Probability & Measure

Università della Svizzera italiana

Fabian Bosshard

June 30, 2026

Contents

Preface ... **ii**

References ... **ii**

0 Useful Notions ... **1**

1 Probability Spaces ... **4**
- 1.1 Measurable Spaces ... 4
- 1.2 Measures ... 8

2 Lebesgue measure ... **13**
- 2.1 Borel sets ... 13
- 2.2 Construction of the Lebesgue-Borel measure ... 17
- 2.3 Null sets ... 26

3 Random variables ... **28**
- 3.1 Measurable functions ... 28
- 3.2 Construction of measurable functions ... 29

4 Independence ... **35**
- 4.1 Independent Random Variables ... 35
- 4.2 Borel-Cantelli lemmas ... 42

5 The integral ... **48**
- 5.1 Construction of the integral ... 48
- 5.2 Properties of the Integral ... 55
- 5.3 L^p-spaces ... 60
- 5.4 Convergence theorems ... 61

6 Laws of Large Numbers ... **64**
- 6.1 Weak Law of Large Numbers ... 64
- 6.2 Strong Law of Large Numbers ... 66

7 Central Limit Theorem ... **68**

Preface

This document contains unofficial student-made notes for the course Probability & Measure taught by Michael Multerer with the assistance of Jacopo Quizi in Spring 2026 at the Università della Svizzera italiana. These notes are mostly based on the course materials, especially [1], but they also include additional explanations, examples, and intermediate steps to aid understanding. The textbooks listed for the course are [2, 3]. The counterexample in Example 3.6 is adapted from [4]. The visualization in Definition 2.3 is adapted from [5]; the sketches in Proposition 2.3 are inspired by [6, 7]; and the monkey illustration in Example 4.7 is adapted from [8]. If you spot an error, please report it to `fabianlucasbosshard@gmail.com`. The LaTeX source is available at `https://github.com/fabianbosshard/usi-informatics-course-summaries`.

References

[1] Michael Multerer. *Probability & Measure*. Università della Svizzera italiana, 2026. URL: `https://muchip.github.io/resources/PM_notes.pdf`.

[2] Paul R. Halmos. *Measure Theory*. 1st ed. New York: Springer-Verlag, 1974. URL: `https://reperio.usi.ch/permalink/41SLSP_USI/17vpo4o/alma991170317226105501`.

[3] Kenneth Shum. *Measure-Theoretic Probability: With Applications to Statistics, Finance, and Engineering*. Cham, Switzerland: Birkhäuser, 2023. URL: `https://reperio.usi.ch/permalink/41SLSP_USI/17vpo4o/alma991171870308905501`.

[4] *A Counterexample Showing That Separate Measurability Need Not Imply Joint Measurability*. Math Stack Exchange. URL: `https://math.stackexchange.com/a/647242/1692832` (visited on 06/26/2026).

[5] user121799. *A TikZ Visualization of a Preimage*. TeX Stack Exchange. URL: `https://tex.stackexchange.com/a/457595` (visited on 06/26/2026).

[6] user121799. *An Epsilon-Delta Diagram in TikZ*. TeX Stack Exchange. URL: `https://tex.stackexchange.com/a/485692` (visited on 06/26/2026).

[7] Evion and Stack Exchange Community. *An Epsilon-Delta Plot in TikZ*. TeX Stack Exchange. URL: `https://tex.stackexchange.com/a/563373` (visited on 06/26/2026).

[8] Mark Wibrow and Stack Exchange Community. *A Monkey Icon in TikZ*. TeX Stack Exchange. URL: `https://tex.stackexchange.com/a/414059` (visited on 06/26/2026).

0 Useful Notions

Contingency table: displays the multivariate frequency distribution of the variables.

Conditional Probability and Bayes' Theorem

Events

Conditional probability

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$$

Bayes' theorem

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\,\mathrm{P}(A)}{\mathrm{P}(B)}$$

Total probability

Let $(A_i)_i$ be a partition of Ω. Then

$$\begin{aligned}\mathrm{P}(B) &= \sum_i \mathrm{P}(B \cap A_i) \\ &= \sum_i \mathrm{P}(B \mid A_i)\,\mathrm{P}(A_i).\end{aligned}$$

From Events to Random Variables

If X, Y are random variables, the same structure appears at the level of distributions.

Random Variables

Discrete case (pmf)

Joint distribution
$p_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y)$

Marginal distribution
$p_Y(y) = \sum_x p_{X,Y}(x,y)$

Conditional distribution
$p_{X|Y}(x \mid y) = \dfrac{p_{X,Y}(x,y)}{p_Y(y)}$

Bayes' theorem
$p_{X|Y}(x \mid y) = \dfrac{p_{Y|X}(y \mid x)\, p_X(x)}{\sum_u p_{Y|X}(y \mid u)\, p_X(u)}$

Continuous case (pdf)

Joint distribution
$f_{X,Y}(x,y)$

Marginal distribution
$f_Y(y) = \int f_{X,Y}(x,y)\,\mathrm{d}x$

Conditional distribution
$f_{X|Y}(x \mid y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

Bayes' theorem
$f_{X|Y}(x \mid y) = \dfrac{f_{Y|X}(y \mid x)\, f_X(x)}{\int f_{Y|X}(y \mid u)\, f_X(u)\,\mathrm{d}u}$

Rational Numbers

For any $a, b \in \mathbb{R}$ with $a < b$

$$\bigcup_{\substack{a<q_1<q_2<b \\ q_1,q_2 \in \mathbb{Q}}} (q_1, q_2) = (a, b)$$

We have $\overline{\mathbb{Q}} = \mathbb{R}$.

Telescopic Identity

Let $(a_n)_{n \in \mathbb{N}_0}$ be a sequence. Then, for every $N \in \mathbb{N}$, there holds

$$\sum_{n=1}^{N} (a_n - a_{n-1}) = a_N - a_0 \qquad \Longleftrightarrow \qquad a_N = a_0 + \sum_{n=1}^{N} (a_n - a_{n-1})$$

If $a_n \to a$, then letting $N \to \infty$ yields

$$a = a_0 + \sum_{n=1}^{\infty} (a_n - a_{n-1}) \tag{0.1}$$

Open rectangles are open

To check that a rectangle is open, it suffices to show that every point inside it admits a small ball still contained in the rectangle, see Remark 2.3.

Consider the rectangle

$$Q = (\boldsymbol{a}, \boldsymbol{b}) = \bigtimes_{i=1}^{d} (a_i, b_i)$$

where $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ with $\boldsymbol{a} < \boldsymbol{b}$. Given $\boldsymbol{x} \in Q$, choose $r(\boldsymbol{x}) > 0$ such that

$$0 < r(\boldsymbol{x}) < \min_{1 \leq i \leq d} \{x_i - a_i,\, b_i - x_i\}$$

which is possible because $\boldsymbol{x} \in Q$ implies $a_i < x_i < b_i$ for every $i = 1, \ldots, d$. Now consider $\boldsymbol{z} \in B_{r(\boldsymbol{x})}(\boldsymbol{x})$. Then

$$|z_i - x_i| \leq \|\boldsymbol{z} - \boldsymbol{x}\|_2 < r(\boldsymbol{x})$$

for every $i = 1, \ldots, d$. Hence

$$a_i < x_i - r(\boldsymbol{x}) < z_i < x_i + r(\boldsymbol{x}) < b_i$$

for every $i = 1, \ldots, d$. Thus $\boldsymbol{z} \in Q$, and consequently $B_{r(\boldsymbol{x})}(\boldsymbol{x}) \subseteq Q$. Since every $\boldsymbol{x} \in Q$ lies in such a ball, we can cover the rectangle by these balls,

$$Q = \bigcup_{\boldsymbol{x} \in Q} B_{r(\boldsymbol{x})}(\boldsymbol{x})$$

and deduce $Q \in \mathcal{T}$, i.e. Q is open in the canonical topology (Example 2.1). In particular, every rational open rectangle $Q \in \mathcal{R}$ is open, hence Borel by Definition 2.2. Thus $\mathcal{R} \subseteq \mathcal{T}$ (see also Proposition 2.1).

Set Operations

If P_1 and P_2 are properties of elements of Ω, then we have

- $\{\omega \in \Omega : P_1(\omega) \wedge P_2(\omega)\} = \{\omega \in \Omega : P_1(\omega)\} \cap \{\omega \in \Omega : P_2(\omega)\}$ (0.2)
- $\{\omega \in \Omega : P_1(\omega) \vee P_2(\omega)\} = \{\omega \in \Omega : P_1(\omega)\} \cup \{\omega \in \Omega : P_2(\omega)\}$ (0.3)
- $\{\omega \in \Omega : \neg P_1(\omega)\} = \{\omega \in \Omega : P_1(\omega)\}^{\complement}$ (0.4)

Cartesian Product

Definition 0.1. The *Cartesian product* of two sets A and B is the set

$$A \times B = \{(a, b) : a \in A \wedge b \in B\}$$

i.e., the set of all ordered pairs (a, b) with $a \in A$ and $b \in B$. ◀

We have

$$(A_1 \times A_2) \cap (B_1 \times B_2) = (A_1 \cap B_1) \times (A_2 \cap B_2) \tag{0.5}$$

since

$$\begin{aligned}
(x, y) \in (A_1 \times A_2) \cap (B_1 \times B_2) &\iff (x, y) \in A_1 \times A_2 \wedge (x, y) \in B_1 \times B_2 \\
&\iff (x \in A_1 \wedge y \in A_2) \wedge (x \in B_1 \wedge y \in B_2) \\
&\iff (x \in A_1 \wedge x \in B_1) \wedge (y \in A_2 \wedge y \in B_2) \\
&\iff x \in A_1 \cap B_1 \wedge y \in A_2 \cap B_2 \\
&\iff (x, y) \in (A_1 \cap B_1) \times (A_2 \cap B_2)
\end{aligned}$$

But in general

$$(A_1 \times A_2) \cup (B_1 \times B_2) \neq (A_1 \cup B_1) \times (A_2 \cup B_2) \tag{0.6}$$

De Morgan's Laws

Let $(A_i)_{i\in I}$ be a family of subsets of Ω, where I is an arbitrary, possibly countably or uncountably infinite, indexing set.

