss of generality that $\mu^*(A) < \infty$. Let $A_n \in \mathcal{R}$ be a sequence of sets with $A \subseteq \bigcup_{n\geq 1} A_n$. Then

$$\begin{array}{rcl} \mu^*(A) & \leq & \mu^*(A \cap E) + \mu^*(A \setminus E) \\ & \leq & \sum_{n=1}^{\infty} \mu(A_n \cap E) + \mu(A_n \setminus E) \\ & = & \sum_{n=1}^{\infty} \mu(A_n). \end{array}$$

If we take the infimum over all possible such sequences A_n, then the right side approaches the value $\mu^*(A)$, and hence we get the equality $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \setminus E)$. Since $A \subseteq \Omega$ and $E \in \mathfrak{R}$ were arbitrary, this finishes the proof. $\square$

Definition 2.16. Let $\mathfrak{A}$ be a semi-ring over Ω, and μ a measure on $\mathfrak{A}$. We say that μ is σ-finite, if there is a sequence $E_n \in \mathfrak{A}$ with $\mu(E_n) < \infty$ and $\Omega = \bigcup_{n\geq 1} E_n$.

Theorem 2.17. *Let $\mathfrak{R}$ be a ring on a set Ω, and μ a σ-finite measure on $\mathfrak{R}$. Let $\mathfrak{M}_0$ be the σ-algebra generated by $\mathfrak{R}$. Then the measure extension $\mu^*|_{\mathfrak{M}_0}$ from $\mathfrak{R}$ to $\mathfrak{M}_0$ is the unique measure on $\mathfrak{M}_0$ extending μ on $\mathfrak{R}$.*

Proof. Suppose that μ_1 is any measure on $\mathfrak{M}_0$ with $\mu_1|_{\mathfrak{R}} = \mu$. We claim $\mu_1 \leq \mu^*|_{\mathfrak{M}_0}$. Let $E \in \mathfrak{M}_0$, and assume without loss of genereality $\mu^*(E) < \infty$. So in particular there exist sequences $E_n \in \mathfrak{R}$ with $E \subseteq \bigcup_{n\geq 1} E_n$. By σ-subadditivity of μ_1, it follows that $\mu_1(E) \leq \sum_{n=1}^{\infty} \mu(E_n)$. But since this holds for any choice of $(E_n)_n$, we obtain $\mu_1(E) \leq \mu^*(E)$.

Since we assumed μ to be σ-finite, we can choose some sequence $F_n \in \mathfrak{R}$ with $\mu(F_n) < \infty$ and $\Omega = \bigcup_{n\geq 1} F_n$. Without loss of generality we may assume that $F_n \subseteq F_{n+1}$ for all $n \geq 1$. For every $A \in \mathfrak{M}_0$ we have

$$\mu_1(F_n \cap A) + \mu_1(F_n \setminus A) = \mu_1(F_n) = \mu^*(F_n) = \mu^*(F_n \cap A) + \mu^*(F_n \setminus A).$$

Note that all these summands are finite, and we know from above $\mu_1(F_n \cap A) \leq \mu^*(F_n \cap A)$. It follows that necessarily $\mu_1(F_n \cap A) = \mu^*(F_n \cap A)$. By taking the supremum over n, we conclude $\mu_1(A) = \mu^*(A)$. $\square$

Corollary 2.18. *Let $\mathfrak{A}$ be a semi-ring over a set Ω, and μ_0 a measure on $\mathfrak{A}$. Let $\mathfrak{M}$ be the σ-algebra generated by $\mathfrak{A}$. Then there exists a measure $\mu : \mathfrak{M} \to [0, \infty]$ extending μ_0. If μ_0 is σ-finite, then μ is unique.*

Proof. We note that if $\mathfrak{R}$ is the ring generated by $\mathfrak{A}$, then $\mathfrak{M}$ is also the σ-algebra generated by $\mathfrak{R}$. So this is a direct consequence of Proposition 2.7, Proposition 2.10, Theorem 2.12, and Theorem 2.15. In the case of μ_0 being σ-finite, uniqueness of μ is exactly Theorem 2.17. □

2.3 Application: The Lebesgue measure on $\mathbb{R}$

Notation. In what follows, we will fix a bijection φ on a set Ω. We will also denote by φ the induced bijection on 2^Ω, which associates to every subset $E \subseteq \Omega$ its image $\varphi(E)$ under φ.

Proposition 2.19. *Let $\mathfrak{A}$ be semi-ring on a set Ω, and let $\mu_0 : \mathfrak{A} \to [0, \infty]$ be a σ-finite measure. Let $\mathfrak{R}$ be the ring generated by $\mathfrak{A}$, and $\mu : \mathfrak{R} \to [0, \infty]$ the unique extension to a measure. Suppose that φ is a bijection on Ω that restricts to a bijection on $\mathfrak{A}$, and suppose that $\mu_0 = \mu_0 \circ \varphi$. Then φ restricts to a bijection on $\mathfrak{R}$, and $\mu = \mu \circ \varphi$.*

Proof. This is immediate from the definition of both $\mathfrak{R}$ and μ, and is left as an exercise. □

Proposition 2.20. *Let $\mathfrak{R}$ be a ring on a set Ω, and let $\mu : \mathfrak{R} \to [0, \infty]$ be a measure. Suppose that φ is a bijection on Ω that restricts to a bijection on $\mathfrak{R}$, and suppose that $\mu = \mu \circ \varphi$. Then the outer measure μ^* satisfies $\mu^* \circ \varphi = \mu^*$. Moreover, φ restricts to a bijection on the σ-algebra of μ^*-measurable subsets of Ω.*

Proof. Let $A \subseteq \Omega$ be an arbitrary subset, and let $E_n \subseteq \Omega$ be any sequence of sets. Then clearly $A \subseteq \bigcup_{n \geq 1} E_n$ if and only if $\varphi(A) \subseteq \bigcup_{n \geq 1} \varphi(E_n)$. If φ defines a bijection on $\mathfrak{R}$, then also $E_n \in \mathfrak{R}$ if and only if $\varphi(E_n) \in \mathfrak{R}$. Furthermore we have from assumption that $\mu(\varphi(E_n)) = \mu(E_n)$. By the definition of μ^*, we immediately get $\mu^*(A) = \mu^*(\varphi(A))$.

Now assume additionally that E is μ^*-measurable. Then we have

$$\begin{aligned}
\mu^*(A) &= \mu^*(\varphi^{-1}(A)) \\
&= \mu^*(\varphi^{-1}(A) \cap E) + \mu^*(\varphi^{-1}(A) \setminus E) \\
&= \mu^*(A \cap \varphi(E)) + \mu^*(A \setminus \varphi(E)).
\end{aligned}$$

Since A is arbitrary, it follows that $\varphi(E)$ is μ^*-measurable. The reverse argument shows that if $\varphi(E)$ is μ^*-measurable, then E was μ^*-measurable to begin with. This finishes the proof. □

Proposition 2.21. *Consider the set $\mathfrak{A} \subseteq 2^{\mathbb{R}}$ of half-open intervals $\mathfrak{A} = \{(a,b] \mid a,b \in \mathbb{R},\ a \leq b\}$.*[8] *Then $\mathfrak{A}$ is a semi-ring, and the map $\mu_0 : \mathfrak{A} \to [0,\infty]$ given by $\mu_0\big((a,b]\big) = b - a$ is a measure on $\mathfrak{A}$.*

Proof. Evidently $\emptyset \in \mathfrak{A}$. We have $(a,b] \cap (c,d] = \Big(\max(a,c), \min(b,d)\Big]$, so $\mathfrak{A}$ is closed under intersections. The set-difference is given as a disjoint union $(a,b] \setminus (c,d] = (a,c] \cup (d,b]$, so we see that $\mathfrak{A}$ is a semi-ring. Evidently $\mu_0(\emptyset) = 0$, so we only need to show the σ-additivity.

Suppose that $E = (a,b] \in \mathfrak{A}$, and let $E_n = (a_n, b_n] \in \mathfrak{A}$ be a sequence of pairwise disjoint sets with $E = \bigcup_{n\geq 1} E_n$. Given any $M \geq 1$, we have in particular $\bigcup_{n=1}^{M} E_n \subset E$. By reordering E_n for $1 \leq n \leq M$ and discarding any empty sets if necessary, we may assume $b_n \leq a_{n+1}$ for all $n < M$. This leads to

$$\sum_{n=1}^{M} \mu_0(E_n) = \sum_{n=1}^{M} b_n - a_n \leq b_M - a_M + \sum_{n=1}^{M-1} a_{n+1} - a_n = b_M - a_1 \leq b - a = \mu_0(E).$$

Since M is arbitrary, this leads to $\sum_{n=1}^{\infty} \mu_0(E_n) \leq \mu_0(E)$. Let $\varepsilon > 0$ with $\varepsilon < b - a$. Then in particular

$$[a+\varepsilon, b] \subset (a,b] = \bigcup_{n\geq 1} (a_n, b_n] \subset \bigcup_{n\geq 1} (a_n, b_n + 2^{-n}\varepsilon).$$

The right-hand side is an open covering of the compact set on the left side, and hence there is some $N \geq 1$ such that $[a+\varepsilon, b] \subset \bigcup_{n=1}^{N} (a_n, b_n + 2^{-n}\varepsilon)$. We change the ordering of the intervals appearing in this union by the following inductive procedure: Choose $k_1 \in \{1, \dots, N\}$ to be the index so that

$$a_{k_1} = \max\left\{a_j \mid a + \varepsilon \in (a_j, b_j + 2^{-j}\varepsilon)\right\}.$$

If $b < b_{k_1} + 2^{-k_1}\varepsilon$, then the procedure stops here. Otherwise, choose $k_2 \in \{1, \dots, N\}$ to be the index so that

$$a_{k_2} = \max\left\{a_j \mid b_{k_1} + 2^{-k_1} \in (a_j, b_j + 2^{-j}\varepsilon)\right\}.$$

If $b < b_{k_2} + 2^{-k_2}\varepsilon$, then the procedure stops here. Otherwise one continues inductively until the procedure stops after $L \leq N$ steps. This yields an injective map $k : \{1, \dots, L\} \to \{1, \dots, N\}$ such that $[a+\varepsilon, b] \subset \bigcup_{n=1}^{L} (a_{k_n}, b_{k_n} + 2^{-k_n}\varepsilon)$ and such that for all $n < L$, we have $a_{k_{n+1}} < b_{k_n} + 2^{-k_n}\varepsilon$. From this we can

[8] By convention, we set $(a,b] := \emptyset$ if $a \geq b$.

deduce (with the help of the first part of the proof)

$$\begin{aligned}
b - a - \varepsilon &< b_{k_L} + 2^{-k_L}\varepsilon - a_{k_1} \\
&= b_{k_L} + 2^{-k_L}\varepsilon - a_{k_L} + \sum_{n=1}^{L-1} a_{k_{n+1}} - a_{k_n} \\
&< b_{k_L} + 2^{-k_L}\varepsilon - a_{k_L} + \sum_{n=1}^{L-1} b_{k_n} + 2^{-k_n} - a_{k_n} \\
&= \sum_{n=1}^{L} b_{k_n} + 2^{-k_n}\varepsilon - a_{k_n} \\
&\leq \varepsilon + \sum_{n=1}^{N} b_n - a_n \\
&\leq \varepsilon + b - a.
\end{aligned}$$

From this we may conclude

$$b - a \leq 2\varepsilon + \sum_{n=1}^{N} b_n - a_n \leq 2\varepsilon + \sum_{n=1}^{\infty} \mu_0(E_n).$$

Since $\varepsilon > 0$ was arbitrary, this finally implies $\mu_0(E) = \sum_{n=1}^{\infty} \mu_0(E_n)$ and finishes the proof. $\square$

Theorem 2.22 (Lebesgue)**.** *The measure $\mu_0 : \mathfrak{A} \to [0, \infty]$ defined on the semi-ring of half-open intervals $\mathfrak{A} \subset 2^{\mathbb{R}}$ extends to a translation-invariant measure $\lambda : \mathfrak{L} \to [0, \infty]$ on the σ-algebra $\mathfrak{L} \subset 2^{\mathbb{R}}$ of Lebesgue-measurable sets, which contains the Borel σ-algebra of $\mathbb{R}$. On the Borel σ-algebra, λ is the unique measure extending μ_0, and in fact the unique translation-invariant measure with $\lambda((0, 1]) = 1$.*

Proof. We use Proposition 2.7 and extend μ_0 to a measure $\mu : \mathfrak{R} \to [0, \infty]$ on the ring $\mathfrak{R} \subset 2^{\mathbb{R}}$ generated by $\mathfrak{A}$. We consider the outer measure μ^* on $\mathbb{R}$ as in Proposition 2.10. We appeal to Theorem 2.12 and define $\mathfrak{L}$ as the σ-algebra of μ^*-measurable sets (which we call Lebesgue-measurable), which contains $\mathfrak{A}$. Since $\mathfrak{A}$ generates the Borel σ-algebra, it follows that $\mathfrak{L}$ contains the Borel σ-algebra. By Theorem 2.15, the measure $\lambda = \mu^*|_{\mathfrak{L}}$ indeed extends μ on $\mathfrak{A}$.