We have

$$\left(\bigcap_{i\in I} A_i\right)^{\complement} = \bigcup_{i\in I} A_i^{\complement} \tag{0.7}$$

since

$$\begin{aligned} x \in \left(\bigcap_{i\in I} A_i\right)^{\complement} &\iff x \notin \bigcap_{i\in I} A_i \\ &\iff \neg\big(\forall i \in I : x \in A_i\big) \\ &\iff \exists i \in I : x \notin A_i \\ &\iff \exists i \in I : x \in A_i^{\complement} \\ &\iff x \in \bigcup_{i\in I} A_i^{\complement} \end{aligned}$$

We have

$$\left(\bigcup_{i\in I} A_i\right)^{\complement} = \bigcap_{i\in I} A_i^{\complement} \tag{0.8}$$

since

$$\begin{aligned} x \in \left(\bigcup_{i\in I} A_i\right)^{\complement} &\iff x \notin \bigcup_{i\in I} A_i \\ &\iff \neg\big(\exists i \in I : x \in A_i\big) \\ &\iff \forall i \in I : x \notin A_i \\ &\iff \forall i \in I : x \in A_i^{\complement} \\ &\iff x \in \bigcap_{i\in I} A_i^{\complement} \end{aligned}$$

Limit Inferior and Limit Superior

Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of subsets of Ω and let $N \in \mathbb{N}$ be fixed.

We have

$$\liminf_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n\geq k} A_n = \bigcup_{k=N}^{\infty} \bigcap_{n\geq k} A_n$$

because removing finitely many initial conditions does not change the set of elements that belong to all A_n eventually.

We have

$$\limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n\geq k} A_n = \bigcap_{k=N}^{\infty} \bigcup_{n\geq k} A_n$$

because removing finitely many initial conditions does not change the set of elements that belong to infinitely many A_n.

Convergence of Functions

Let E be a set and let $f, (f_n)_{n\in\mathbb{N}}$ be functions $E \to \mathbb{R}$.

Definition 0.2 (pointwise convergence). We say that f_n converges to f pointwise on E, denoted by $f_n \xrightarrow{\text{p.w.}} f$, if

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for every $x \in E$.

Equivalently, for every $x \in E$ and every $\varepsilon > 0$, there exists $N = N(x, \varepsilon) \in \mathbb{N}$ such that

$$n \geq N \quad\Longrightarrow\quad |f_n(x) - f(x)| < \varepsilon$$ ◀

Definition 0.3 (uniform convergence). We say that f_n converges to f uniformly on E, denoted by $f_n \xrightarrow{\text{unif.}} f$, if

$$\sup_{x\in E} |f_n(x) - f(x)| \xrightarrow{n\to\infty} 0$$

Equivalently, for every $\varepsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$ such that

$$n \geq N \quad\Longrightarrow\quad |f_n(x) - f(x)| < \varepsilon$$

for every $x \in E$. ◀

In particular, uniform convergence implies pointwise convergence, but the converse is false in general.

Example 0.1. Let $f_n : [0,1] \to \mathbb{R}$ be given by $f_n(x) = x^n$. Then $f_n \xrightarrow{\text{p.w.}} \mathbf{1}_{\{1\}}$ on $[0,1]$, but the convergence is not uniform, since $\sup_{x\in[0,1]} |f_n(x) - \mathbf{1}_{\{1\}}(x)| = 1$ for every $n \in \mathbb{N}$. ◀

1 Probability Spaces

1.1 Measurable Spaces

Let Ω be a non-empty set. We call Ω the sample space, which specifies the possible outcomes of a random trial.

Based on the *elementary outcomes* $\omega \in \Omega$, one can also consider more complex *events* $A \subseteq \Omega$, which are based on the usual set operations.

If P is a property of elements of Ω, then the set

$$\{\omega \in \Omega : P(\omega)\} \subset \Omega$$

is the subset of all elements $\omega \in \Omega$ for which P is true. Furthermore, for $A \subset \Omega$, we write

$$A^{\complement} := \{\omega \in \Omega : \omega \notin A\}$$

for the *complement* of A in Ω. The complement of Ω is the empty set $\varnothing$. Given two sets $A, B \subset \Omega$, we further define their *union* by

$$A \cup B := \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$$

their *intersection* by

$$A \cap B := \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$$

and their *difference* by

$$A \setminus B := A \cap B^{\complement}$$

The collection of all possible subsets of Ω is called the *power set* of Ω and is defined as

$$2^{\Omega} := \{A : A \subseteq \Omega\}$$

We are often interested in certain subsets of the sample space. To work with these subsets, we define a collection of events that is closed under the usual set operations.

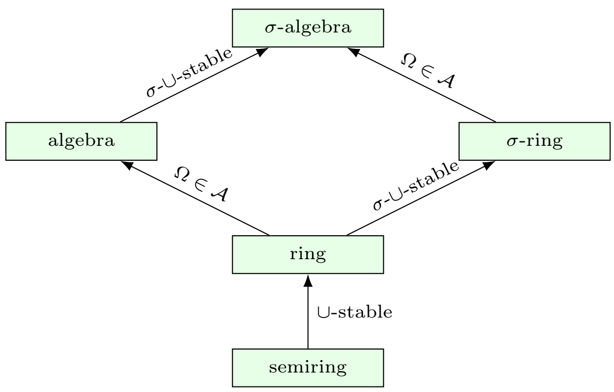
Definition 1.1 (Algebra). A set $\mathcal{F} \subseteq 2^{\Omega}$ is called an *algebra* on Ω if

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^{\complement} \in \mathcal{F}$

(iii) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

The sets $A \in \mathcal{F}$ are called *events*. ◀



The smallest algebra is $\{\varnothing, \Omega\}$. Since for any algebra $\mathcal{F} \subseteq 2^{\Omega}$, 2^{Ω} is the largest algebra. Note that any algebra is also closed under intersection, since $A \cap B = (A^{\complement} \cup B^{\complement})^{\complement}$ (De Morgan).

Proposition 1.1 (Algebras from partitions). For a partition Π of a finite set Ω, the associated collection

$$\mathcal{F}(\Pi) := \left\{ \bigcup_{E \in \Pi'} E : \Pi' \subseteq \Pi \right\}$$

is an algebra on Ω. ◁

Proof.
- Since Π is a partition of Ω, we have $\Omega = \bigcup_{E\in\Pi} E$, so by choosing $\Pi' = \Pi$, we obtain $\Omega \in \mathcal{F}(\Pi)$. ✓
- Now let $A \in \mathcal{F}(\Pi)$. Then there exists $\Pi_A \subseteq \Pi$ such that $A = \bigcup_{E\in\Pi_A} E$. Since the blocks of a partition are pairwise disjoint and cover Ω, we obtain

$$A^{\complement} = \Omega \setminus A = \left(\bigcup_{E\in\Pi} E\right) \setminus \left(\bigcup_{E\in\Pi_A} E\right) = \bigcup_{E\in\Pi\setminus\Pi_A} E$$

which is again a union of blocks. Hence $A^{\complement} \in \mathcal{F}(\Pi)$. ✓
- Finally, let $A, B \in \mathcal{F}(\Pi)$. Then there exist $\Pi_A, \Pi_B \subseteq \Pi$ such that $A = \bigcup_{E\in\Pi_A} E$ and $B = \bigcup_{E\in\Pi_B} E$. Therefore,

$$A \cup B = \left(\bigcup_{E\in\Pi_A} E\right) \cup \left(\bigcup_{E\in\Pi_B} E\right) = \bigcup_{E\in\Pi_A\cup\Pi_B} E,$$

and since $\Pi_A \cup \Pi_B \subseteq \Pi$, this shows that $A \cup B \in \mathcal{F}(\Pi)$. ✓ □

Example 1.1 (Different information ⇒ different algebras)**.** Alice generates two random bits. The sample space therefore is

$$\Omega = \{00, 01, 10, 11\}$$

Alice observes the exact outcome

$$\Pi_A = \big\{\{00\}, \{01\}, \{10\}, \{11\}\big\}$$

hence her observable events are all subsets of Ω:

$$\mathcal{F}_A = \mathcal{F}(\Pi_A) = 2^{\Omega}$$

Bob is only told the *number of ones* in the two bits. This induces the partition

$$\Pi_B = \big\{\{00\}, \{01, 10\}, \{11\}\big\}$$

of the sample space Ω. Therefore Bob can only observe events that are unions of blocks of Π_B, i.e., the algebra generated by Π_B:

$$\mathcal{F}_B = \mathcal{F}(\Pi_B) = \{\varnothing, \{00\}, \{01, 10\}, \{11\}, \{00, 11\}, \{00, 01, 10\}, \{01, 10, 11\}, \Omega\}$$

Cindy is only told whether the bits are *equal or different*. This induces the coarser partition

$$\Pi_C = \big\{\{00, 11\}, \{01, 10\}\big\}$$

hence her observable events form the algebra:

$$\mathcal{F}_C = \mathcal{F}(\Pi_C) = \{\varnothing, \{00, 11\}, \{01, 10\}, \Omega\}$$

We have

$$\mathcal{F}_A \supset \mathcal{F}_B \supset \mathcal{F}_C$$

expressing that more information corresponds to a finer partition and thus a larger algebra of observable events. ◀

Often, we encounter situations with infinitely many events and we want to assign probabilities to the case that all of them occur simultaneously. This is possible, as long as the number of events involved is at most countably infinite.

Definition 1.2. A set S is *countable* if $S = \varnothing$ or there exists a surjection $f : \mathbb{N} \to S$. ◀

Theorem 1.2. The Cartesian product of two countably infinite sets is countably infinite. ◁

Proof. Let A and B be two countably infinite sets. Then there exist surjections $f_A : \mathbb{N} \to A$ and $f_B : \mathbb{N} \to B$. We can define a surjection $f : \mathbb{N} \to A \times B$ by

$$f(n) := (f_A(p_1(n)), f_B(p_2(n)))$$

where $p_1(n)$ and $p_2(n)$ are the row and column indices of the n-th element in the diagonal enumeration of $\mathbb{N} \times \mathbb{N}$, i.e.,

n	1	2	3	4	5	6	7	8	9	$\cdots$
$p_1(n)$	1	2	1	3	2	1	4	3	2	$\cdots$
$p_2(n)$	1	1	2	1	2	3	1	2	3	$\cdots$

This shows that $A \times B$ is countable. □

Definition 1.3 (σ-algebra). A set $\mathcal{F} \subseteq 2^\Omega$ is called a *σ-algebra* on Ω if

(i) $\mathcal{F}$ is an algebra (Definition 1.1) on Ω

(ii) for any $A_1, A_2, \dots \in \mathcal{F}$ we have $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$

The sets in $\mathcal{F}$ are called measurable and the tuple $(\Omega, \mathcal{F})$ is called a *measurable space*. ◀

Remark 1.2. Let $\mathcal{I}$ be an index set, which may be uncountably infinite, and let $A_i \subseteq \Omega$ for each $i \in \mathcal{I}$. We denote the union of A_i over $i \in \mathcal{I}$ by

$$\bigcup_{i \in \mathcal{I}} A_i := \{\omega \in \Omega : \omega \in A_i \text{ for some } i \in \mathcal{I}\}$$

and the intersection of A_i over $i \in \mathcal{I}$ by

$$\bigcap_{i \in \mathcal{I}} A_i := \{\omega \in \Omega : \omega \in A_i \text{ for all } i \in \mathcal{I}\}$$

If $\mathcal{I}$ is countable, e.g. $\mathcal{I} = \mathbb{N}$, we often write $\bigcup_{i=1}^\infty A_i$ and $\bigcap_{i=1}^\infty A_i$ to represent the union and intersection of A_i over $i \in \mathcal{I}$, respectively. ◀

Example 1.3 (Co-finite algebra). Let $\Omega = \mathbb{N}$ and denote by $\mathcal{F}$ the collection of subsets of $\mathbb{N}$ that are finite or co-finite, i.e.,

$$\mathcal{F} = \{A \subseteq \mathbb{N} : |A| < \infty \text{ or } |A^\complement| < \infty\}$$

Then $\mathcal{F}$ is an algebra, but not a σ-algebra, e.g. the sequence $\{2\}, \{4\}, \{6\}, \dots$ is contained in $\mathcal{F}$, but $\bigcup_{i=1}^\infty \{2i\} \notin \mathcal{F}$. ◀

Recall that for sequences of real numbers, the limit inferior and limit superior are defined as

$$\liminf_{i \to \infty} x_i := \lim_{n \to \infty} \inf_{i \geq n} x_i \quad \text{and} \quad \limsup_{i \to \infty} x_i := \lim_{n \to \infty} \sup_{i \geq n} x_i$$

Definition 1.4. Let $A_1, A_2, \dots$ be a sequence of subsets of Ω. The *limit inferior* and *limit superior* of this sequence are defined as

$$\liminf_{i \to \infty} A_i := \bigcup_{i=1}^\infty \bigcap_{k=i}^\infty A_k \quad \text{and} \quad \limsup_{i \to \infty} A_i := \bigcap_{i=1}^\infty \bigcup_{k=i}^\infty A_k$$

respectively. ◀

If $\omega \in \liminf_{i\to\infty} A_i$, we say that "$\omega$ is in A_i eventually", i.e., for all but finitely many i, whereas $\omega \in \limsup_{i\to\infty} A_i$ means that "ω is in A_i infinitely often", i.e. for infinitely many i. We have

$$\liminf_{i \to \infty} A_i \subseteq \limsup_{i \to \infty} A_i \tag{1.1}$$

The power set 2^Ω is always the largest σ-algebra on Ω. In contrast, the smallest σ-algebra is given by $\mathcal{F} = \{\varnothing, \Omega\}$.

Theorem 1.3 (Generated σ-algebra)**.** Let $\mathcal{A} \subseteq 2^\Omega$. If we define the set

$\mathcal{A}$ is a family of subsets of Ω

$$\mathcal{M} := \{\mathcal{F} \mid \mathcal{A} \subseteq \mathcal{F} \subseteq 2^\Omega \text{ and } \mathcal{F} \text{ is a } \sigma\text{-algebra}\}$$

the smallest σ-algebra such that $\mathcal{A} \subseteq \sigma(\mathcal{A})$ is given by

$$\sigma(\mathcal{A}) := \bigcap_{\mathcal{F} \in \mathcal{M}} \mathcal{F}$$

which is called the σ-algebra generated by $\mathcal{A}$. ◁

Proof. $\mathcal{M}$ is non-empty since $\mathcal{A} \subset 2^\Omega$. $\sigma(\mathcal{A})$ is a σ-algebra:

- By Definition 1.1 (i), $\Omega \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{M}$, which means that $\Omega \in \sigma(\mathcal{A})$. ✓
- If $A \in \sigma(\mathcal{A})$, then $A \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{M}$. By Definition 1.1 (ii), this implies $A^\complement \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{M}$ and, hence, $A^\complement \in \sigma(\mathcal{A})$. ✓
- Let $A_n \in \sigma(\mathcal{A})$ for all $n \in \mathbb{N}$. Then $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ and each $\mathcal{F} \in \mathcal{M}$. By Definition 1.3 (ii), this implies
$$A := \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$$
for each such $\mathcal{F}$. Therefore, $A \in \sigma(\mathcal{A})$. ✓

Moreover, if $A \in \mathcal{A}$, then $A \in \mathcal{F}$ for all $\mathcal{F} \in \mathcal{M}$, so $A \in \sigma(\mathcal{A})$.

For minimality, let $\mathcal{G}$ be a σ-algebra with $\mathcal{A} \subseteq \mathcal{G}$. Then $\mathcal{G} \in \mathcal{M}$, hence

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{F} \in \mathcal{M}} \mathcal{F} \subseteq \mathcal{G}$$

implying that $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$. □

Lemma 1.4 (Monotonicity)**.** Let $\mathcal{A}, \mathcal{C} \subseteq 2^\Omega$ with $\mathcal{A} \subseteq \mathcal{C}$. Then $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{C})$. ◁

Proof. Since $\sigma(\mathcal{C})$ is a σ-algebra containing $\mathcal{C}$, it also contains $\mathcal{A}$. By the minimality of $\sigma(\mathcal{A})$ (Theorem 1.3), we obtain $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{C})$. □

Corollary 1.5 (Idempotency)**.** For every $\mathcal{A} \subseteq 2^\Omega$, $\sigma(\sigma(\mathcal{A})) = \sigma(\mathcal{A})$. ◁

Proof. Since $\mathcal{A} \subseteq \sigma(\mathcal{A})$, monotonicity (Lemma 1.4) yields $\sigma(\mathcal{A}) \subseteq \sigma(\sigma(\mathcal{A}))$. On the other hand, $\sigma(\mathcal{A})$ is itself a σ-algebra containing $\sigma(\mathcal{A})$, so by minimality of the generated σ-algebra, $\sigma(\sigma(\mathcal{A})) \subseteq \sigma(\mathcal{A})$. Thus equality holds. □

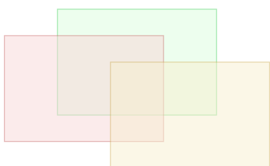
Example 1.4. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ denote two measurable spaces (see Definition 1.3). The *product σ-algebra* is defined as

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma\big(\{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}\big)$$

Note that the family

$$\{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$$

is not a σ-algebra in general. Intuitively, it corresponds to the fact that while rectangles are closed under intersections the same is not true for unions. The union of two rectangles in general is not a rectangle.