Let us argue why λ is translation-invariant. Let $t \in \mathbb{R}$. Consider the bijection φ on $\mathbb{R}$ given by $\varphi(a) = t + a$. Then evidently φ restricts to a bijection on $\mathfrak{A}$, and we have $\mu = \mu \circ \varphi$. By combining Proposition 2.19 and Proposition 2.20, it follows that φ restrict to a bijection on $\mathfrak{L}$, and $\lambda = \lambda \circ \varphi$. In other words, we have $\lambda(E + t) = \lambda(E)$ for every Lebesgue-measurable set $E \in \mathfrak{L}$. Since $t \in \mathbb{R}$ is arbitrary, this shows the claim.

Lastly, the measure μ_0 on $\mathfrak{A}$ is σ-finite. Since $\mathfrak{A}$ generates the Borel σ-algebra, the uniqueness of the measure follows from Theorem 2.17. On the other hand, if we have a translation-invariant measure μ with $\mu((0,1]) = 1$, then it is easy to see that $\mu((0, \frac{1}{n}]) = \frac{1}{n}$ for all $n \geq 1$, which one can use to show that μ agrees with λ on all sets in $\mathfrak{A}$ with rational endpoints. If $a, b \in \mathbb{R}$ with $a < b$, pick some number $c \in \mathbb{Q}$ strictly between them. Choose decreasing sequences of rational numbers $a_n, b_n \in \mathbb{Q}$ such that $a_n \to a$ and $b_n \to b$. Then

$$(a, c] = \bigcup_{n \geq 1} (a_n, c], \quad (c, b] = \bigcap_{n \geq 1} (c, b_n].$$

By the continuity of measures, we can conclude that $\mu((a,b]) = b - a$, so μ agrees with λ on all of $\mathfrak{A}$, hence $\mu = \lambda$ on the Borel σ-algebra. $\square$

Remark 2.23. WARNING! The σ-algebra of Lebesgue sets is indeed bigger than the Borel σ-algebra.[9] Nevertheless, one sometimes refers to the Lebesgue measure to mean its restriction on the Borel σ-algebra, and denotes that also by λ. In the exercise sessions, we have already discussed an example of a set $A \subset [0,1]$ that is not even Lebesgue-measurable.

2.4 Application: Lebesgue–Stieltjes Measures

Definition 2.24. Let X be a locally compact, σ-compact Hausdorff space. A Borel measure on X is a measure on the Borel σ-algebra that assigns a finite value to every compact subset of X. In the special case $X = \mathbb{R}$, these are called Lebesgue–Stieltjes measures.

Proposition 2.25. *Let μ be a Lebesgue–Stieltjes measure. Then there exists a unique increasing right-continuous function $F : \mathbb{R} \to \mathbb{R}$ such that $F(0) = 0$ and $\mu((a,b]) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$, $a < b$.*

Proof. From these properties it follows that if such a function F exists at all, then it has to be given by the formula

$$F(t) = \begin{cases} \mu((0,t]) & , \quad t > 0 \\ 0 & , \quad t = 0 \\ -\mu((t,0]) & , \quad t < 0. \end{cases}$$

We claim that this function has indeed the right properties. Let $a, b \in \mathbb{R}$ with $a < b$. We aim to show $\mu((a,b]) = F(b) - F(a)$. If $a \geq 0$, then

[9] This is not so easy to see, but an example is discussed here, for whoever is interested: `https://www.math3ma.com/blog/lebesgue-but-not-borel`.

$\mu((a,b]) = \mu((0,b] \setminus (0,a]) = \mu((0,b]) - \mu((0,a]) = F(b) - F(a)$. If $b < 0$, we can prove this analogously. If $a < 0 \leq b$, then we have $\mu((a,b]) = \mu((a,0] \cup (0,b]) = \mu((a,0]) + \mu((0,b]) = F(b) - F(a)$.

The fact that F is increasing follows immediately from the fact that μ has nonnegative values. The right-continuity follows from

$$\lim_{n\to\infty} F(a+\varepsilon_n) = \lim_{n\to\infty} \mu((0,a+\varepsilon_n]) = \mu((0,a]) = F(a)$$

for any sequence $\varepsilon_n > 0$ with $\varepsilon_n \to 0$. Here we used the continuity property of μ as a measure with respect to countable decreasing intersections. $\square$

Theorem 2.26. *The assignment $\mu \mapsto F$, which assigns to every Lebesgue–Stieltjes measure its increasing right-continuous function as in Proposition 2.25, is a bijection. In particular, whenever $F : \mathbb{R} \to \mathbb{R}$ is an increasing right-continuous function with $F(0) = 0$, there exists a unique Lebesgue–Stieltjes measure μ such that $\mu((a,b]) = F(b) - F(a)$ for all $a, b \in \mathbb{R}$ with $a < b$.*

Proof. We have already seen that every Lebesgue–Stieltjes measure gives an increasing right-continuous function with these properties. Furthermore, the assignment $\mu \to F$ is injective. This is because F uniquely determined how μ is defined on the semi-ring of half-open intervals. Since μ restricted to this semi-ring is σ-finite, it follows from Theorem 2.17 that μ is hence uniquely determined by F.

It remains to be shown that for every choice of F, there exists a corresponding Lebesgue–Stieltjes measure. Indeed, let us define $\mathfrak{A}$ as the semi-ring of bounded half-open intervals. As in Proposition 2.21, we define $\mu : \mathfrak{A} \to [0,\infty]$ via $\mu((a,b]) = F(b) - F(a)$. For σ-additivity, assume $E = (a,b]$ can be expressed as the disjoint union of $E_n = (a_n, b_n]$. If $M \geq 1$ is any number and we reorder the E_n for $n \leq M$ to ensure $b_n \leq a_{n+1}$, then it follows from the fact that F is increasing that

$$\begin{aligned}
\sum_{n=1}^{M} \mu(E_n) &= \sum_{n=1}^{M} F(b_n) - F(a_n) \\
&\leq F(b_M) - F(a_M) + \sum_{n=1}^{M-1} F(a_{n+1}) - F(a_n) \\
&= F(b_M) - F(a_1) \\
&\leq F(b) - F(a) \;=\; \mu(E).
\end{aligned}$$

Since M was arbitrary, we have $\sum_{n=1}^{\infty} \mu(E_n) \leq \mu(E)$.

For the reverse inequality, choose $\varepsilon > 0$. We use right-continuity to pick for every $n \geq 1$ a small $\delta_n > 0$ such that $F(b_n + \delta_n) - F(b_n) \leq 2^{-n}\varepsilon$. Then

$$[a+\varepsilon, b] \subset E = \bigcup_{n=1}^{\infty} E_n \subset \bigcup_{n=1}^{\infty} (a_n, b_n + \delta_n)$$

Since the left side is compact, we obtain some $N \geq 1$ such that $[a+\varepsilon, b] \subset \bigcup_{n=1}^{N}(a_n, b_n + \delta_n)$. Now repeating exactly the same argument as in the proof of Proposition 2.21, we may deduce $F(b)-F(a+\varepsilon) \leq \varepsilon + \sum_{n=1}^{\infty} F(b_n)-F(a_n)$. Letting $\varepsilon \to 0$, we obtain the σ-additivity for μ.

The rest of the claim follows from the Carathéodory construction, exactly as in the proof of Theorem 2.22, see also Corollary 2.18. $\square$

Example 2.27. For the choice $F = \mathrm{id}_{\mathbb{R}}$, we recover the Lebesgue measure. On the other hand, if $a \in \mathbb{R}$ is some chosen number and we set

$$F(t) = \begin{cases} 0 & , \quad t < a \\ 1 & , \quad t \geq a, \end{cases}$$

then one can show that we recover the Dirac measure $\mu = \delta_a$. More generally, if μ is the measure corresponding to an increasing right-continuous function F, one can show that $\mu(\{a\}) = 0$ if and only if F is continuous at a.

2.5 Product measures

Definition 2.28. Let $(\Omega_i, \mathfrak{M}_i)$ be two measurable spaces for $i = 1, 2$. Then the product σ-algebra $\mathfrak{M}_1 \otimes \mathfrak{M}_2$ on $\Omega_1 \times \Omega_2$ is defined as the σ-algebra generated by all sets of the form $E_1 \times E_2$ for $E_1 \in \mathfrak{M}_1$ and $E_2 \in \mathfrak{M}_2$, the so-called measurable rectangles in $\Omega_1 \times \Omega_2$.

Proposition 2.29. *Let $(\Omega_i, \mathfrak{M}_i, \mu_i)$ be two measure spaces for $i = 1, 2$. Then the set of measurable rectangles*

$$\mathfrak{A} = \{E_1 \times E_2 \mid E_1 \in \mathfrak{M}_1,\ E_2 \in \mathfrak{M}_2\} \subseteq 2^{\Omega_1 \times \Omega_2}$$

is a semi-ring. Furthermore, the map $\mu_0 : \mathfrak{A} \to [0, \infty]$ given by $\mu_0(E_1 \times E_2) = \mu_1(E_1)\mu_2(E_2)$ is a measure on $\mathfrak{A}$.

Proof. This is proved in the exercise sessions. $\square$

Theorem 2.30. *Let $(\Omega_i, \mathfrak{M}_i, \mu_i)$ be two measure spaces for $i = 1, 2$. Then the measure μ_0 on the measurable rectangles extends to a measure on the product σ-algebra $\mu_1 \otimes \mu_2 : \mathfrak{M}_1 \otimes \mathfrak{M}_2 \to [0, \infty]$.*

Proof. This is a special case of Corollary 2.18. $\square$

Definition 2.31. Let $d \geq 1$. Then the Lebesgue measure on $\mathbb{R}^d$ is defined as the d-fold product measure

$$\lambda^{(d)} = \underbrace{\lambda \otimes \lambda \otimes \cdots \otimes \lambda}_{d \text{ times}}$$

with respect to the measure space $(\mathbb{R}, \mathfrak{L}, \lambda)$. The Lebesgue σ-algebra $\mathfrak{L}^{(d)}$ on $\mathbb{R}^d$ is the one consisting of all $\lambda^{(d)}$-measurable sets in the sense of Definition 2.11, which contains the Borel σ-algebra. If the dimension d is clear from context, we may sometimes slightly abuse notation and just write λ for the Lebesgue measure on $\mathbb{R}^d$.

Corollary 2.32. *The Lebesgue measure $\lambda^{(d)} : \mathfrak{L}^{(d)} \to [0,\infty]$ is translation invariant for all $d \geq 1$.*

Proof. We already know that the Lebesgue measure on $\mathbb{R}$ is translation invariant, so there is nothing to prove when $d = 1$. We proceed by induction and assume that $d \geq 2$ is a number so that $\lambda^{(d-1)}$ is translation invariant.

Let $t = (t_1, \dots, t_d) = (t', t_d) \in \mathbb{R}^d = \mathbb{R}^{d-1} \times \mathbb{R}$. Denote by μ_0 the product measure defined on the measurable rectangles. If $E \in \mathfrak{L}^{(d-1)}$ and $F \in \mathfrak{L}$ are measurable, then $(E \times F) + t = (E + t') \times (F + t_d)$, and so $\mu_0((E \times F) + t) = \lambda^{(d-1)}(E + t')\lambda(F + t_d) = \lambda^{(d-1)}(E)\lambda(F) = \mu_0(E \times F)$. It now follows directly from Proposition 2.19 and Proposition 2.20 that $\lambda^{(d)}(A + t) = \lambda^{(d)}(A)$ for all $A \in \mathfrak{L}^{(d)}$, concluding the proof. □

We conclude this section with Fubini's theorem, which is a fundamental result that tells us how to compute integrals over product measures.