Hence the family is not even an algebra, and therefore certainly not a σ-algebra. If it were closed under complements, closure under unions would follow by De Morgan's laws, leading to a contradiction. It is, however, a semiring (Definition 2.5) of subsets of $\Omega_1 \times \Omega_2$. ◀

Example 1.5. Let $\Omega_1 = \Omega_2 = \mathbb{N}$ and $\mathcal{F}_1 = \mathcal{F}_2 = 2^{\mathbb{N}}$. Consider

$$\mathcal{R} := \{A_1 \times A_2 : A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\}$$

Define $R_n := \{n\} \times \{n\}$, $n \in \mathbb{N}$. Clearly, $\{n\} \in 2^{\mathbb{N}}$, so $R_n \in \mathcal{R}$ for all $n \in \mathbb{N}$. Consider

$$D := \bigcup_{n \in \mathbb{N}} R_n = \{(n, n) : n \in \mathbb{N}\}$$

which is the set of integer points on the diagonal of $\mathbb{N} \times \mathbb{N}$.

We claim that $D \notin \mathcal{R}$. Indeed, suppose $D = A \times B$ for some $A, B \subseteq \mathbb{N}$. Now take any $n_1, n_2 \in \mathbb{N}$ with $n_1 \neq n_2$. Since $(n_1, n_1) \in D$, we have $n_1 \in A$ and $n_1 \in B$. Since $(n_2, n_2) \in D$, we have $n_2 \in A$ and $n_2 \in B$. Thus, $n_1, n_2 \in A$ and $n_1, n_2 \in B$, hence $(n_1, n_2), (n_2, n_1) \in A \times B = D$, a contradiction.

Thus $\mathcal{R}$ is not closed under countable unions and therefore not a σ-algebra. ◀

1.2 Measures

Introducing the concept of a *measure*, we can assign volumes to sets.

Definition 1.5 (Measure)**.** Let $(\Omega, \mathcal{F})$ be a measurable space (see Definition 1.3). A set function $\mu : \mathcal{F} \to [0, \infty]$ is called a *measure* if

(i) $\mu(\varnothing) = 0$,

(ii) for any $A_1, A_2, \dots \in \mathcal{F}$ with $A_i \cap A_j = \varnothing$ for $i \neq j$, there holds the *σ-additivity*

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i) \tag{1.2}$$

A measure μ is called *σ-finite* if there exists a sequence $A_1, A_2, \dots \in \mathcal{F}$ such that $\Omega = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$. It is *finite* if $\mu(\Omega) < \infty$. The triple $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*. ◀

Remark 1.6. If a measure assumes the value ∞, we make the convention $\infty + \infty = \infty$ and $\infty + c = \infty$ for any $c \in \mathbb{R}$. ◀

Definition 1.6 (Counting measure)**.** Let Ω be a countable set and consider the measurable space $(\Omega, 2^{\Omega})$. We define the set function μ which assigns a set $A \in \mathcal{F}$ its cardinality, i.e., $\mu(A) = |A|$. The set function μ is a measure called the *counting measure*. ◀

Example 1.7. Consider $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$ where $\mu(A) := |A|$ is the counting measure on $\mathbb{N}$.

(a) μ is a measure because:

  (i) $\mu(\varnothing) = |\varnothing| = 0$. ✓

  (ii) For pairwise disjoint $(A_i)_{i \in \mathbb{N}}$ we have

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \left|\bigcup_{i=1}^{\infty} A_i\right| \overset{\text{disjoint}}{=} \sum_{i=1}^{\infty} |A_i| = \sum_{i=1}^{\infty} \mu(A_i)$$

  in all possible scenarios:
  - $|A_k| = \infty$ for some k's, i.e. some A_k have countably infinitely many elements. ✓
  - $|A_k| < \infty$ for all k and $|A_k| > 0$ for infinitely many k. ✓
  - $|A_k| < \infty$ for all k and $|A_k| > 0$ for finitely many k. ✓

(b) We can choose $A_i := \{i\}$. Then $\bigcup_{i=1}^{\infty} A_i = \mathbb{N}$ and $\mu(A_i) = |\{i\}| = 1 < \infty$ for all $i \in \mathbb{N}$. Hence the counting measure is σ-finite.

(c) But $\mu(\mathbb{N}) = |\mathbb{N}| = \infty$, so it is not finite. ◀

To assign *probabilities* to events, we consider the following specialization of a measure.

Definition 1.7 (Probability space)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. If $\mu(\Omega) = 1$, we call $(\Omega, \mathcal{F}, \mu)$ a *probability space* and μ a *probability measure*. In this case, we usually denote μ by $\mathbb{P}$. ◀

Example 1.8 (Discrete probability space)**.** Let $\Omega = \mathbb{N}$. Suppose $p_1, p_2, \dots$ are a sequence of nonnegative real numbers such that

$$\sum_{i=1}^{\infty} p_i = 1$$

For any $A \subset \Omega$, we define

$$\mathbb{P}(A) = \sum_{i \in A} p_i$$

Then, the set function $\mathbb{P} : 2^\Omega \to [0,1]$ is a probability measure and $(\Omega, 2^\Omega, \mathbb{P})$ is a probability space. ◀

Theorem 1.6 (monotonicity)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $A, B \in \mathcal{F}$ with $A \subseteq B$. Then,

$$\mu(A) \leq \mu(B) \tag{1.3}$$

◁

Proof. We have $B = A \cup (B \setminus A)$, whereby the sets A and $B \setminus A$ are disjoint. By (1.2), we have

$$\mu(B) = \mu(A) + \mu(B \setminus A)$$

and since $\mu(B \setminus A) \geq 0$ by definition, it follows that $\mu(B) \geq \mu(A)$. □

Corollary 1.7. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $A, B \in \mathcal{F}$ with $A \subset B$ and $\mu(A) < \infty$. Then,

$$\mu(B \setminus A) = \mu(B) - \mu(A) \tag{1.4}$$

◁

Proof. The sets A and $B \setminus A = B \cap A^{\complement}$ are disjoint. Since $A \subset B$, we have $B = (B \setminus A) \cup A$. By (1.2),

$$\mu(B) = \mu\big((B \setminus A) \cup A\big) = \mu(B \setminus A) + \mu(A)$$

Since $\mu(A) < \infty$, subtracting $\mu(A)$ yields the claim. □

Theorem 1.8 ($\triangle$-inequality)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $A, B \in \mathcal{F}$. Then,

$$\mu(A \cup B) \leq \mu(A) + \mu(B)$$

which can be interpreted as a *triangle inequality*. ◁

Proof. There holds $A \cup B = A \cup (B \setminus A)$, which yields

$$\mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(\underbrace{B \setminus A}_{\subset B}) \leq \mu(A) + \mu(B)$$

by Theorem 1.6. □

Theorem 1.9 (countable subadditivity)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $A_1, A_2, \dots \in \mathcal{F}$. Then

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) \leq \sum_{i=1}^{\infty} \mu(A_i) \tag{1.5}$$

◁

Proof. Define

$$B_i := A_i \setminus \bigcup_{j=1}^{i-1} A_j$$

where we set $\bigcup_{j=1}^{0} A_j := \varnothing$ for $i = 1$. Then $B_i \in \mathcal{F}$, the sets B_i are pairwise disjoint, $B_i \subseteq A_i$ for every i, and

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$$

Hence by σ-additivity (1.2) and monotonicity from Theorem 1.6,

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) = \sum_{i=1}^{\infty} \mu(B_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

□

Theorem 1.10. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let

$$A_1 \subset A_2 \subset \cdots \in \mathcal{F}$$

be an increasing sequence of sets. Then,

$$\boxed{\lim_{i\to\infty} \mu(A_i) = \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big)} \tag{1.6}$$

which can be seen as *continuity from below* for sequences of sets. ◁

Proof. We set $B_1 := A_1$ and $B_i := A_i \setminus A_{i-1}$ for $i > 1$. Then, the sets B_i are mutually disjoint by construction and $A_i = \bigcup_{j=1}^{i} B_j$. By σ-additivity (1.2), we infer

$$\begin{aligned}\lim_{i\to\infty} \mu(A_i) &= \lim_{i\to\infty} \mu\Big(\bigcup_{j=1}^{i} B_j\Big) = \lim_{i\to\infty} \sum_{j=1}^{i} \mu(B_j)\\ &= \sum_{j=1}^{\infty} \mu(B_j) \overset{(1.2)}{=} \mu\Big(\bigcup_{j=1}^{\infty} B_j\Big) = \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big)\end{aligned}$$

where we used the fact that $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ in the last step. □

Theorem 1.11. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let

$$A_1 \supset A_2 \supset \cdots \in \mathcal{F}$$

be a decreasing sequence of sets. Further assume that $\mu(A_{i_0}) < \infty$ for some $i_0 \in \mathbb{N}$. Then,

$$\boxed{\lim_{i\to\infty} \mu(A_i) = \mu\Big(\bigcap_{i=1}^{\infty} A_i\Big)} \tag{1.7}$$

which can be seen as *continuity from above* for sequences of sets. ◁

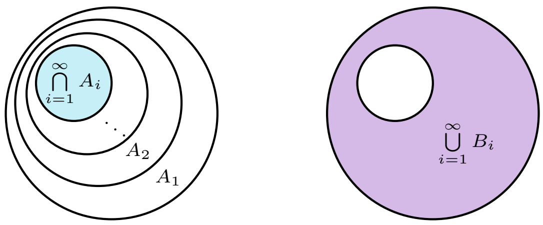
Proof. If $i_0 = 1$, set $B_i := A_1 \setminus A_i$ for $i = 1, 2, \ldots$. The sequence $B_1 \subset B_2 \subset \ldots$ is increasing and satisfies

$$\mu(B_i) = \mu(A_1) - \mu(A_i) \tag{1.8}$$

by (1.4) since $A_i \subset A_1$. In addition, we have

$$\begin{aligned}A_1 \setminus \Big(\bigcup_{i=1}^{\infty} B_i\Big) &= A_1 \cap \Big(\bigcup_{i=1}^{\infty} B_i\Big)^{\complement} \overset{(0.8)}{=} A_1 \cap \Big(\bigcap_{i=1}^{\infty} B_i^{\complement}\Big) = \bigcap_{i=1}^{\infty} (A_1 \cap B_i^{\complement})\\ &= \bigcap_{i=1}^{\infty} (A_1 \cap (A_1 \cap A_i^{\complement})^{\complement}) = \bigcap_{i=1}^{\infty} (A_1 \cap A_i) = \bigcap_{i=1}^{\infty} A_i\end{aligned} \tag{1.9}$$

which can be visualized as follows:



Thus,

$$\begin{aligned}\lim_{i\to\infty} \mu(A_i) &= \mu(A_1) - \big(\mu(A_1) - \lim_{i\to\infty} \mu(A_i)\big) = \mu(A_1) - \lim_{i\to\infty} \big(\mu(A_1) - \mu(A_i)\big)\\ &\overset{(1.8)}{=} \mu(A_1) - \lim_{i\to\infty} \mu(B_i) \overset{(1.6)}{=} \mu(A_1) - \mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) \overset{(1.4)}{=} \mu\Big(A_1 \setminus \big(\bigcup_{i=1}^{\infty} B_i\big)\Big)\\ &\overset{(1.9)}{=} \mu\Big(\bigcap_{i=1}^{\infty} A_i\Big)\end{aligned}$$

If $i_0 > 1$, we apply the same argument to the tail sequence starting at i_0. Since removing finitely many sets from a decreasing sequence does not change the intersection, we have $\bigcap_{i=i_0}^{\infty} A_i = \bigcap_{i=1}^{\infty} A_i$. Hence, $\lim_{i\to\infty} \mu(A_i) = \mu(\bigcap_{i=i_0}^{\infty} A_i) = \mu(\bigcap_{i=1}^{\infty} A_i)$. □

Example 1.9 (Why finiteness is necessary). Consider $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$ with counting measure $\mu(A) = |A|$ as in Example 1.7, and define $A_i := \{n \in \mathbb{N} : n \geq i\}$.

(a) The sequence $(A_i)_{i\in\mathbb{N}}$ is decreasing, since

$$A_i = \{i\} \cup \{n \in \mathbb{N} : n \geq i+1\} = \{i\} \cup A_{i+1}$$

and therefore $A_{i+1} \subseteq A_i$ for all $i \in \mathbb{N}$.

(b) Every A_i is (countably) infinite, so $\mu(A_i) = \infty$ for all $i \in \mathbb{N}$.

(c) Assume for contradiction that this is not the case. Then there exists some $n \in \mathbb{N}$ such that $n \in A_i$ for all $i \in \mathbb{N}$. However, $n \notin A_m$ for all $m > n$ by definition, a contradiction. ↯ Thus $\bigcap_{i=1}^{\infty} A_i = \varnothing$.

(d) From (b), we have $\mu(A_i) = \infty$ for all $i \in \mathbb{N}$. Thus $\lim_{i\to\infty} \mu(A_i) = \infty$, but from (c) we have

$$\mu\Big(\bigcap_{i=1}^{\infty} A_i\Big) = \mu(\varnothing) = |\varnothing| = 0$$

so continuity from above fails here. This does not contradict Theorem 1.11, since the finiteness assumption fails: there is no $i_0 \in \mathbb{N}$ such that $\mu(A_{i_0}) < \infty$. ◀

The properties in Theorems 1.10 and 1.11 are known as continuity from below and from above. If we want to indicate that $(A_i)_{i=1}^{\infty}$ is an increasing sequence of sets with union A, we write $A_i \nearrow A$. To indicate that it is a decreasing sequence of sets with intersection A, we write $A_i \searrow A$.

We have the following unifying result.

Theorem 1.12. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $A_1, A_2, \dots \in \mathcal{F}$ be a sequence of sets. Suppose that

$$\liminf_{i\to\infty} A_i = \limsup_{i\to\infty} A_i$$

and denote that set by $\lim\limits_{i\to\infty} A_i$. Then, if $\mu(\bigcup_{i=i_0}^{\infty} A_i) < \infty$ for some $i_0 \in \mathbb{N}$, we have

$$\lim_{i\to\infty} \mu(A_i) = \mu\Big(\lim_{i\to\infty} A_i\Big)$$

More generally, if $\mu(\bigcup_{i=i_0}^{\infty} A_i) < \infty$ for some $i_0 \in \mathbb{N}$, we have

$$\boxed{\mu\Big(\liminf_{i\to\infty} A_i\Big) \leq \liminf_{i\to\infty} \mu(A_i) \leq \limsup_{i\to\infty} \mu(A_i) \leq \mu\Big(\limsup_{i\to\infty} A_i\Big)}$$

◁

Proof. We prove the general statement, which implies the first one in the case that the limit inferior and limit superior coincide.

A general property for any sequence of real numbers is that $\liminf_{i\to\infty} x_i \leq \limsup_{i\to\infty} x_i$, so the central '$\leq$' in the statement is satisfied.

For the remaining inequalities, we introduce

$$B_i := \bigcap_{k=i}^{\infty} A_k \qquad C_i := \bigcup_{k=i}^{\infty} A_k$$

Note that $\{B_i\}_{i=1}^{\infty}$ is an increasing, while $\{C_i\}_{i=1}^{\infty}$ is a decreasing sequence of sets. By Definition 1.4, we have

$$\liminf_{i\to\infty} A_i = \bigcup_{i=1}^{\infty} B_i \qquad \limsup_{i\to\infty} A_i = \bigcap_{i=1}^{\infty} C_i$$

Left '$\leq$': For $k \geq i$, we have $B_i \subseteq A_k$, and hence, by Theorem 1.6, $\mu(B_i) \leq \mu(A_k)$, whenever $k \geq i$. This yields

$$\mu(B_i) \leq \inf_{k\geq i} \mu(A_k)$$

Taking the limit $i \to \infty$ and using Theorem 1.10, we obtain

$$\mu\Big(\liminf_{i\to\infty} A_i\Big) \overset{\text{def}}{=} \mu\Big(\bigcup_{i=1}^{\infty} B_i\Big) \overset{(1.6)}{=} \lim_{i\to\infty} \mu(B_i) \leq \lim_{i\to\infty} \inf_{k\geq i} \mu(A_k) \overset{\text{def}}{=} \liminf_{i\to\infty} \mu(A_i)$$

Right '$\leq$': For $k \geq i$, we have $C_i \supseteq A_k$, and hence, by Theorem 1.6, $\mu(C_i) \geq \mu(A_k)$, whenever $k \geq i$. This yields

$$\mu(C_i) \geq \sup_{k \geq i} \mu(A_k)$$

Taking the limit $i \to \infty$ and using Theorem 1.11, we obtain

$$\mu\Big(\limsup_{i\to\infty} A_i\Big) \overset{\text{def}}{=} \mu\Big(\bigcap_{i=1}^{\infty} C_i\Big) \overset{(1.7)}{=} \lim_{i\to\infty} \mu(C_i) \geq \lim_{i\to\infty} \sup_{k\geq i} \mu(A_k) \overset{\text{def}}{=} \limsup_{i\to\infty} \mu(A_i)$$

This concludes the proof. □

To conclude this section, we remark that is impossible to define a probability measure that models the unifrom distribution for all subsets of $[0,1]$. A non-measurable subset is, for example, the *Vitali set*.

Fact 1.13. There is no *translation invariant* measure $\mu : 2^{[0,1]} \to [0,1]$, i.e.,

$$\mu(A + x) = \mu(A)$$

for all $A \subseteq [0,1]$ with $A + x \in [0,1]$, $x \in \mathbb{R}$ such that

$$\mu([a,b]) = b - a$$

for all $a < b$. ◁

2 Lebesgue measure

2.1 Borel sets

While σ-algebras formalize measurability, we often require concepts related to continuity. The latter is assessed by means of a topology. To bring these two concepts together, one considers the σ-algebra generated by a *topology*.