Theorem 2.33 (Tonelli; see exercises)**.** *Let $(\Omega_i, \mathfrak{M}_i, \mu_i)$ be two σ-finite measure spaces for $i = 1, 2$. Let $f : \Omega_1 \times \Omega_2 \to [0, \infty]$ be a $\mathfrak{M}_1 \otimes \mathfrak{M}_2$-measurable function. Denote $f_x = f(x, \_) : \Omega_2 \to [0,\infty]$ and $f^y = f(\_, y) : \Omega_1 \to [0, \infty]$. Then:*

(a) *For every $x \in \Omega_1$, the function f_x is $\mathfrak{M}_2$-measurable.*

(b) *For every $y \in \Omega_2$, the function f^y is $\mathfrak{M}_1$-measurable.*

(c) *The function $\Omega_1 \to [0,\infty]$ given by $x \mapsto \int_{\Omega_2} f_x \, d\mu_2$ is $\mathfrak{M}_1$-measurable.*

(d) *The function $\Omega_2 \to [0,\infty]$ given by $y \mapsto \int_{\Omega_1} f^y \, d\mu_1$ is $\mathfrak{M}_2$-measurable.*

(e) *One has the equalities*

$$\int_{\Omega_1 \times \Omega_2} f \, d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \Big(\int_{\Omega_2} f_x \, d\mu_2 \Big) \, d\mu_1(x) = \int_{\Omega_2} \Big(\int_{\Omega_1} f^y \, d\mu_1 \Big) \, d\mu_2(y).$$

Theorem 2.34 (Fubini)**.** *Let $(\Omega_i, \mathfrak{M}_i, \mu_i)$ be two σ-finite measure spaces for $i = 1, 2$. Let $f : \Omega_1 \times \Omega_2 \to \mathbb{C}$ be a $\mathfrak{M}_1 \otimes \mathfrak{M}_2$-measurable function. Denote $f_x = f(x, \_) : \Omega_2 \to \mathbb{C}$ and $f^y = f(\_, y) : \Omega_1 \to \mathbb{C}$. Then the following are equivalent:*

(A) $\displaystyle\int_{\Omega_1\times\Omega_2} |f| \; d(\mu_1 \otimes \mu_2) < \infty;$

(B) $\displaystyle\int_{\Omega_1} \left(\int_{\Omega_2} |f_x| \; d\mu_2 \right) d\mu_1(x) < \infty;$

(C) $\displaystyle\int_{\Omega_2} \left(\int_{\Omega_1} |f^y| \; d\mu_1 \right) d\mu_2(y) < \infty.$

If any (or every) one of these statements holds, then we have that

(a) *The function $\Omega_1 \to \mathbb{C}$ given by $x \mapsto \int_{\Omega_2} f_x \; d\mu_2$ is well-defined on a conull set and μ_1-integrable.*

(b) *The function $\Omega_2 \to \mathbb{C}$ given by $y \mapsto \int_{\Omega_1} f^y \; d\mu_1$ is well-defined on a conull set and μ_2-integrable.*

(c) *One has the equalities*

$$\int_{\Omega_1\times\Omega_2} f \; d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f_x \; d\mu_2 \right) d\mu_1(x) = \int_{\Omega_2} \left(\int_{\Omega_1} f^y \; d\mu_1 \right) d\mu_2(y).$$

Proof. It follows from Tonelli's theorem that the integrals appearing in (A), (B), and (C) are always the same, so their finiteness are indeed equivalent. Now let us assume that these statements are true. By splitting f up into its real and imaginary parts, let us assume without loss of generality that f is a real function.

By part (c) in Tonelli's theorem and Proposition 1.38, it follows that the set E of all $x \in \Omega_1$, for which f_x is integrable, is in $\mathfrak{M}_1$ and its complement is a null set. Let $f = f^+ - f^-$ be the canonical decomposition into positive functions as in Remark 1.39. Then it is very easy to see that $(f^+)_x = (f_x)^+$ and $(f^-)_x = (f_x)^-$. So f_x^+ and f_x^- are both integrable whenever $x \in E$. Moreover the map

$$E \ni x \mapsto \int_{\Omega_2} f_x \; d\mu_2 = \int_{\Omega_2} f_x^+ \; d\mu_2 - \int_{\Omega_2} f_x^- \; d\mu_2$$

is $\mathfrak{M}_1$-measurable by part (c) in Tonelli's theorem, and this function is in fact μ_1-integrable because

$$\int_{\Omega_1} \left| \int_{\Omega_2} f_x \; d\mu_2 \right| d\mu_1(x) \leq \int_{\Omega_1} \left(\int_{\Omega_2} |f_x| \; d\mu_2 \right) d\mu_1(x) < \infty.$$

Hence it follows that

$$
\begin{aligned}
&\int_{\Omega_1} \left(\int_{\Omega_2} f_x \; d\mu_2 \right) d\mu_1(x) \\
= \;& \int_{\Omega_1} \chi_E \Big(\int_{\Omega_2} f_x \; d\mu_2 \Big) \; d\mu_1(x) \\
= \;& \int_{\Omega_1} \chi_E \Big(\int_{\Omega_2} f_x^+ \; d\mu_2 - \int_{\Omega_2} f_x^- \; d\mu_2 \Big) \; d\mu_1(x) \\
= \;& \int_{\Omega_1} \chi_E \Big(\int_{\Omega_2} f_x^+ \; d\mu_2 \Big) \; d\mu_1(x) - \int_{\Omega_1} \chi_E \Big(\int_{\Omega_2} f_x^- \; d\mu_2 \Big) \; d\mu_1(x) \\
= \;& \int_{\Omega_1} \Big(\int_{\Omega_2} f_x^+ \; d\mu_2 \Big) \; d\mu_1(x) - \int_{\Omega_1} \Big(\int_{\Omega_2} f_x^- \; d\mu_2 \Big) \; d\mu_1(x) \\
= \;& \int_{\Omega_1 \times \Omega_2} f^+ \; d(\mu_1 \otimes \mu_2) - \int_{\Omega_1 \times \Omega_2} f^- \; d(\mu_1 \otimes \mu_2) \\
= \;& \int_{\Omega_1 \times \Omega_2} f \; d(\mu_1 \otimes \mu_2).
\end{aligned}
$$

The remaining equality follows exactly in the same way by exchanging the roles of Ω_1 and Ω_2. □

2.6 Infinite products of probability measures

Lemma 2.35. *Let Ω be a non-empty set and $\mathfrak{A} \subseteq 2^\Omega$ a semi-ring with $\Omega \in \mathfrak{A}$. Suppose that $\mu : \mathfrak{A} \to [0,1]$ is a map with $\mu(\Omega) = 1$. Then μ is a measure if and only if for every sequence $A_n \in \mathfrak{A}$ of pairwise disjoint sets with $\Omega = \bigcup_{n \geq 1} A_n$, one has $\sum_{n=1}^\infty \mu(A_n) = 1$.*

Proof. Clearly any measure must satisfy this property, so the "only if" part is clear.

For the "if" part, we may first take $A_1 = \Omega$ and $A_n = \emptyset$ for all $n \geq 2$, which immediately implies $\mu(\emptyset) = 0$. We need to show that μ is σ-additive. Let $B_n \in \mathfrak{A}$ be a sequence of pairwise disjoint sets such that $B = \bigcup_{n \geq 1} B_n \in \mathfrak{A}$. As $\mathfrak{A}$ is a semi-ring, we have $\Omega \setminus B = A_1 \cup \cdots \cup A_\ell$ for pairwise disjoint sets $A_1, \dots, A_\ell \in \mathfrak{A}$. Then the assumption implies on the one hand that $1 = \mu(B) + \sum_{n=1}^\ell \mu(A_n)$. On the other hand, if we set $A_{\ell+k} = B_k$ for $k \geq 1$, then the sequence $(A_n)_{n \geq 1}$ defines a pairwise disjount covering of Ω, so $1 = \sum_{n=1}^\infty \mu(A_n)$. Comparing these two equations, we see that $\mu(B) = \sum_{n > \ell} \mu(A_n) = \sum_{n=1}^\infty \mu(B_n)$, which shows our claim. □

Definition 2.36. Let I be a non-empty index set and $(\Omega_i, \mathfrak{M}_i, \mu_i)$ a probability space for every $i \in I$. We denote $\Omega = \prod_{i \in I} \Omega_i$, and set

$$
\mathfrak{A} = \left\{ \prod_{i \in I} E_i \mid E_i \in \mathfrak{M}_i \text{ for all } i \in I, \; E_i = \Omega_i \text{ for all but finitely many } i \in I \right\}.
$$

The σ-algebra generated by $\mathfrak{A}$ is denoted by $\bigotimes_{i\in I} \mathfrak{M}_i$. We consider the map $\mu : \mathfrak{A} \to [0,1]$ given by $\mu(\prod_{i\in I} E_i) = \prod_{i\in I} \mu_i(E_i)$. Note that these products are well-defined as all but finitely many factors are assumed to be 1.

Theorem 2.37. *Adopt the notation from the above definition. Then $\mathfrak{A}$ is a semi-ring and μ is a measure on $\mathfrak{A}$. In particular, it extends uniquely to a probability measure μ on $\bigotimes_{i\in I} \mathfrak{M}_i$. One also denotes $\mu = \bigotimes_{i\in I} \mu_i$.*

Proof. The "in particular" part is due to Corollary 2.18. By definition, every set in $\mathfrak{A}$ is a product of subsets of the spaces Ω_i which are proper subsets only over finitely many indices. In particular, given any sequence A_n, there are countably many indices in I so that over every other index $i \in I$, the projection of every set A_n to the i-th coordinate yields Ω_i. Considering the semi-ring axioms for $\mathfrak{A}$ and the axioms of being a measure for μ, we can see that it is enough to consider the case where I is countable. As we already know that the claim is true if I is finite, we may from now on assume $I = \mathbb{N}$.

The fact that $\mathfrak{A}$ is a semi-ring now follows directly from the exercises. We hence need to show that μ is a measure on $\mathfrak{A}$, where our goal is to appeal to the condition in the above lemma. Suppose that $A_n \in \mathfrak{A}$ is a sequence of pairwise disjoint sets with $\Omega = \bigcup_{n\geq 1} A_n$. It is our intention to show $\sum_{n=1}^\infty \mu(A_n) = 1$.

For each $n \geq 1$ let us write $A_n = \prod_{i=1}^\infty A_{n,i}$ and pick $i_n \geq 1$ such that $A_{n,i} = \Omega_i$ whenever $i > i_n$. For all $m, n \in \mathbb{N}$ and $\omega = (\omega_i)_i \in A_m$ we claim to have the equation

$$\prod_{i\leq i_m} \chi_{A_{n,i}}(\omega_i) \cdot \prod_{i>i_m} \mu_i(A_{n,i}) = \delta_{n,m}.$$

Indeed, if $n = m$, then all the involved factors are equal to one, so the equation holds. Assume $n \neq m$. We observe $\chi_{A_k}((\omega_i)_i) = \prod_{i\leq i_k} \chi_{A_{k,i}}(\omega_i)$. As $1 = \sum_{k=1}^\infty \chi_{A_k}$, we conclude for $(\omega_i)_i \in A_m$ that $0 = \prod_{i\leq i_n} \chi_{A_{n,i}}(\omega_i)$. So either $i_n \leq i_m$, in which case the desired inequality above follows immediately. Or, if $i_n > i_m$, then every tuple of the form $(\omega_1, \dots, \omega_{i_m}, \alpha_{i_m+1}, \dots)$ is also in A_m, so analogously

$$0 = \prod_{i\leq i_m} \chi_{A_{n,i}}(\omega_i) \cdot \prod_{i=i_m+1}^{i_n} \chi_{A_{n,i}}(\alpha_i).$$

By consecutively integrating over α_i for $i = i_m + 1, \dots, i_n$, we recover the equation we wish to show. Finally, let us assume that $\sum_{n=1}^\infty \mu(A_n) \neq 1$. We have

$$\begin{aligned}\sum_{n=1}^{\infty} \mu(A_n) &= \sum_{n=1}^{\infty} \prod_{i=1}^{\infty} \mu_i(A_{n,i}) \\ &= \sum_{n=1}^{\infty} \int_{\Omega_1} \chi_{A_{n,1}}(\omega_1) \cdot \Big(\prod_{i=2}^{\infty} \mu_i(A_{n,i}) \Big) \, d\mu_1(\omega_1) \\ &= \int_{\Omega_1} \sum_{n=1}^{\infty} \chi_{A_{n,1}}(\omega_1) \cdot \Big(\prod_{i=2}^{\infty} \mu_i(A_{n,i}) \Big) \, d\mu_1(\omega_1)\end{aligned}$$

As this quantity is not 1, there must be some $\omega_1 \in \Omega_1$ for which

$$\sum_{n=1}^{\infty} \chi_{A_{n,1}}(\omega_1) \cdot \Big(\prod_{i=2}^{\infty} \mu_i(A_{n,i}) \Big) \neq 1.$$

By iterating this process, we can come up with a tuple $\omega = (\omega_i)_i \in \Omega$ satisfying

$$\sum_{n=1}^{\infty} \Big[\prod_{i \leq k} \chi_{A_{n,i}}(\omega_i) \cdot \prod_{i > k} \mu_i(A_{n,i}) \Big] \neq 1$$

for all $k \geq 1$. However, since $\omega \in A_m$ for some m, it follows from our previous observation for $k = i_m$ that precisely one of these summands is 1 and the others are zero. This leads to a contradiction. Hence we have indeed that $\sum_{n=1}^{\infty} \mu(A_n) = 1$. □

3 Probability

3.1 Probability spaces and random variables

Definition 3.1. A probability space is a measure space $(\Omega, \mathfrak{M}, \mu)$ with $\mu(\Omega) = 1$. In this context,

- μ is called a probability measure,
- Ω is called a sample space,
- the elements of Ω are called outcomes,
- the elements of $\mathfrak{M}$ are called events,
- the value $\mu(A)$ is called the probability of (the event) A. (To be even more suggestive, one often writes $\mathbb{P}$ instead of μ.)

The following easy special case illustrates how the most elementary probabilistic experiments, such as tossing a coin only finitely many times, can be thought of in this framework. We omit the easy proof.

Proposition 3.2. *Let Ω be a non-empty countable set and $\mathfrak{M} = 2^\Omega$. Let $p : \Omega \to [0,1]$ be any map with the property $\sum_{\omega\in\Omega} p(\omega) = 1$. Then the assignment $2^\Omega \ni A \mapsto \mathbb{P}(A) = \sum_{\omega\in A} p(\omega)$ defines a probability measure. Conversely, if $\mathbb{P}$ is a probability measure on Ω, then $p(\omega) = \mathbb{P}(\{\omega\})$ defines a map with $\sum_{\omega\in\Omega} p(\omega) = 1$.*

For more involved probabilistic questions, the sample space is not necessarily countable, which makes it somewhat more difficult to come up with the right choices of events and probability measures. The following can be seen as a model for *tossing a coin infinitely often.*

Example 3.3 (infinite coin tossing)**.** For each toss, a coin can only come up as heads or tails, which we conveniently denote as the outcomes 0 and 1. Since we want to model tossing the coin infinitely often, the outcomes are sequences having value 0 or 1. This gives rise to the sample space $\Omega = \{0,1\}^{\mathbb{N}}$.

A rather obvious example of an event is when the first k tosses are equal to some k-tuple $\omega \in \{0,1\}^k$. The associated subset of Ω is given as $A_\omega = \{\omega\} \times \{0,1\}^{\mathbb{N}^{>k}}$. For example, A_1 is the event where the first coin toss comes up as tails, or $A_{0,1,1}$ is the event where the first three tosses come up as heads $\to$ tails $\to$ tails. Let $\mathfrak{M}$ be the event σ-algebra generated by all such sets. We note that a lot of other natural choices for events are automatically in $\mathfrak{M}$. For example, the event that the n-th coin toss comes up as heads is given by

$$\{0,1\}^{n-1} \times \{0\} \times \{0,1\}^{\mathbb{N}^{>n}} = \bigcup_{\omega\in\{0,1\}^{n-1}} A_{(\omega,0)}.$$

As for determining the probability, we want all of our coin tosses to be *fair*, meaning that there is always an equal chance that either heads or tails comes up. In particular, the coin tosses should all be independent from each other. If $\mu : \mathfrak{M} \to [0,1]$ is supposed to be a probability measure modelling this behavior, then we can agree on $\mu(A_0) = \mu(A_1) = \frac{1}{2}$. Inductively, we may conclude that for all $\omega \in \{0,1\}^n$, one should have $\mu(A_\omega) = 2^{-n}$. In this case, if we view $\{0,1\}$ as a discrete space, we may give Ω the product topology, in which case $\mathfrak{M}$ is the Borel σ-algebra. If $\mu_2 : \{0,1\} \to [0,1]$ is the measure that assigns the value $\frac{1}{2}$ to both singletons, then μ is in fact the infinite product measure $\mu = \bigotimes_{n=1}^\infty \mu_2$.

Definition 3.4. Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space and W a topological space. A W-valued random variable X is a measurable map $X : \Omega \to W$.

For a Borel subset $B \subseteq W$, we will frequently use the popular notation $(X \in B)$ for the preimage $X^{-1}(B)$.

Remark 3.5. One sometimes also says "stochastic variable". In the particular case $W = \mathbb{R}^n$, one calls it a vector random variable, and for $W = \mathbb{R}$, a real random variable. The case $W = \mathbb{R}^{\mathbb{N}}$ is referred to as a random sequence. In the case of a real random variable X, we will also freely play around with the above notation, for example $(|X| \leq 1)$ is written instead of $(X \in [-1, 1])$, etc.

Remark 3.6. From our previous study on measurable maps we can observe that real random variables are closed under addition and multiplication, and pointwise limits. General random variables are closed under the same type of operations under which measurable maps are closed.

Definition 3.7. Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space and X a W-valued random variable. The distribution of X is the Borel probability measure $\mathbb{P}_X$ on W given by $\mathbb{P}_X(B) = \mathbb{P}(X \in B)$. If X and Y are two W-valued random variables (defined on possibly different probability spaces), we say that X and Y are identically distributed if $\mathbb{P}_X = \mathbb{P}_Y$.

Definition 3.8. Let X be a real random variable. Then its (cumulative) distribution function $F_X : \mathbb{R} \to \mathbb{R}$ is defined as $F_X(t) = \mathbb{P}_X((-\infty, t]) = \mathbb{P}_X(X \leq t)$.

Indeed, in the situation above, the measure $\mathbb{P}_X$ on $\mathbb{R}$ is a Lebesgue–Stieltjes measure, so we know from the previous chapter that F_X is an increasing and right-continuous function.

Definition 3.9. Let X be a real random variable on a probability space $(\Omega, \mathfrak{M}, \mathbb{P})$. We say that X has an expected value (or mean), if X is integrable as a function $X : \Omega \to \mathbb{R}$. In this case, its expected value (or mean) is defined as

$$\mathbb{E}(X) = \int_\Omega X \, d\mathbb{P}.$$

Proposition 3.10. *Let X be a real random variable. Then X has an expected value if and only if $id : \mathbb{R} \to \mathbb{R}$ is $\mathbb{P}_X$-integrable. In this case, its expected value is $\mathbb{E}(X) = \int_{\mathbb{R}} x \, d\mathbb{P}_X(x)$. In particular, the expected value only depends on the distribution of X.*

Proof. This is a direct consequence of Theorem 1.54 and Remark 1.55. □

Proposition 3.11. *Let W be a topological space and X a W-valued random variable with distribution $\mathbb{P}_X$. Let $f : W \to \mathbb{R}$ be a Borel measurable function. Then $f \circ X = f(X)$ has an expected value if and only if f is $\mathbb{P}_X$-integrable. In that case, $\mathbb{E}(f(X)) = \int_W f \, d\mathbb{P}_X$.*

Notation 3.12. Certain expected values get a special name. Let X be a real random variable.

(i) Given $k \geq 0$, we call $\mathbb{E}(|X|^k)$ the k-th absolute moment. If it exists, then $\mathbb{E}(X^k)$ is called the k-th moment.

(ii) Suppose X has an expected value $m_X = \mathbb{E}(X)$. Then $\mathrm{Var}(X) = \mathbb{E}((X - m_X)^2)$ is called the variance of X.

(iii) More generally, suppose X and Y are two real random variables on the same probability space, for which the second moments exist. Then it follows from Hölder's inequality that all of X, Y, XY have expected values, and the value $\mathrm{Cov}(X, Y) = \mathbb{E}((X - m_X)(Y - m_Y))$ is called the covariance of X and Y.

3.2 Independence

Definition 3.13. Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space. Two events $A, B \in \mathfrak{M}$ are called independent, if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Example 3.14. In the context of seeing measure theory as a way to assign a "volume" to certain sets, this notion has no clear meaning. The concept does however become relevant if one adopts the probabilistic point of view. For example, keep in mind Example 3.3 modelling the infinite coin tossing. In that context, we may for example define A to be the event where both the first and second toss comes up as heads, and B the event where both the third and fourth toss comes up as tails. In other words,

$$A = \{0\} \times \{0\} \otimes \{0,1\}^{\mathbb{N}^{\geq 3}}, \quad B = \{0,1\} \times \{0,1\} \times \{1\} \times \{1\} \otimes \{0,1\}^{\mathbb{N}^{\geq 5}}.$$

In our model we assume that the individual coin tosses are not supposed to influence each other, and hence we should certainly view these events as independent. Indeed, we have here

$$A \cap B = \{0\} \times \{0\} \times \{1\} \times \{1\} \otimes \{0,1\}^{\mathbb{N}^{\geq 5}},$$

so by the properties of the product measure μ we can see here that $\mu(A) = \frac{1}{4}$, $\mu(B) = \frac{1}{4}$, and $\mu(A \cap B) = \frac{1}{16}$. Similarly, all pairs of events defined by the outcomes of coin tosses happening at distinct times will be independent in this model.

Definition 3.15. Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space. A family of events $\{A_i\}_{i \in I} \subseteq \mathfrak{M}$ is called independent, if for all pairwise distinct indices $i_1, \dots, i_n \in I$, one has $\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_i}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_n})$.

Definition 3.16. Let Ω be a non-empty set. For a sequence $A_n \in 2^\Omega$, we define

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k \quad \text{and} \quad \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k.$$

Example 3.17. Let us motivate this again from the point of view of our infinite coin tossing model. Let H_n be the event that where the n-th toss comes up heads. Then the event $\limsup_{n\to\infty} H_n$ describes the situation where heads comes up infinitely many times, and the event $\liminf_{n\to\infty} H_n$ describes the situation where heads comes up in all but finitely many tosses. For these reasons, it is not uncommon in a probabilistic context to use the notation

$$\limsup_{n\to\infty} A_n = (A_n, \text{infinitely often}) = (A_n, \text{i.o.})$$

and

$$\liminf_{n\to\infty} A_n = (A_n, \text{almost always}) = (A_n, \text{a.a.}).$$

Proposition 3.18. *Let* $(\Omega, \mathfrak{M}, \mathbb{P})$ *be a probability space, and* $A_n \in \mathfrak{M}$ *a sequence. Then*

$$\mathbb{P}(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_{n\to\infty} A_n).$$

Proof. See exercises. □

Lemma 3.19 (Borel–Cantelli Lemma)**.** *Let* $(\Omega, \mathfrak{M}, \mathbb{P})$ *be a probability space, and* $A_n \in \mathfrak{M}$ *a sequence.*

(i) *Suppose* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. *Then* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.

(ii) *Suppose* $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. *If* $\{A_n\}_{n\in\mathbb{N}}$ *is an independent family, then* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Proof. We prove (i) in the exercise sessions, so we only need to prove (ii).

We will use the fact from the exercise sessions that the family of complements $\{A_n^c\}_{n\in\mathbb{N}}$ is also independent. Moreover, we are about to use the well-known inequality $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$. We observe for all $n \geq 1$

that

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{k=n}^{\infty} A_k^c\Big) &= \lim_{m\to\infty} \mathbb{P}\Big(\bigcap_{k=n}^{m} A_k^c\Big) \\
&= \lim_{m\to\infty} \prod_{k=n}^{m} \mathbb{P}(A_k^c) \\
&= \lim_{m\to\infty} \prod_{k=n}^{m} (1-\mathbb{P}(A_k)) \\
&\leq \lim_{m\to\infty} \prod_{k=n}^{m} \exp(-\mathbb{P}(A_k)) \\
&= \lim_{m\to\infty} \exp\Big(-\sum_{k=n}^{m} \mathbb{P}(A_k)\Big) \\
&= 0.
\end{aligned}$$

We conclude that $\bigcup_{n\geq 1}\bigcap_{k\geq n} A_k^c = \liminf_{n\to\infty} A_n^c$ is a null set. Therefore $\limsup_{n\to\infty} A_n = (\liminf_{n\to\infty} A_n^c)^c$ is co-null, which shows our claim. □

Example 3.20. Let us have yet another look at our infinite coin tossing model and what we observed in Example 3.14. Let H_n denote the event where the n-th coin toss comes up as heads. Then the family $\{H_n\}_{n\in\mathbb{N}}$ is independent, each event has probability $\frac{1}{2}$, and hence it follows from the above that the event $(H_n, \text{infinitely often})$ has probability 1.

Definition 3.21. Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space. A family of sub-σ-algebras $\mathfrak{M}_i \subseteq \mathfrak{M}$ for $i \in I$ is called independent, if for every collection $\{A_i\}_{i\in I} \subseteq \mathfrak{M}$ with $A_i \in \mathfrak{M}_i$ is independent.

Definition 3.22. Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space. Let I be an index set, and let X_i be a random variable on $(\Omega, \mathfrak{M}, \mathbb{P})$ with values in the topological space W_i for $i \in I$. We call the family $\{X_i\}_{i\in I}$ independent, if for every collection $\{S_i\}_{i\in I}$ of Borel sets $S_i \subseteq W_i$, we have that the family of events $\{(X_i \in S_i)\}_{i\in I}$ is independent. (In other words, if we denote by $\mathfrak{B}_i$ the Borel-σ-algebra on W_i, this condition means that the family of pullback σ-algebras $\{X_i^*(\mathfrak{B}_i)\}_{i\in I}$ is independent in the above sense.)

Theorem 3.23. *Let $\{\mathbb{P}_i\}_{i\in I}$ be a family of Borel probability measures on $\mathbb{R}$. Then there exists a probability space $(\Omega, \mathfrak{M}, \mathbb{P})$ with an independent family of real random variables $\{X_i\}_{i\in I}$ such that $\mathbb{P}_i = \mathbb{P}_{X_i}$.*

Proof. We set $(\Omega, \mathfrak{M}, \mathbb{P}) = \bigotimes_{i\in I}(\mathbb{R}, \mathfrak{B}, \mathbb{P}_i)$ and define $X_i : \Omega \to \mathbb{R}$ to be the projection onto the i-th component. It is an easy exercise to see that this yields an independent family with the desired property. □

Since the proof of the following is very easy, we leave it as an exercise.