Definition 2.1 (Topology)**.** A set $\mathcal{T} \subseteq 2^\Omega$ is called a *topology* on Ω if

(i) $\varnothing, \Omega \in \mathcal{T}$

(ii) $O_1, O_2 \in \mathcal{T} \implies O_1 \cap O_2 \in \mathcal{T}$

(iii) $\mathcal{A} \subseteq \mathcal{T} \implies \left(\bigcup_{O \in \mathcal{A}} O\right) \in \mathcal{T}$

Sets $O \in \mathcal{T}$ are called *open* and sets A with $A^\complement \in \mathcal{T}$ are called *closed*. The tuple $(\Omega, \mathcal{T})$ is a *topological space*. ◀

Example 2.1. In the case $\Omega = \mathbb{R}^d$ for some $d \in \mathbb{N}$, the canonical topology is induced by the Euclidean norm

$$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$$

This is seen as follows. Define the ball

$$B_r(\boldsymbol{x}) := \{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{y} - \boldsymbol{x}\|_2 < r\}$$

with radius $r > 0$ and center $\boldsymbol{x} \in \mathbb{R}^d$. Then, for the union of all these balls

$$\mathcal{G} = \{B_r(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d, r > 0\} \subseteq 2^{\mathbb{R}^d}$$

there holds

$$\mathcal{T} = \left\{\left(\bigcup_{A \in \mathcal{A}} A\right) : \mathcal{A} \subseteq \mathcal{G}\right\}$$

◀

Definition 2.2 (Borel sets)**.** Let $(\Omega, \mathcal{T})$ be a topological space (Definition 2.1). The σ-field generated by $\mathcal{T}$ (Theorem 1.3), i.e., $\mathcal{B}(\Omega) := \sigma(\mathcal{T})$, is called the *Borel σ-field* and the sets in $\mathcal{B}(\Omega)$ are called *Borel sets*. ◀

Caution 2.2. Let $(\Omega, \mathcal{T})$ be a topological space. Since the σ-algebra generated by $\mathcal{T}$ contains $\mathcal{T}$, we have

$$\mathcal{T} \subseteq \sigma(\mathcal{T}) = \mathcal{B}(\Omega)$$

i.e. every open set is a Borel set. In general, however, not every Borel set is open! ◀

Example 2.1 (continuing)**.** The Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ can also be generated by any of the following sets

$$\{B_r(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^d, r > 0\}$$

$$\{A \subseteq \mathbb{R}^d : A \text{ is open}\} \qquad \{A \subseteq \mathbb{R}^d : A \text{ is closed}\} \qquad \{A \subseteq \mathbb{R}^d : A \text{ is compact}\}$$

$$\{B_r(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{Q}^d, r \in \mathbb{Q}^+\}$$

$$\{(\boldsymbol{a}, \boldsymbol{b}) : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{Q}^d, \boldsymbol{a} < \boldsymbol{b}\} \qquad \{[\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{Q}^d, \boldsymbol{a} < \boldsymbol{b}\}$$

$$\{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{Q}^d, \boldsymbol{a} < \boldsymbol{b}\} \qquad \{[\boldsymbol{a}, \boldsymbol{b}) : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{Q}^d, \boldsymbol{a} < \boldsymbol{b}\}$$

$$\{(-\infty, \boldsymbol{b}) : \boldsymbol{b} \in \mathbb{Q}^d\} \qquad \{(\boldsymbol{a}, \infty) : \boldsymbol{a} \in \mathbb{Q}^d\}$$

$$\{(-\infty, \boldsymbol{b}] : \boldsymbol{b} \in \mathbb{Q}^d\} \qquad \{[\boldsymbol{a}, \infty) : \boldsymbol{a} \in \mathbb{Q}^d\}$$

that is

$$\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{E})$$

for any of the above sets $\mathcal{E}$. Recall that a set is *compact* if it is closed and bounded, where the latter means that it fits inside a ball of finite radius (Definition 2.7). Moreover, for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, we write $\boldsymbol{a} < \boldsymbol{b}$ if and only if $a_i < b_i$ for all $i = 1, \dots, d$. For $\boldsymbol{a} < \boldsymbol{b}$, we define the open rectangle as the Cartesian product

$$(\boldsymbol{a}, \boldsymbol{b}) = \bigtimes_{i=1}^{d} (a_i, b_i)$$

and $[\boldsymbol{a}, \boldsymbol{b}]$, $(\boldsymbol{a}, \boldsymbol{b}]$, $[\boldsymbol{a}, \boldsymbol{b})$ are defined analogously. At the open end of intervals, we allow for $-\infty$ and ∞. ◀

Remark 2.3 (Open sets in $\mathbb{R}^d$)**.** Let $G \subseteq \mathbb{R}^d$. By Example 2.1, G is open in the canonical topology if and only if for every $\boldsymbol{x} \in G$ there exists $r > 0$ such that $B_r(\boldsymbol{x}) \subseteq G$. Equivalently, every point of an open set is an interior point. ◀

Proposition 2.1 (Rational open rectangles generate the Borel σ-algebra)**.** Let

$$\mathcal{R} = \{(\boldsymbol{a}, \boldsymbol{b}) : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{Q}^d,\ \ \boldsymbol{a} < \boldsymbol{b}\}$$

Then

(a) $\mathcal{R}$ is countable

(b) every open rectangle $(\boldsymbol{a}, \boldsymbol{b}) \subset \mathbb{R}^d$ can be written as a countable union of elements of $\mathcal{R}$

(c) $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{R})$ ◁

Proof.

(a) An element of $\mathcal{R}$ is determined by a pair $(\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{Q}^d \times \mathbb{Q}^d$ with $\boldsymbol{a} < \boldsymbol{b}$. Since $\mathbb{Q}$ is countable, $\mathbb{Q}^d$ is countable, and therefore $\mathbb{Q}^d \times \mathbb{Q}^d$ is countable as well (see Theorem 1.2). Hence every subset of it, in particular $\mathcal{R}$, is countable.

(b) Now let $(\boldsymbol{a}, \boldsymbol{b})$ be an open rectangle in $\mathbb{R}^d$, i.e. $(\boldsymbol{a}, \boldsymbol{b}) = \times_{i=1}^{d} (a_i, b_i)$. We claim that

$$(\boldsymbol{a}, \boldsymbol{b}) = \bigcup_{\substack{\boldsymbol{q}, \boldsymbol{r} \in \mathbb{Q}^d \\ \boldsymbol{a} < \boldsymbol{q} < \boldsymbol{r} < \boldsymbol{b}}} (\boldsymbol{q}, \boldsymbol{r})$$

where we note that $\{(\boldsymbol{q}, \boldsymbol{r}) \in \mathbb{Q}^d \times \mathbb{Q}^d : \boldsymbol{a} < \boldsymbol{q} < \boldsymbol{r} < \boldsymbol{b}\} \subset \mathcal{R} \subset \mathbb{Q}^d \times \mathbb{Q}^d$, i.e. the index set of the union is a subset of a countable set. Hence the union is over a countable index set.

'⊇': Let $\boldsymbol{x} \in (\boldsymbol{q}, \boldsymbol{r})$ for some $\boldsymbol{a} < \boldsymbol{q} < \boldsymbol{r} < \boldsymbol{b}$. Then for every $i = 1, \ldots, d$,

$$a_i < q_i < x_i < r_i < b_i$$

and therefore $\boldsymbol{x} \in (\boldsymbol{a}, \boldsymbol{b})$.

'⊆': Let $\boldsymbol{x} \in (\boldsymbol{a}, \boldsymbol{b})$. Then for every coordinate i, by density of $\mathbb{Q}$ in $\mathbb{R}$, there exist $q_i, r_i \in \mathbb{Q}$ such that $a_i < q_i < x_i < r_i < b_i$. Setting $\boldsymbol{q} = [q_1, \ldots, q_d]$, $\boldsymbol{r} = [r_1, \ldots, r_d]$, we obtain

$$\boldsymbol{a} < \boldsymbol{q} < \boldsymbol{x} < \boldsymbol{r} < \boldsymbol{b}$$

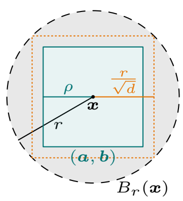
and hence $\boldsymbol{x} \in (\boldsymbol{q}, \boldsymbol{r})$. Since this rectangle appears in the union on the right-hand side, we conclude that $\boldsymbol{x} \in \bigcup_{\substack{\boldsymbol{q}, \boldsymbol{r} \in \mathbb{Q}^d \\ \boldsymbol{a} < \boldsymbol{q} < \boldsymbol{r} < \boldsymbol{b}}} (\boldsymbol{q}, \boldsymbol{r})$.

Thus every open rectangle in $\mathbb{R}^d$ is a countable union of elements of $\mathcal{R}$.

(c) '⊆': Let $U \subseteq \mathbb{R}^d$ be open, i.e. $U \in \mathcal{T}$. By Remark 2.3, for every $\boldsymbol{x} \in U$, there exists $r > 0$ such that $B_r(\boldsymbol{x}) \subseteq U$.

We construct an open rectangle around $\boldsymbol{x}$ contained in $B_r(\boldsymbol{x})$. Let $\rho > 0$ such that

$$\rho < \frac{r}{\sqrt{d}}$$



and define $(\boldsymbol{a}, \boldsymbol{b}) := (\boldsymbol{x} - \rho\mathbf{1},\ \boldsymbol{x} + \rho\mathbf{1})$. Then $\forall \boldsymbol{z} \in (\boldsymbol{a}, \boldsymbol{b})$, we have

$$\|\boldsymbol{z} - \boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{d} (z_i - x_i)^2} < \sqrt{\sum_{i=1}^{d} \rho^2} = \sqrt{d}\,\rho < r$$

so $\boldsymbol{z} \in B_r(\boldsymbol{x})$. Hence $(\boldsymbol{a}, \boldsymbol{b}) \subseteq B_r(\boldsymbol{x}) \subseteq U$.

By part (b), $(\boldsymbol{a}, \boldsymbol{b})$ can be written as a countable union of elements of $\mathcal{R}$. Thus for every $\boldsymbol{x} \in U$, there exists $Q \in \mathcal{R}$ with $\boldsymbol{x} \in Q \subseteq U$.

All these rectangles cover U, i.e.,

$$U = \bigcup_{\substack{Q \in \mathcal{R} \\ Q \subseteq U}} Q$$

and this union is countable since $\mathcal{R}$ is countable. Therefore $U \in \sigma(\mathcal{R})$. Hence $\mathcal{T} \subseteq \sigma(\mathcal{R})$ and by Lemma 1.4, $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{T}) \subseteq \sigma(\mathcal{R})$.