Proposition 3.24. *Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space. Let I be an index set, and let X_i be a random variable on $(\Omega, \mathfrak{M}, \mathbb{P})$ with values in the topological space W_i for $i \in I$. For each $i \in I$, let V_i be another topological space and $f_i : W_i \to V_i$ a Borel measurable map. If the family of random variables $\{X_i\}_{i\in I}$ is independent, then so is $\{f_i(X_i)\}_{i\in I}$.*

Proposition 3.25. *Let X and Y be two mutually independent, real random variables defined on a probability space $(\Omega, \mathfrak{M}, \mathbb{P})$. If both X and Y have a mean, then so does XY, and in fact $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.*

Proof. Consider the pullback σ-algebras

$$\mathfrak{M}_X = \{(X \in S) \mid S \subseteq \mathbb{R} \text{ Borel}\}, \quad \mathfrak{M}_Y = \{(Y \in S) \mid S \subseteq \mathbb{R} \text{ Borel}\},$$

which are both sub-σ-algebras of $\mathfrak{M}$. Given any $A \in \mathfrak{M}_X$ and $B \in \mathfrak{M}_Y$, it follows by independence that

$$\int_\Omega \chi_A \chi_B \, d\mathbb{P} = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \int_\Omega \chi_A \, d\mathbb{P} \cdot \int_\Omega \chi_B \, d\mathbb{P}.$$

By linearity, it follows that for any $\mathfrak{M}_X$-measurable simple function $s : \Omega \to [0, \infty)$ and $\mathfrak{M}_Y$-measurable simple function $t : \Omega \to [0, \infty)$, one has the equation

$$\int_\Omega st \, d\mathbb{P} = \int_\Omega s \, d\mathbb{P} \cdot \int_\Omega t \, d\mathbb{P}.$$

By the Monotone Convergence Theorem, this even holds when s and t are not assumed to be simple. Applying this to $s = |X|$ and $t = |Y|$ yields $\mathbb{E}(|XY|) = \mathbb{E}(|X|)\mathbb{E}(|Y|) < \infty$, so indeed XY has a mean. Furthermore if we decompose $X = X^+ - X^-$ and $Y = Y^+ - Y^-$, then

$$\begin{aligned}
\mathbb{E}(XY) &= \textstyle\int_\Omega XY \, d\mathbb{P} \\
&= \textstyle\int_\Omega (X^+ - X^-)(Y^+ - Y^-) \, d\mathbb{P} \\
&= \textstyle\int_\Omega X^+Y^+ + X^-Y^- - X^-Y^+ - X^+Y^- \, d\mathbb{P} \\
&= \mathbb{E}(X^+)\mathbb{E}(Y^+) + \mathbb{E}(X^-)\mathbb{E}(Y^-) - \mathbb{E}(X^-)\mathbb{E}(Y^+) - \mathbb{E}(X^+)\mathbb{E}(Y^-) \\
&= (\mathbb{E}(X^+) - \mathbb{E}(X^-))(\mathbb{E}(Y^+) - \mathbb{E}(Y^-)) \\
&= \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

□

3.3 Law of Large Numbers

Before we come to the main point of this subsection, we need to cover some observations that are useful in computations.

Theorem 3.26 (Chebyshev's inequality)**.** *Let X be a real random variable on the probability space $(\Omega, \mathfrak{M}, \mathbb{P})$. Suppose that $f : \mathbb{R} \to \mathbb{R}^{\geq 0}$ is an even measurable function whose restriction to $\mathbb{R}^{\geq 0}$ is increasing. Let $a \geq 0$ be any number with $f(a) > 0$. Then*

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{f(a)} \mathbb{E}(f(X)).$$

Proof. We simply observe:

$$\begin{aligned}
\mathbb{E}(f(X)) &= \int_\Omega f \circ X \; d\mathbb{P} \\
&\geq \int_\Omega (f \circ X)\chi_{|X| \geq a} \; d\mathbb{P} \\
&\geq f(a) \int_\Omega \chi_{|X| \geq a} \; d\mathbb{P} \\
&= f(a)\mathbb{P}(|X| \geq a).
\end{aligned}$$

□

Corollary 3.27. *Let X be a real random variable on the probability space $(\Omega, \mathfrak{M}, \mathbb{P})$.*

(i) For all $a, p > 0$, one has $\mathbb{P}(|X| \geq a) \leq \dfrac{1}{a^p}\mathbb{E}(|X|^p)$.

(ii) If X has an expected value, then for all $a > 0$, one has

$$\mathbb{P}(|X - m_X| \geq a) \leq \frac{1}{a^2} \operatorname{Var}(X).$$

Proof. The first part follows from the general Chebyshev inequality for $f(x) = |x|^p$. The second part follows when applying it to $X - m_X$ as the real random variable and the function $f(x) = x^2$. □

Definition 3.28. Let X_n be a sequence of real random variables defined on a probability space $(\Omega, \mathfrak{M}, \mathbb{P})$. Given a real random variable X on the same probability space, we say that X_n converges to X in probability, if for all $\varepsilon > 0$, one has $\mathbb{P}(|X - X_n| > \varepsilon) \xrightarrow{n\to\infty} 0$. One writes $X_n \xrightarrow{p} X$.

Theorem 3.29 (Weak Law of Large Numbers)**.** *Let X_n be a sequence of identically distributed real random variables defined on a probability space $(\Omega, \mathfrak{M}, \mathbb{P})$ that are pairwise independent. Suppose that every (or any) X_n has a mean $m \in \mathbb{R}$. Then it follows that $\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{p} m$.*

Proof. Since they are identically distributed, it follows that both the mean and the variance of X_n are the same for every $n \geq 1$. We may assume without loss of generality $m = 0$. Let us first assume that the variance of X_n is finite and equal to σ^2 for all $n \geq 1$. For every $n \geq 1$, denote $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. We have that $\bar{X}_n$ also has mean 0, and hence

$$\operatorname{Var}(\bar{X}_n) = \mathbb{E}(\bar{X}_n^2) = \frac{1}{n^2}\mathbb{E}\Big(\sum_{j,k=1}^n X_j X_k\Big) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Here we used Proposition 3.25 for $j \neq k$. By applying Corollary 3.27(ii), we may hence see

$$\mathbb{P}(|\bar{X}_n| \geq \varepsilon) \leq \frac{\mathbb{E}(\bar{X}_n^2)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \overset{n\to\infty}{\longrightarrow} 0$$

for all $\varepsilon > 0$.

Now for the general case, we allow the possibility that X_n has infinite variance. Let us still assume $m = 0$. For all $n, N \geq 1$, write $X_n = X_n^{\leq N} + X_n^{>N}$, where $X_n^{\leq N} = X_n \cdot \chi_{(|X_n| \leq N)}$. Then by the Monotone Convergence Theorem

$$a_N := \mathbb{E}(|X_n^{>N}|) = \mathbb{E}(|X_n|) - \mathbb{E}(|X_n^{\leq N}|) \overset{N\to\infty}{\longrightarrow} 0,$$

where we are using that these values do not depend on n as the X_n were identically distributed. Let us also denote

$$\bar{X}_n^{\leq N} = \frac{1}{n}\sum_{j=1}^n X_j^{\leq N} \quad \text{and} \quad \bar{X}_n^{>N} = \frac{1}{n}\sum_{j=1}^n X_j^{>N}.$$

We have in particular for any $\varepsilon > 0$ that

$$\mathbb{P}(|\bar{X}_n^{>N}| \geq \varepsilon) \leq \frac{\mathbb{E}(|\bar{X}_n^{>N}|)}{\varepsilon} \leq \frac{a_N}{\varepsilon}.$$

So, given any $\delta > 0$ with $\delta < 1$ and any $N \geq 1$ large enough such that $a_N \leq \frac{1}{2}\delta\varepsilon$ for all $n \geq 1$, it follows that $\mathbb{P}(|\bar{X}_n^{>N}| \geq \varepsilon/2) \leq \delta$. Hence

$$\begin{aligned}
\mathbb{P}(|\bar{X}_n| \geq \varepsilon) &\leq \mathbb{P}\Big(|\bar{X}_n^{>N}| + |\bar{X}_n^{\leq N}| \geq \varepsilon\Big) \\
&\leq \delta + \mathbb{P}\Big(|\bar{X}_n^{>N}| + |\bar{X}_n^{\leq N}| \geq \varepsilon \wedge |\bar{X}_n^{>N}| \leq \varepsilon/2\Big) \\
&\leq \delta + \mathbb{P}\Big(|\bar{X}_n^{\leq N}| \geq \varepsilon/2\Big) \\
&\leq \delta + \mathbb{P}\Big(|\bar{X}_n^{\leq N} - (\mathbb{E}(\bar{X}_n^{\leq N}) + \mathbb{E}(\bar{X}_n^{>N}))| \geq \varepsilon/2\Big) \\
&\leq \delta + \mathbb{P}\Big(|\bar{X}_n^{\leq N} - \mathbb{E}(\bar{X}_n^{\leq N})| \geq \tfrac{\varepsilon}{2}(1-\delta)\Big) \\
&\overset{n\to\infty}{\longrightarrow} \delta.
\end{aligned}$$

Here we have used that $X_n^{\leq N}$ has finite variance and the above falls into our previous subcase. As $\delta > 0$ was arbitrary, this verifies $\bar{X}_n \xrightarrow{p} 0$ in general. $\square$

There is also a stronger law, i.e., a theorem with a stronger conclusion than the weak law of large numbers. We will prove this strong law under an additional assumption.[10]

Theorem 3.30 (Strong Law of Large Numbers)**.** *Let X_n be an independent sequence of identically distributed real random variables defined on a probability space $(\Omega, \mathfrak{M}, \mathbb{P})$. Suppose that every (or any) X_n has a fourth moment, meaning that $\mathbb{E}(X_n^4) < \infty$. If $m = \mathbb{E}(X_n)$ is the mean of all these random variables, then it follows that $\mathbb{P}\Big(\frac{1}{n}\sum_{k=1}^n X_k \to m\Big) = 1$.*

Proof. We first note that applying Hölder's inequality twice yields $\mathbb{E}(|X_n|)^4 \leq \mathbb{E}(X_n^2)^2 \leq \mathbb{E}(X_n^4) =: M_4$, hence the assumptions on X_n ensure that m exists.

As before, we assume without loss of generality $m = 0$ by replacing each X_n by $X_n - m$, if necessary. Denote $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Then

$$\mathbb{E}(\bar{X}_n^4) = \frac{1}{n^4}\sum_{i,j,k,l=1}^n \mathbb{E}(X_i X_j X_k X_l).$$

Considering the summand over the tuple (i, j, k, l), it follows from the independence of the sequence X_n and Proposition 3.25 that it is zero if there is one entry in this tuple that is different from all the other three. In other words, the summand can only be non-zero if all four entries agree, or if the tuple has two different indices occuring exactly twice. Two different entries can possibly occur in the pattern (k, k, l, l), (k, l, k, l) or (k, l, l, k), leading to $3n(n-1)$ possible summands. Using Proposition 3.25 once again, the expression hence simplifies to

$$\mathbb{E}(\bar{X}_n^4) = \frac{1}{n^4}\Big(\sum_{k=1}^n \mathbb{E}(X_k^4) + 3\sum_{\substack{k,l=1\\k\neq l}}^n \mathbb{E}(X_k^2)\mathbb{E}(X_l^2)\Big)$$

By the very beginning of the proof, each summand in the bracket can be estimated above by M_4, and there are a total of $n + 3n(n-1) \leq 3n^2$ summands, hence

$$\mathbb{E}(\bar{X}_n^4) \leq \frac{3n^2 M_4}{n^4} = \frac{3M_4}{n^2}.$$

If we apply Corollary 3.27 for $p = 4$, it follows for every $\varepsilon > 0$ that

$$\mathbb{P}(|\bar{X}_n| \geq \varepsilon) \leq \frac{3M_4}{n^2\varepsilon^4}.$$

[10] The conclusion of the theorem is actually true under the same assumptions we had for the weak law. The proof of the general case is however quite demanding, which is why we add the stronger assumption here. Anyone who is interested in the proof of the general case is referred to, for example, "Theorem 22.1" in the book "Probability and Measure" by Patrick Billingsley.

The right side is a summable sequence of non-negative numbers. Hence it follows by the Borel–Cantelli Lemma that

$$\mathbb{P}(|\bar{X}_n| \geq \varepsilon, \text{infinitely often}) = 0.$$

However, we note that by the definition of convergence of sequences, we have

$$(\bar{X}_n \not\to 0) = \bigcup_{\substack{\varepsilon>0 \\ \varepsilon\in\mathbb{Q}}} (|\bar{X}_n| \geq \varepsilon, \text{infinitely often}),$$

and hence it really follows that $\bar{X}_n \to 0$ occurs almost surely. □

3.4 Central Limit Theorem

Definition 3.31. Given a real random variable X, one defines its characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ via $\varphi_X(t) = \mathbb{E}(e^{itX})$.

Theorem 3.32 (Lévy's Inversion Theorem)**.** *Let X be a real random variable with characteristic function φ_X and distribution $\mathbb{P}_X$. Then one has for all $a, b \in \mathbb{R}$ with $a < b$ that*

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\ dt = \frac{1}{2}\mathbb{P}_X(\{a\}) + \mathbb{P}_X\big((a,b)\big) + \frac{1}{2}\mathbb{P}_X(\{b\}).$$

Proof of Theorem 3.32. Note first that due to the l'Hôpital rule, we have $\lim_{t\to 0} \frac{e^{-ita}-e^{-itb}}{t} = i(b-a)$, so the function in the integral is to be understood as a continuous function in this sense. Using Fubini's theorem, we may write

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \underbrace{\varphi_X(t)}_{=\mathbb{E}(e^{itX})}\ dt = \int_{\mathbb{R}} \Big(\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\ dt \Big)\ d\mathbb{P}_X(x).$$

For the purpose of the proof, we define the function $\kappa : \mathbb{R} \to \mathbb{R}$ via $\kappa(x) = \int_0^x \frac{\sin t}{t}\ dt$ for $x \geq 0$, and $\kappa(x) = -\kappa(-x)$ when $x < 0$. We will appeal (without proof) to a fact from calculus stating that $\lim_{x\to\infty} \kappa(x) = \frac{\pi}{2}$. We obviously have the identity $\frac{d}{dx}\kappa(x) = \frac{\sin(x)}{x}$ for all $x \geq 0$. Using the fact that cos is an even function, we hence observe that

$$\begin{aligned}
&\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\ dt \\
=\ &\int_{-T}^{T} \frac{\cos(t(x-a)) - \cos(t(x-b)) + i\sin(t(x-a)) - i\sin(t(x-b))}{it}\ dt \\
=\ &\int_{-T}^{T} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t}\ dt \\
=\ &2\kappa(T(x-a)) - 2\kappa(T(x-b)).
\end{aligned}$$

In particular,

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\ dt = 2 \int_{\mathbb{R}} \kappa(T(x-a)) - \kappa(T(x-b))\ d\mathbb{P}_X(x)$$

In order to determine what happens within the integral when $T \to \infty$, we note that due to the initial remark about κ we obtain

$$\lim_{T\to\infty} \kappa(T(x-a)) - \kappa(T(x-b)) = \begin{cases} 0 & , \quad x < a \text{ or } x > b \\ \pi & , \quad a < x < b \\ \pi/2 & , \quad x = a \text{ or } x = b. \end{cases}$$

Since $\mathbb{P}_X$ is a probability measure, it follows from the Dominated Convergence Theorem that

$$\lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\ dt = \int_{\mathbb{R}} \frac{1}{2}\chi_{\{a,b\}} + \chi_{(a,b)}\ d\mathbb{P}_x,$$

which is precisely the claim. $\square$

We will use the following fact from real analysis without proof:[11]

Theorem 3.33. *A monotone function $f : \mathbb{R} \to \mathbb{R}$ can have at most countably many points of discontinuity.*

Corollary 3.34. *Two real random variables are identically distributed if and only if their characteristic functions are equal.*

Proof. The "only if" part is clear, so we show the "if" part. Let X and Y be two real random variables. The distribution function $F_X(t) = \mathbb{P}_X((-\infty, t])$ is increasing, and therefore by the above theorem, it is discontinuous in at most countably many points. In analogy to Example 2.27, this means that we may have $\mathbb{P}_X(\{t\}) \neq 0$ for at most countably many $t \in \mathbb{R}$. The same observation is of course true for Y in place of X. Let $W \subseteq \mathbb{R}$ be the subset of all points t with $\mathbb{P}_X(\{t\}) = 0 = \mathbb{P}_Y(\{t\})$. Then W is co-countable and in particular dense. It is then an easy exercise to show that the semi-ring

$$\mathfrak{A}_W = \{(a, b] \mid a, b \in W,\ a < b\} \subseteq 2^{\mathbb{R}}$$

generates the Borel σ-algebra. By Theorem 2.17, it follows that $\mathbb{P}_X = \mathbb{P}_Y$ holds if and only if $\mathbb{P}_X$ and $\mathbb{P}_Y$ agree on $\mathfrak{A}_W$. If we assume $\varphi_X = \varphi_Y$, then this is a direct consequence of Theorem 3.32. $\square$

[11] For those interested, see for example "Theorem 4.30" in the book "Principles of Mathematical Analysis" (third edition) by Walter Rudin.

Definition 3.35. A sequence $\mathbb{P}_n$ of Borel probability measures on $\mathbb{R}$ is said to weakly converge to a Borel probability measure $\mathbb{P}$, if one has

$$\lim_{n\to\infty} \int_{\mathbb{R}} f \; d\mathbb{P}_n = \int_{\mathbb{R}} f \; d\mathbb{P}$$

for every compactly supported continuous function $f : \mathbb{R} \to \mathbb{R}$. We write $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

If F_n is the distribution function for $\mathbb{P}_n$ and F the distribution function for $\mathbb{P}$, we say that F_n converges to F weakly, written $F_n \xrightarrow{w} F$, if $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

Definition 3.36. A sequence of real random variables X_n is said to converge in distribution to a real random variable X, written $X_n \xrightarrow{d} X$, if $\mathbb{P}_{X_n} \xrightarrow{w} \mathbb{P}_X$.

Remark. WARNING! Unlike the notions of convergence we considered before, convergence of real random variables in distribution is *not* well-behaved with respect to standard operations such as addition or multiplication. That is, if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$ holds, it is not necessarily true that $X_n + Y_n \xrightarrow{d} X + Y$ or $X_n Y_n \xrightarrow{d} XY$.

We also note that it is shown in the exercises that for a sequence $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, the same kind of limit behavior follows for f being any bounded continuous function on $\mathbb{R}$, not necessarily compactly supported. We will subsequently use this characterization of weak convergence without further mention.

Theorem 3.37. *Let F_n, $n \geq 1$, and F be distribution functions with respect to Borel probability measures $\mathbb{P}_n$ and $\mathbb{P}$. Let $C(F) \subseteq \mathbb{R}$ be the set of all points over which F is continuous. Then $F_n \xrightarrow{w} F$ if and only if $F_n(x) \to F(x)$ for all $x \in C(F)$.*

Proof. Let us first assume $F_n \xrightarrow{w} F$ holds. We shall first show the following intermediate claim. If $A \subseteq \mathbb{R}$ is a closed interval, then $\limsup_{n\to\infty} \mathbb{P}_n(A) \leq \mathbb{P}(A)$. Indeed, we may define a pointwise decreasing sequence of piecewise linear functions $f_k : \mathbb{R} \to \mathbb{R}$ with $f_k|_A = 1$ and $\lim_{k\to\infty} f_k(t) = 0$ for all $t \notin A$. Then

$$\limsup_{n\to\infty} \mathbb{P}_n(A) \leq \limsup_{n\to\infty} \int_{\mathbb{R}} f_k \; d\mathbb{P}_n = \int_{\mathbb{R}} f_k \; d\mathbb{P}, \quad k \geq 1.$$

Here we used the fact that $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$. Letting $k \to \infty$ on the right side, we obtain $\mathbb{P}(A)$ as the limit by the Dominated Convergence Theorem, so the intermediate claim is proved.

Now let $x \in C(F)$. Then it follows from the intermediate claim that

$$\limsup_{n\to\infty} F_n(x) = \limsup_{n\to\infty} \mathbb{P}_n((-\infty, x]) \leq \mathbb{P}((-\infty, x]) = F(x).$$

Since x is a continuous point of F, we also have $\mathbb{P}(\{x\}) = 0$ and hence

$$\begin{aligned} F(x) = \mathbb{P}((-\infty, x)) &= 1 - \mathbb{P}([x, \infty)) \\ &\leq 1 - \limsup_{n\to\infty} \mathbb{P}_n([x, \infty)) \\ &= \liminf_{n\to\infty} \mathbb{P}_n((-\infty, x)) \\ &\leq \liminf_{n\to\infty} F_n(x). \end{aligned}$$

Conversely, let us assume that $F_n(x) \to F(x)$ holds for all $x \in C(F)$. By definition, we may easily observe $\mathbb{P}_n((a, b]) \to \mathbb{P}((a, b])$ for all $a, b \in C(F)$ with $a < b$. Therefore, it follows that

$$\int_{\mathbb{R}} f \; d\mathbb{P}_n \to \int_{\mathbb{R}} f \; d\mathbb{P}$$

holds whenever f belongs to the linear subspace spanned by all indicator functions $\chi_{(a,b]}$ for $a, b \in C(F)$. Since $C(F)$ is dense in $\mathbb{R}$, it is easy to see that the closure of this linear subspace with respect to the sup-norm $\| \cdot \|_\infty$ contains the space of continuous compactly supported functions. So (with a standard $\varepsilon/2$-argument) we may conclude that the above limit behavior even holds for f being any compactly supported continuous function. □

Definition 3.38. A sequence of Borel probability measures $(\mathbb{P}_n)_n$ on $\mathbb{R}$ is called tight, if

$$\lim_{R\to\infty} \liminf_{n\to\infty} \; \mathbb{P}_n([-R, R]) = 1.$$

Remark. In the exercise sessions, we will prove the following analytical statement which is useful for the proof of the next theorem. Let $F_n : \mathbb{R} \to [0, 1]$ be a sequence of increasing right-continuous functions satisfying the limit formula $\lim_{R\to\infty} F_n(R) - F_n(-R) = 1$.

- If the sequence $F_n(x)$ is convergent for a dense set of numbers $x \in \mathbb{R}$, then there exists an increasing right-continuous function $F : \mathbb{R} \to [0, 1]$ such that $F_n(x) \to F(x)$ holds whenever F is continuous in x.

- Suppose that a limit function F as above exists and that

  $$\lim_{R\to\infty} \liminf_{n\to\infty} F_n(R) - F_n(-R) = 1.$$

  Then F is the distribution function for a Borel probability measure on $\mathbb{R}$.

Theorem 3.39 (Helly's Selection Theorem)**.** *For any tight sequence $(\mathbb{P}_n)_n$ of Borel probability measures on $\mathbb{R}$, there exists a subsequence $(\mathbb{P}_{n_k})_k$ and a Borel probability measure $\mathbb{P}$ on $\mathbb{R}$ such that $P_{n_k} \xrightarrow{w} \mathbb{P}$.*

Proof. Let F_n be the distribution function of $\mathbb{P}_n$ for every $n \geq 1$. Then $\mathbb{P}_n$ being tight translates to

$$\lim_{R\to\infty} \liminf_{n\to\infty} F_n(R) - F_n(-R) = 1.$$

Of course this tightness criterion holds for every subsequence as well. By Theorem 3.37 and the above remark, it suffices to show that there is some increasing sequence of natural numbers n_k such that F_{n_k} converges pointwise on a dense set of real numbers, for example on the rational numbers $\mathbb{Q}$. Let $\mathbb{N} \ni \ell \mapsto q_\ell$ be an enumeration of $\mathbb{Q}$. By Bolzano–Weierstrass, we know that there is some increasing sequence of numbers $n_{(1,k)}$ such that $F_{n(1,k)}(q_1)$ converges as $k \to \infty$. Applying Bolzano–Weierstrass again, we know that $(n(1,k))_k$ admits a subsequence $(n(2,k))_k$ such that $F_{n(2,k)}(q_2)$ converges as $k \to \infty$. Proceed like this inductively, and find finer and finer subsequences $(n(\ell,k))_k$ such that $F_{n(\ell,k)}(q_\ell)$ converges as $k \to \infty$. Finally define $n_k = n(k,k)$. Then $(n_k)_k$ is a subsequence of $(n(\ell,k))_k$ (up to the finitely many indices $k \leq \ell$) for every $\ell \geq 1$, so indeed $F_{n_k}(q_\ell)$ converges as $k \to \infty$, for every $\ell \geq 1$. This finishes the proof. $\square$

Lemma 3.40. *Let X be a real random variable. Then we have for all $R > 0$ the estimate*

$$\mathbb{P}(|X| > 2R) \leq R \int_{-1/R}^{1/R} 1 - \varphi_X(t)\ dt.$$

Proof. Using Fubini's theorem we compute

$$\begin{aligned}
R \int_{-1/R}^{1/R} 1 - \varphi_X(t)\ dt &= 2 - R \int_{-1/R}^{1/R} \varphi_X(t)\ dt \\
&= 2 - R \int_{-1/R}^{1/R} \int_{\mathbb{R}} e^{itx}\ d\mathbb{P}_X(x)\ dt \\
&= 2 - R \int_{\mathbb{R}} \int_{-1/R}^{1/R} \cos(tx)\ dt\ d\mathbb{P}_X(x) \\
&= 2 - 2R \int_{\mathbb{R}} \int_{0}^{1/R} \cos(tx)\ dt\ d\mathbb{P}_X(x) \\
&= 2 - 2R \int_{\mathbb{R}} \frac{\sin(x/R)}{x}\ d\mathbb{P}_X(x) \\
&= 2\mathbb{E}\Big(1 - \frac{\sin(X/R)}{X/R}\Big)
\end{aligned}$$