'⊇': Every set in $\mathcal{R}$ is an open rectangle, hence open in $\mathbb{R}^d$, and therefore Borel by Definition 2.2. Thus $\mathcal{R} \subseteq \mathcal{B}(\mathbb{R}^d)$. By Lemma 1.4, $\sigma(\mathcal{R}) \subseteq \mathcal{B}(\mathbb{R}^d)$. □

Proposition 2.2 (Open sets as countable unions of balls)**.** Let $G \subseteq \mathbb{R}^d$ be an open set in the canonical topology (see Example 2.1). Then:

(a) for every $\boldsymbol{x} \in G$ there exists $r > 0$ such that the open ball

$$B_r(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{y} - \boldsymbol{x}\|_2 < r\}$$

is contained in G, i.e. $B_r(\boldsymbol{x}) \subseteq G$.

(b) G can be written as a union of such balls, i.e.

$$G = \bigcup_{\boldsymbol{x} \in G} B_{r(\boldsymbol{x})}(\boldsymbol{x})$$

where for each $\boldsymbol{x} \in G$, we choose $r(\boldsymbol{x}) > 0$ such that $B_{r(\boldsymbol{x})}(\boldsymbol{x}) \subseteq G$.

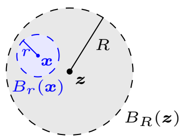
(c) G can be written as a countable union of balls with rational centers and rational radii, i.e.

$$G = \bigcup_{n=1}^{\infty} B_{r_n}(\boldsymbol{x}_n)$$

with $\boldsymbol{x}_n \in \mathbb{Q}^d$ and $r_n \in \mathbb{Q}^+$. ◁

Proof.

(a) Since $G \in \mathcal{T}$, $\exists \mathcal{A} \subseteq \mathcal{G}$ such that $G = \bigcup_{A \in \mathcal{A}} A$. So if $\boldsymbol{x} \in G$, then $\boldsymbol{x} \in A$ for some $A \in \mathcal{A}$, i.e. $\boldsymbol{x} \in B_R(\boldsymbol{z})$ for some $\boldsymbol{z} \in \mathbb{R}^d$, $R > 0$.



If we choose $r = R - \|\boldsymbol{x} - \boldsymbol{z}\|_2 > 0$, then $B_r(\boldsymbol{x}) \subseteq B_R(\boldsymbol{z}) \subseteq G$, since for any $\boldsymbol{y} \in B_r(\boldsymbol{x})$,

$$\|\boldsymbol{y} - \boldsymbol{z}\|_2 \leq \|\boldsymbol{y} - \boldsymbol{x}\|_2 + \|\boldsymbol{x} - \boldsymbol{z}\|_2 = \|\boldsymbol{y} - \boldsymbol{x}\|_2 + R - r < r + R - r = R$$

by the $\triangle$-inequality.

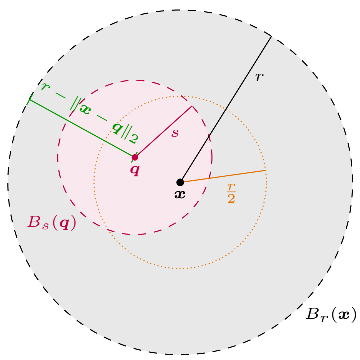
(b) '$\subseteq$': If $\boldsymbol{x} \in G$, then $\|\boldsymbol{x} - \boldsymbol{x}\|_2 = 0 < r(\boldsymbol{x})$, hence $\boldsymbol{x} \in B_{r(\boldsymbol{x})}(\boldsymbol{x})$, so $\boldsymbol{x} \in \bigcup_{\boldsymbol{z} \in G} B_{r(\boldsymbol{z})}(\boldsymbol{z})$.

'$\supseteq$': If $\boldsymbol{x} \in \bigcup_{\boldsymbol{z} \in G} B_{r(\boldsymbol{z})}(\boldsymbol{z})$, then $\exists \boldsymbol{y} \in G$ such that $\boldsymbol{x} \in B_{r(\boldsymbol{y})}(\boldsymbol{y})$. Since $B_{r(\boldsymbol{y})}(\boldsymbol{y}) \subseteq G$ by (a), we get $\boldsymbol{x} \in G$.

(c) Let $\boldsymbol{x} \in G$. By (a), there exists $r > 0$ such that $B_r(\boldsymbol{x}) \subseteq G$. By the density of $\mathbb{Q}^d$ in $\mathbb{R}^d$, choose $\boldsymbol{q} \in \mathbb{Q}^d$ with $\|\boldsymbol{x} - \boldsymbol{q}\|_2 < \frac{r}{2}$. By the density of $\mathbb{Q}$ in $\mathbb{R}$, choose $s \in \mathbb{Q}^+$ such that $\|\boldsymbol{x} - \boldsymbol{q}\|_2 < s < r - \|\boldsymbol{x} - \boldsymbol{q}\|_2$. Then $\boldsymbol{x} \in B_s(\boldsymbol{q})$. Moreover, by the $\triangle$-inequality, for every $\boldsymbol{y} \in B_s(\boldsymbol{q})$,

$$\|\boldsymbol{y} - \boldsymbol{x}\|_2 \leq \|\boldsymbol{y} - \boldsymbol{q}\|_2 + \|\boldsymbol{q} - \boldsymbol{x}\|_2 < s + \|\boldsymbol{q} - \boldsymbol{x}\|_2 < r$$

hence $\boldsymbol{y} \in B_r(\boldsymbol{x})$. Therefore $B_s(\boldsymbol{q}) \subseteq B_r(\boldsymbol{x}) \subseteq G$:



Thus every $\boldsymbol{x} \in G$ lies in some ball $B_s(\boldsymbol{q})$ with $\boldsymbol{q} \in \mathbb{Q}^d$, $s \in \mathbb{Q}^+$, and $B_s(\boldsymbol{q}) \subseteq G$. Hence

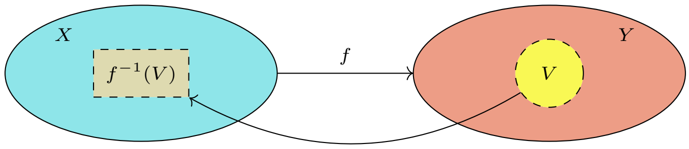
$$G = \bigcup_{\substack{\boldsymbol{q} \in \mathbb{Q}^d,\ s \in \mathbb{Q}^+ \\ B_s(\boldsymbol{q}) \subseteq G}} B_s(\boldsymbol{q})$$

and since $\mathbb{Q}^d \times \mathbb{Q}^+$ is countable (Theorem 1.2), the union is over a countable set. Hence we can enumerate it as $G = \bigcup_{n=1}^{\infty} B_{r_n}(\boldsymbol{x}_n)$ with $\boldsymbol{x}_n \in \mathbb{Q}^d$ and $r_n \in \mathbb{Q}^+$. □

Definition 2.3 (Topological Continuity)**.** Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces. A function $f : X \to Y$ is called *topologically continuous* if for all open subsets $V \in \mathcal{T}_Y$ the preimage

$$f^{-1}(V) := \{x \in X : f(x) \in V\}$$

is an open subset of X, i.e., $f^{-1}(V) \in \mathcal{T}_X$, visualized as follows:



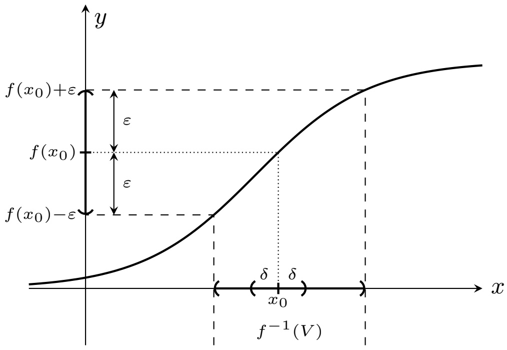
◀

Definition 2.4 (ε-δ Continuity)**.** A function $f : \mathbb{R} \to \mathbb{R}$ is called *continuous* at $x_0 \in \mathbb{R}$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon$. A function is called *continuous* if it is continuous at every point in its domain. ◀

Proposition 2.3. For a function $f : \mathbb{R} \to \mathbb{R}$ topological continuity is equivalent to the usual ε-δ-continuity. ◁

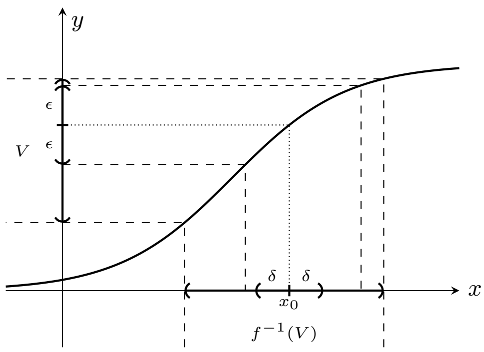
Proof. We show both implications.

⇒: Fix $x_0 \in \mathbb{R}$ and $\varepsilon > 0$ and consider $V := (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$, which is an open subset of $\mathbb{R}$. Hence by topological continuity also $f^{-1}(V)$ is open. Since $f(x_0) \in V$, we have $x_0 \in f^{-1}(V)$. As $f^{-1}(V)$ is open, there exists $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subseteq f^{-1}(V)$:



Therefore, if $|x - x_0| < \delta$, then $x \in f^{-1}(V)$, so $f(x) \in V = (f(x_0) - \varepsilon, f(x_0) + \varepsilon)$, and thus $|f(x) - f(x_0)| < \varepsilon$. Hence f is ε-δ-continuous at x_0. Since x_0 was arbitrary, f is ε-δ-continuous everywhere.