Note that we have $\frac{\sin(X/R)}{X/R} \leq 1$. Moreover, if any $x \in \mathbb{R}$ satisfies $|x| > 2R$, then $|x/R| > 2$, in which case $|\sin(x/R)| \leq 1 \leq |x|/2R$, hence $1 - \frac{\sin(x/R)}{x/R} \geq \frac{1}{2}$. This leads to the inequality $\chi_{(|X|>2R)} \leq 2(1 - \frac{\sin(X/R)}{X/R})$, and hence forming the expected value on both sides yields the claim. $\square$

Theorem 3.41 (Lévy's Continuity Theorem)**.** *Let X_n and X be real random variables for $n \geq 1$. Then $X_n \xrightarrow{d} X$ holds if and only if $\varphi_{X_n}(t) \to \varphi_X(t)$ for all $t \in \mathbb{R}$.*

Proof. The direction "⇒" holds by definition of the characteristic functions and by what it means to converge in distribution. For the converse, let us assume $\varphi_{X_n}(t) \to \varphi_X(t)$ for all $t \in \mathbb{R}$. Let us first show that the sequence $\mathbb{P}_{X_n}$ is tight. Indeed, using the above lemma we observe for $R > 0$ that

$$\begin{array}{rcl} \mathbb{P}_{X_n}([-R,R]) & = & 1 - \mathbb{P}_{X_n}(|X_n| > R) \\ & \geq & 1 - \dfrac{R}{2} \displaystyle\int_{-2/R}^{2/R} 1 - \varphi_{X_n}(t)\ dt \\ & \overset{n\to\infty}{\longrightarrow} & 1 - 2 \cdot \dfrac{R}{4} \displaystyle\int_{-2/R}^{2/R} 1 - \varphi_X(t)\ dt. \end{array}$$

Here we also used the Dominated Convergence Theorem. Using the fact that φ_X is a continuous function with $\varphi_X(0) = 1$, we see that the integral $\frac{R}{4}\int_{-2/R}^{2/R} 1 - \varphi_X(t)\ dt$ goes to zero as $R \to \infty$. Hence the above inequality yields

$$\lim_{R\to\infty} \liminf_{n\to\infty} \mathbb{P}_{X_n}([-R,R]) = 1,$$

or in other words the tightness of the sequence $\mathbb{P}_{X_n}$.

In order to show $X_n \xrightarrow{d} X$ we proceed by contradiction, and suppose that X_n does not converge to X in distribution. This means that for some compactly supported continuous function $f : \mathbb{R} \to \mathbb{R}$, one has that $\int_{\mathbb{R}} f\ d\mathbb{P}_{X_n}$ does not converge to $\int_{\mathbb{R}} f\ d\mathbb{P}_X$. Thus, after passing to a subsequence, we may assume without loss of generality that

$$\liminf_{n\to\infty} \Big| \int_{\mathbb{R}} f\ d\mathbb{P}_{X_n} - \int_{\mathbb{R}} f\ d\mathbb{P}_X \Big| > 0.$$

By Theorem 3.39, we may again pass to a subsequence and assume without loss of generality that $X_n \xrightarrow{d} Y$ for some real random variable Y. By the "⇒" part, we obtain $\varphi_X(t) = \lim_{n\to\infty} \varphi_{X_n}(t) = \varphi_Y(t)$ for all $t \in \mathbb{R}$. By Corollary 3.34, it follows that $\mathbb{P}_X = \mathbb{P}_Y$, but this leads to a contradiction to what we assumed above. We conclude that $X_n \xrightarrow{d} X$ must hold in the first place. □

Lemma 3.42. *For all $x \in \mathbb{R}$ and $n \geq 0$, we have the inequality*

$$\Big| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \Big| \leq \min\Big(\frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!} \Big).$$

Proof. We proceed by induction in n. We note that it suffices to consider the case $x \geq 0$ since we may obtain the claim for $x < 0$ by complex conjugation of the involved terms. Let us first proceed with proving that $\frac{2|x|^n}{n!}$ is an upper bound. For $n = 0$, this is equal to 2, and since e^{ix} has norm one, this follows from the triangle inequality. Now assume that this inequality holds for a given number $n \geq 0$ and all $x \geq 0$. We then compute

$$\begin{aligned} \left| e^{ix} - \sum_{k=0}^{n+1} \frac{(ix)^k}{k!} \right| &= \left| \int_0^x e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \, dt \right| \\ &\leq \int_0^x \left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| dt \\ &\leq \int_0^x \frac{2t^n}{n!} \, dt = \frac{2x^{n+1}}{(n+1)!}. \end{aligned}$$

We can also proceed by induction to show the upper bound $\frac{|x|^{n+1}}{(n+1)!}$. Then one can first consider $n = -1$, in which case both the left and right side is 1. One can then perform the induction step $(n-1) \to n$ with a completely analogous calculation as above, namely

$$\begin{aligned} \left| e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!} \right| &= \left| \int_0^x e^{it} - \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \, dt \right| \\ &\leq \int_0^x \left| e^{it} - \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \right| dt \\ &\leq \int_0^x \frac{t^n}{n!} \, dt = \frac{x^{n+1}}{(n+1)!}. \end{aligned}$$

□

Lemma 3.43. *Let X be a real random variable such that $\mathbb{E}(X^2) = 1$ and $\mathbb{E}(X) = 0$. Then one has the limit formula*

$$\lim_{t \to \infty} \varphi_X(t/\sqrt{n})^n = e^{-t^2/2}, \quad t \in \mathbb{R}.$$

Proof. As an application of Lemma 3.42, we obtain for all $t \in \mathbb{R}$ the estimate

$$\left| e^{itX} - \left(1 + itX - \frac{t^2 X^2}{2}\right) \right| \leq \min\left(t^2 X^2, \frac{t^3 |X^3|}{6}\right) = t^2 \min\left(X^2, \frac{t|X^3|}{6}\right).$$

In particular, since X has a second absolute moment we see that the random variable on the left has a mean. Moreover the expected value of the random variable $X_0^t = \min\left(X^2, \frac{t|X^3|}{6}\right)$ tends to zero as $t \to 0$ as a consequence of the Dominated Convergence Theorem. So by integrating and applying the triangle inequality, we see

$$\Big|\varphi_X(t) - 1 + \frac{t^2}{2}\Big| \leq \mathbb{E}\Big(\Big|e^{itX} - 1 - itX + \frac{t^2X^2}{2}\Big|\Big) \leq t^2\mathbb{E}(X_0^t).$$

If we substitute $t \to t/\sqrt{n}$, this leads to the observation that

$$x_n = n(\varphi_X(t/\sqrt{n}) - \Big(1 - \frac{t^2}{2n}\Big)) \overset{n\to\infty}{\longrightarrow} 0.$$

Using the fact from calculus that one has convergence

$$e^x = \lim_{n\to\infty}\Big(1 + \frac{x}{n}\Big)^n$$

uniformly over bounded intervals, we finally conclude that

$$\begin{aligned}
\varphi_X(t/\sqrt{n})^n &= \Big(1 + (\varphi_X(t/\sqrt{n}) - 1)\Big)^n \\
&= \Big(1 - \frac{\frac{t^2}{2} - n(\varphi_X(t/\sqrt{n}) - 1 + \frac{t^2}{2n})}{n}\Big)^n \\
&= \Big(1 - \frac{\frac{t^2}{2} - x_n}{n}\Big)^n \\
&\overset{n\to\infty}{\longrightarrow} e^{-t^2/2}.
\end{aligned}$$

□

Theorem 3.44 (Central Limit Theorem)**.** *Let X_n be an independent sequence of identically distributed real random variables with mean $\mathbb{E}(X_n) = 0$ and variance $\mathbb{E}(X_n^2) = 1$. Consider the standard normal variable $\mathcal{N}$ given by $\mathbb{P}_{\mathcal{N}}((a,b]) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}\,dt$. Then one has*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \xrightarrow{d} \mathcal{N}.$$

Proof. Since the X_n are identically distributed, they all have the same characteristic function, which we shall denote by φ_X. Denote $Y_n = n^{-1/2}\sum_{k=1}^n X_k$, so that we wish to show $Y_n \xrightarrow{d} \mathcal{N}$. Using that the X_n are independent, we observe for all $t \in \mathbb{R}$ that

$$\begin{aligned}
\varphi_{Y_n}(t) &= \mathbb{E}(e^{-itY_n}) \\
&= \mathbb{E}(e^{-i\frac{t}{\sqrt{n}}\sum_{k=1}^n X_k}) \\
&= \varphi_{\sum_{k=1}^n X_k}(t/\sqrt{n}) \\
&= \prod_{k=1}^n \varphi_{X_k}(t/\sqrt{n}) \\
&= \varphi_X(t/\sqrt{n})^n
\end{aligned}$$

Given that the standard normal variable has the characteristic function $\varphi_{\mathcal{N}}(t) = e^{-t^2/2}$, the proof is complete by combining Lemma 3.43 and Theorem 3.41. $\square$

Corollary 3.45. *Let X_n be an independent sequence of identically distributed real random variables with a finite mean $\mathbb{E}(X_n) = 0$ and finite variance $\mathbb{E}(X_n^2) = \sigma^2$. Then one has*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \xrightarrow{d} \mathcal{N}(0, \sigma),$$

where $\mathcal{N}(0, \sigma)$ is the real random variable given by the distribution $\mathbb{P}((a, b]) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2\sigma^2}}\, dt$.

4 Selected further topics

In this final chapter we treat a few selected further topics, both from probability theory and general measure theory.

4.1 Tail events

Definition 4.1. Let $(\Omega, \mathfrak{M})$ be a measurable space and $A_n \in \mathfrak{M}$ a sequence of measurable sets. For each $n \geq 1$, denote by $\mathfrak{M}_n \subseteq \mathfrak{M}$ the σ-algebra generated by $\{A_k\}_{k \geq n}$. The tail σ-algebra for the sequence A_n is defined as

$$\mathfrak{M}_\infty = \bigcap_{n \geq 1} \mathfrak{M}_n.$$

Example 4.2. As an important example notice in the above definition that one always has that $\limsup\limits_{n\to\infty} A_n$ and $\liminf\limits_{n\to\infty} A_n$ belong to the tail σ-algebra.

Lemma 4.3. *Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space. Let $A \in \mathfrak{M}$, $\{B_i\}_{i\in I} \subseteq \mathfrak{M}$, and $\{C_j\}_{j\in J} \subseteq \mathfrak{M}$ be given.*

(i) *Suppose $\{A\} \cup \{B_i\}_{i\in I}$ is an independent family. Then for every set B belonging to the σ-algebra generated by $\{B_i\}$, one has that A and B are independent.*

(ii) *Suppose that $\{B_i\}_{i\in I} \cup \{C_j\}_{j\in J}$ is an independent family. Let $\mathfrak{B}$ be the σ-algebra generated by $\{B_i\}$ and $\mathfrak{C}$ the σ-algebra generated by $\{C_j\}$. Then $\mathfrak{B}$ and $\mathfrak{C}$ are independent sub-σ-algebras of $\mathfrak{M}$.*

Proof. (i): We denote by $\mathfrak{B}$ the σ algebra generated by $\{B_i\}$. If A is a null set, there is nothing to show, so let us assume $\mathbb{P}(A) > 0$. Consider a new measure $\mathbb{P}_1$ on $(\Omega, \mathfrak{M})$ defined by $\mathbb{P}_1(C) = \frac{\mathbb{P}(A\cap C)}{\mathbb{P}(A)}$.

Let $\mathfrak{A} \subseteq \mathfrak{M}$ be the semi-ring defined by

$$\mathfrak{A} = \left\{ \bigcap_{\ell=1}^{m} D_\ell \mid m \geq 0,\ i_1, \dots, i_m \in I \text{ are distinct and } D_\ell \in \left\{B_{i_\ell}, B_{i_\ell}^c\right\} \right\}.$$

Note that we use the convention that this intersection is defined to be empty when $m = 0$. We leave the details why $\mathfrak{A}$ is a semi-ring as an exercise. Then $\mathfrak{B}$ is also generated by $\mathfrak{A}$ as a σ-algebra. By the independence assumption, it follows that $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $B \in \mathfrak{A}$, and hence $\mathbb{P}_1(B) = \mathbb{P}(B)$. By Theorem 2.17, we see that this implies $\mathbb{P}_1(B) = \mathbb{P}(B)$ for all $B \in \mathfrak{B}$, or in other words $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $B \in \mathfrak{B}$.

(ii): We have to show that for every $B \in \mathfrak{B}$ and $C \in \mathfrak{C}$, it follows that B and C are independent. First of all, it follows from the definition of independence that for any set B belonging to the above defined semi-ring $\mathfrak{A}$, we have that the family $\{B\} \cup \{C_j\}_{j\in J}$ is independent. By (i), it hence follows that for every such $B \in \mathfrak{A}$ and $C \in \mathfrak{C}$ are mutually independent. By the definition of $\mathfrak{A}$ as a set and the definition of independence, we see that hence $\{B_i\}_{i\in I} \cup \{C\}$ is an independent family for all $C \in \mathfrak{C}$. Applying (i) one more time, we can conclude that all $B \in \mathfrak{B}$ and $C \in \mathfrak{C}$ are mutually independent. $\square$

Theorem 4.4 (Kolmogorov's zero-one law)**.** *Let $(\Omega, \mathfrak{M}, \mathbb{P})$ be a probability space and $A_n \in \mathfrak{M}$ an independent sequence of events. Then for any events $A \in \mathfrak{M}_\infty$, one has $\mathbb{P}(A) \in \{0, 1\}$.*

Proof. Suppose $A \in \mathfrak{M}_\infty$. Let $n \geq 1$. Then $A \in \mathfrak{M}_{n+1}$, the σ-algebra generated by $\{A_k\}_{k>n}$. Since the sequence $\{A_k\}_{k\geq 1}$ is independent by assumption, it follows from the above lemma that $\{A\} \cup \{A_k\}_{k\leq n}$ is an independent family. However, as n was arbitrary, it follows that in fact $\{A\} \cup \{A_k\}_{k\geq 1}$ is an independent family. Appealing to the lemma above again, it follows that A and B are independent for every $B \in \mathfrak{M}_1$. Since $\mathfrak{M}_\infty \subseteq \mathfrak{M}_1$, it follows that in particular A is independent of itself. This leads to $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, which implies $\mathbb{P}(A) \in \{0, 1\}$. $\square$

4.2 Radon–Nikodym theorem

Definition 4.5. Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. In analogy to Definition 1.40, we consider

$$\mathcal{L}^2(\Omega, \mathfrak{M}, \mu) = \mathcal{L}^2(\mu) = \Big\{ f : \Omega \to \mathbb{C} \mid f \text{ is measurable and } |f|^2 \text{ is integrable} \Big\}.$$

For $f \in \mathcal{L}^2(\mu)$, one defines the 2-seminorm

$$\|f\|_2 = \Big(\int_\Omega |f|^2 \, d\mu \Big)^{1/2}.$$

Theorem 4.6. *Let $(\Omega, \mathfrak{M}, \mu)$ be a measure space. Then the formula*

$$\langle f \mid g \rangle = \int_\Omega f\bar{g} \, d\mu$$

defines a sesquilinear anti-symmetric positive semi-definite form on $\mathcal{L}^2(\mu)$. If one considers the linear subspace

$$N = \{ f : \Omega \to \mathbb{C} \mid f \textit{ is measurable and } f = 0 \textit{ almost everywhere} \},$$

then $L^2(\mu) = \mathcal{L}^2(\Omega, \mathfrak{M}, \mu)/N$ becomes a Hilbert space with respect the induced inner product and the resulting 2-norm.

Proof. We can conclude from the general Hölder inequality with $p = q = 2$ (see also the version proved in the exercises) that $\langle\cdot|\cdot\rangle$ is well-defined. Its claimed properties are immediate. In particular, it follows from results in basic linear algebra that $\|\cdot\|_2$ is indeed a seminorm. Like before (cf. Proposition 1.45), the subspace N can be identified with those functions $f \in \mathcal{L}^2(\mu)$ with $\|f\|_2 = 0 = \langle f \mid f\rangle$. Therefore the induced sesquilinear form $\langle\cdot|\cdot\rangle$ on $L^2(\mu)$ defines an inner product, so that $\|\cdot\|_2$ becomes a norm in analogy to Definition 1.46. The only thing that remains to be shown is that $L^2(\mu)$ is complete. For this it suffices to prove a statement analogous to Theorem 1.52 (in fact with an analogous proof).

Let $f_n \in \mathcal{L}^2(\mu)$ be a sequence that satisfies the Cauchy criterion in the 2-norm. By applying the Cauchy criterion inductively, we can find a subsequence $(f_{n_k})_k$ such that $\|f_{n_k} - f_{n_{k+1}}\|_2 \leq 2^{-k}$. We define the measurable function $g : \Omega \to [0, \infty]$ as $g(x) = \sum_{k=1}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)|$. Then it follows from the Monotone Convergence Theorem and the triangle inequality for the 2-(semi-)norm that

$$\begin{aligned}
\int_\Omega g^2 \, d\mu &= \lim_{n\to\infty} \int_\Omega \Big(\sum_{k=1}^{n} |f_{n_k} - f_{n_{k+1}}| \Big)^2 \, d\mu \\
&= \lim_{n\to\infty} \Big\| \sum_{k=1}^{n} |f_{n_k} - f_{n_{k+1}}| \Big\|_2^2 \\
&\leq \lim_{n\to\infty} \Big(\sum_{k=1}^{n} \|f_{n_k} - f_{n_{k+1}}\|_2 \Big)^2 \\
&\leq \Big(\sum_{k=1}^{\infty} 2^{-k} \Big)^2 \;=\; 1.
\end{aligned}$$

In particular g^2 is integrable, and by Proposition 1.38, the set $E = g^{-1}([0, \infty))$ has a complement of zero measure. For all $x \in E$, we have by definition that the series $\displaystyle\sum_{k=1}^{\infty} [f_{n_{k+1}}(x) - f_{n_k}(x)]$ converges absolutely, and hence the function

$$f(x) := f_{n_1}(x) + \sum_{k=1}^{\infty} [f_{n_{k+1}}(x) - f_{n_k}(x)] = \lim_{k\to\infty} f_{n_k}(x), \quad x \in E$$

is well-defined and measurable on E. We extend f to a measurable function on Ω by defining it to be zero on the complement of E. We get by the triangle inequality that for all $k \geq 1$, the function f_{n_k} is dominated (on E) by the square-integrable function $|f_{n_1}| + g$. Therefore it follows from the Dominated Convergence Theorem that $\|f - f_{n_k}\|_2 \overset{k\to\infty}{\longrightarrow} 0$. Since the sequence $(f_n)_n$ was assumed to satisfy the Cauchy criterion in $\|\cdot\|_2$ and we just showed that a subsequence converges to f in this norm, it follows that also $\|f - f_n\|_2 \to 0$. (As before this is an easy $\varepsilon/2$-argument.) This finishes the proof. □

Definition 4.7. Let $(\Omega, \mathfrak{M})$ be a measurable space. Given two measures μ, ν on $(\Omega, \mathfrak{M})$, we call ν absolutely continuous with respect to μ, if for all $E \in \mathfrak{M}$, one has that $\mu(E) = 0$ implies $\nu(E) = 0$.

Lemma 4.8. *Let $(\Omega, \mathfrak{M})$ be a measurable space. Let μ be an arbitrary measure and ν a finite measure. Then ν is absolutely continuous with respect to μ if and only if the following condition holds: For every $\varepsilon > 0$, there exists $\delta > 0$, such that for every $E \in \mathfrak{M}$ with $\mu(E) < \delta$, one has $\nu(E) < \varepsilon$.*[12]

Proof. The "if" part is clear, so we shall prove the "only if" part and assume that ν is absolutely continuous with respect to μ. Suppose that the claimed condition were to fail. This means that there exists some $\varepsilon_0 > 0$ for which no $\delta > 0$ satisfies the required property. Let us then choose for every $k \geq 1$ a set $E_k \in \mathfrak{M}$ so that $\nu(E_n) \geq \varepsilon_0$ but $\mu(E_k) < 2^{-k}$. Then the set $A = \limsup_{k\to\infty} E_k$ satisfies $\mu(A) = 0$ by the (first) Borel–Cantelli lemma. So by absolute continuity, we must have $\nu(A) = 0$. However, since ν is a finite measure, we conclude with Proposition 1.20(v) that

$$\nu(A) = \nu\Big(\bigcap_{n\geq 1}\bigcup_{k\geq n} E_k\Big) = \inf_{n\geq 1} \nu\Big(\bigcup_{k\geq n} E_k\Big) \geq \varepsilon_0.$$

This is a contradiction and hence the claimed ε-δ-property must hold. $\square$

We will appeal to the following theorem without proof. It is covered in the course on functional analysis.

Theorem 4.9 (Riesz–Fischer)**.** *Let $\mathcal{H}$ be a Hilbert space. Let $\theta : \mathcal{H} \to \mathbb{C}$ be a bounded linear functional, i.e., a $\mathbb{C}$-linear map such that there exists some constant $C > 0$ such that $\|\theta(x)\| \leq C\|x\|$ for all $x \in \mathcal{H}$. Then there exists an element $y \in \mathcal{H}$ such that $\theta(x) = \langle x \mid y\rangle$ for all $x \in \mathcal{H}$.*

Theorem 4.10 (Radon–Nikodym Theorem)**.** *Let $(\Omega, \mathfrak{M})$ be a measurable space. Let ν and μ both be σ-finite measures on $(\Omega, \mathfrak{M})$. Then ν is absolutely continuous with respect to μ if and only if there exists a non-negative $\mathfrak{M}$-measurable function $h : \Omega \to [0, \infty)$ such that $\nu(E) = \int_E h \, d\mu$ for all $E \in \mathfrak{M}$. In this context, h is called the Radon–Nikodym derivative and is often denoted $h = \frac{d\nu}{d\mu}$. It is uniquely determined up to equality μ-almost everywhere.*

Proof. The "if" part is easy. If such a function h exists, then it follows immediately that ν must be absolutely continuous with respect to μ.

So we need to show the "only if" part and assume that ν is absolutely continuous with respect to μ. We first claim that it is enough to consider

[12] Notice that that this is equivalent to saying that for any sequence $E_n \in \mathfrak{M}$, the condition $\mu(E_n) \to 0$ implies $\nu(E_n) \to 0$.

the case where the involved measures are finite. Indeed, using that ν and μ are σ-finite, we may find a sequence $A_k \in \mathfrak{M}$ with $\mu(A_k), \nu(A_k) < \infty$ and $\Omega = \bigcup_{k\geq 1} A_k$. If we know that the claim is true for finite measures, we can employ it to the restrictions of the measures on $A_k \setminus A_{k-1}$ for all $k \geq 1$ (where $A_0 := \emptyset$) and define find a function h_k on $A_k \setminus A_{k-1}$ with the required property for $E \in \mathfrak{M}$ with $E \subseteq A_k \setminus A_{k-1}$. If extend these functions by zero, we may obtain a non-negative measurable function $h = \sum_{k=1}^{\infty} h_k : \Omega \to [0, \infty)$ satisfying the general formula $\nu(E) = \int_E h \, d\mu$ for all $E \in \mathfrak{M}$ by virtue of the Monotone Convergence Theorem.

So let us from now on assume that both ν and μ are finite measures. Set $\lambda = \nu + \mu$. By the Hölder inequality, it follows that $\mathcal{L}^2(\lambda) \subseteq \mathcal{L}^1(\lambda)$ and $\|f\|_1 \leq \|f\|_2 \sqrt{\lambda(\Omega)}$ for all $f \in \mathcal{L}^2(\lambda)$. Since $\nu \leq \lambda$ by definition, it follows that we can view the assignment

$$L^2(\lambda) \to \mathbb{C}, \quad [f] \mapsto \int_\Omega f \, d\nu$$

as a bounded linear function on a Hilbert space. By the Riesz–Fischer theorem, we therefore find a function $g \in \mathcal{L}^2(\lambda)$ with the property that

$$\int_\Omega f \, d\nu = \int_\Omega fg \, d\lambda.$$

By definition of λ, this leads to the equivalent formula

$$\int_\Omega f(1-g) \, d\nu = \int_\Omega fg \, d\mu.$$

We note also that for all $E \in \mathfrak{M}$ with $\lambda(E) > 0$, we have in particular

$$0 \leq \nu(E) = \int_\Omega \chi_E \, d\nu = \int_\Omega g\chi_E \, d\lambda \leq \lambda(E),$$

so we are in a situation to apply Proposition 1.44 and conclude that $g(x) \in [0, 1]$ holds λ-almost everywhere. Note that by absolute continuity, μ and λ have the same null sets, so this holds in fact μ-almost everywhere. We claim in addition that $\mu(g^{-1}(1)) = 0$. Indeed, for $E = g^{-1}(1) \in \mathfrak{M}$, its characteristic function fits into the equation

$$0 = \int_\Omega \chi_E(1-g) \, d\nu = \int_\Omega \chi_E g \, d\mu = \mu(E).$$

After redefining g on a measurable null set with respect to μ (and hence also with respect to ν, by absolute continuity), we may assume that g takes values in $[0, 1)$. In particular, it makes sense to define $h = \sum_{n=1}^{\infty} g^n$ as a measurable non-negative function on Ω. We claim that this is the desired function. Indeed, we observe due to the equations above that for any $E \in \mathfrak{M}$, we get for every $n \geq 1$ that

$$\int_\Omega \chi_E (1 - g^{n+1})\ d\nu = \int_\Omega \chi_E (1 - g) \sum_{k=0}^{n} g^k\ d\nu = \int_\Omega \chi_E g \sum_{k=0}^{n} g^k\ d\mu.$$

If we let $n \to \infty$, then it follows from the Monotone Convergence Theorem that the left side converges to $\nu(E)$, whereas the right side converges to $\int_E h\ d\mu$. This finishes the proof. $\square$