⇐: Let $V \subseteq \mathbb{R}$ be open. Take any $x_0 \in f^{-1}(V)$. Then $f(x_0) \in V$. Since V is open, there exists $\varepsilon > 0$ such that $(f(x_0) - \varepsilon, f(x_0) + \varepsilon) \subseteq V$. By ε-δ-continuity at x_0, there exists $\delta > 0$ such that $|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$:



This implies that for all $x \in (x_0 - \delta, x_0 + \delta)$, we have $f(x) \in (f(x_0) - \varepsilon, f(x_0) + \varepsilon) \subseteq V$, so $(x_0 - \delta, x_0 + \delta) \subseteq f^{-1}(V)$. Since x_0 was arbitrary in $f^{-1}(V)$, it follows that every point of $f^{-1}(V)$ is an interior point, and therefore $f^{-1}(V)$ is open. Hence f is topologically continuous.

This proves the equivalence. □

2.2 Construction of the Lebesgue-Borel measure

Given a cuboid of the form $[\boldsymbol{a}, \boldsymbol{b}]$, $(\boldsymbol{a}, \boldsymbol{b}]$, $[\boldsymbol{a}, \boldsymbol{b})$ or $(\boldsymbol{a}, \boldsymbol{b})$, it seems natural to assign the volume

$$\mu([\boldsymbol{a}, \boldsymbol{b}]) = \mu((\boldsymbol{a}, \boldsymbol{b}]) = \mu([\boldsymbol{a}, \boldsymbol{b})) = \mu((\boldsymbol{a}, \boldsymbol{b})) = \prod_{i=1}^{d} (b_i - a_i)$$

to it. In the following, we will see that this set function can uniquely be extended to $\mathcal{B}(\mathbb{R}^d)$ in a unique fashion. The resulting extension is called *Lebesgue-Borel measure* (Theorem 2.6).

Definition 2.5 (Semiring)**.** A set $\mathcal{S} \subseteq 2^\Omega$ is called a *semiring* if

(i) $\varnothing \in \mathcal{S}$

(ii) $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$

(iii) $A, B \in \mathcal{S}, B \subseteq A \implies$ there exist $C_1, \ldots, C_n \in \mathcal{S}, C_i \cap C_j = \varnothing$ such that

$$A \setminus B = \bigcup_{k=1}^{n} C_k$$

◀

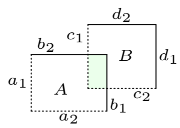
Overview over set systems:

Topology $\mathcal{T} \subseteq 2^\Omega$	**Algebra** $\mathcal{A} \subseteq 2^\Omega$	**σ-algebra** $\mathcal{F} \subseteq 2^\Omega$	**Semiring** $\mathcal{S} \subseteq 2^\Omega$
• $\varnothing, \Omega \in \mathcal{T}$ • finite intersections: $\bigcap_{i=1}^n O_i \in \mathcal{T}$ • arbitrary unions: $\bigcup_{i \in I} O_i \in \mathcal{T}$	• $\varnothing, \Omega \in \mathcal{A}$ • complement: $A^\complement \in \mathcal{A}$ • finite unions/intersec's: $\bigcup_{i=1}^n A_i \in \mathcal{A}$ $\bigcap_{i=1}^n A_i \in \mathcal{A}$	• $\varnothing, \Omega \in \mathcal{F}$ • complement: $A^\complement \in \mathcal{F}$ • countable unions/intersec's: $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$ $\bigcap_{i=1}^\infty A_i \in \mathcal{F}$	• $\varnothing \in \mathcal{S}$ • finite intersections: $\bigcap_{i=1}^n A_i \in \mathcal{S}$ • if $B \subseteq A$, then $A \setminus B = \bigcup_{k=1}^n C_k$ for disjoint $C_k \in \mathcal{S}$
$(\Omega, \mathcal{T})$: topological space		$(\Omega, \mathcal{F})$: measurable space	
Smallest: $\{\varnothing, \Omega\}$ Largest: 2^Ω	Smallest: $\{\varnothing, \Omega\}$ Largest: 2^Ω	Smallest: $\{\varnothing, \Omega\}$ Largest: 2^Ω	Smallest: $\{\varnothing\}$ Largest: 2^Ω

Example 2.4 (Half-open rectangles)**.** Let $d \in \mathbb{N}$. The half-open rectangles

$$\mathcal{C} := \{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d, \boldsymbol{a} \leq \boldsymbol{b}\}$$

form a semiring:



(i) $\varnothing = (\boldsymbol{a}, \boldsymbol{a}] \in \mathcal{C}$. ✓

(ii) Let $A = (\boldsymbol{a}, \boldsymbol{b}], B = (\boldsymbol{c}, \boldsymbol{d}] \in \mathcal{C}$. Then

$$A \cap B = \begin{cases} \varnothing & \text{if } a_i \geq d_i \text{ or } c_i \geq b_i \text{ for some } i = 1, \ldots, d \\ (\max(\boldsymbol{a}, \boldsymbol{c}), \min(\boldsymbol{b}, \boldsymbol{d})] & \text{otherwise} \end{cases}$$

where max and min are taken componentwise. In both cases, $A \cap B \in \mathcal{C}$. ✓

(iii) Let $A = (\boldsymbol{a}, \boldsymbol{b}]$ and $B = (\boldsymbol{\alpha}, \boldsymbol{\beta}]$ be in $\mathcal{C}$ with $B \subseteq A$. For every dimension $i = 1, \ldots, d$, define three intervals

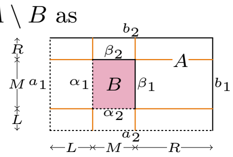
$$I_i(L) = (a_i, \alpha_i] \qquad I_i(M) = (\alpha_i, \beta_i] \qquad I_i(R) = (\beta_i, b_i]$$

where we use the convention that $I_i(\cdot) = \varnothing$ if the left endpoint is equal to the right endpoint (Since $B \subseteq A$, it cannot be greater than the right endpoint). Every vector $\boldsymbol{s} = [s_1, \ldots, s_d] \in \{L, M, R\}^d$ corresponds to a rectangle

$$C(\boldsymbol{s}) := \bigtimes_{i=1}^{d} I_i(s_i)$$

which is in $\mathcal{C}$. Note that rectangles corresponding to different vectors are disjoint. Furthermore, the rectangle corresponding to $\boldsymbol{s} = [M, \ldots, M]$ is exactly B, i.e. $C([M, \ldots, M]) = B$ and the union of all rectangles corresponding to all possible 3^d vectors $\boldsymbol{s} \in \{L, M, R\}^d$ is exactly A, i.e. $\bigcup_{\boldsymbol{s} \in \{L,M,R\}^d} C(\boldsymbol{s}) = A$. Therefore, we can write $A \setminus B$ as

$$A \setminus B = \bigcup_{\substack{\boldsymbol{s} \in \{L,M,R\}^d \\ \boldsymbol{s} \neq [M,\ldots,M]}} C(\boldsymbol{s})$$



where $\boldsymbol{s} = [s_1, \ldots, s_d]$ and $s_i \in \{L, M, R\}$ for $i = 1, \ldots, d$. The union is over (at most) $3^d - 1$ sets, hence finite. ✓ ◀

Similarly to measures (see Definition 1.5), we define the following properties of set functions on semirings

Definition 2.6. Let $\mathcal{S} \subseteq 2^\Omega$ be a semiring and $\mu : \mathcal{S} \to [0,\infty]$ be a set function. We say that μ is

(i) *additive* if for any $A_1, \dots, A_n \in \mathcal{S}$ with $A_i \cap A_j = \varnothing$ for $i \neq j$ and $\bigcup_{i=1}^n A_i \in \mathcal{S}$, there holds

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$$

(ii) *σ-subadditive* if for any $A \in \mathcal{S}$ and any $A_1, A_2, \dots \in \mathcal{S}$ with $A \subseteq \bigcup_{i=1}^\infty A_i$ (i.e. they cover A), there holds

$$\mu(A) \leq \sum_{i=1}^\infty \mu(A_i)$$

(iii) *σ-finite* if there exists a sequence $A_1, A_2, \dots \in \mathcal{S}$ such that $\Omega = \bigcup_{i=1}^\infty A_i$ and $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$. ◀

On a semiring, additivity is sufficient to guarantee finite subadditivity:

Lemma 2.4. Let $\mathcal{S} \subseteq 2^\Omega$ be a semiring and $\mu : \mathcal{S} \to [0,\infty]$ be an additive set function. Then μ is also finitely subadditive, i.e. for any $A \in \mathcal{S}$ and any $A_1, \dots, A_n \in \mathcal{S}$ with $A \subseteq \bigcup_{i=1}^n A_i$, there holds

$$\mu(A) \leq \sum_{i=1}^n \mu(A_i)$$

◁

Proof. Define

$$B_k := A_k \setminus \bigcup_{i=1}^{k-1} A_i = \bigcap_{i=1}^{k-1} \big(A_k \setminus (A_k \cap A_i)\big)$$

for $k = 1, \dots, n$, where we use the convention that $\bigcup_{i=1}^0 A_i = \varnothing$. This is the standard disjointification of $A_1, \dots, A_n$, i.e.

$$\bigsqcup_{k=1}^n B_k = \bigcup_{k=1}^n A_k$$

and $B_{k_1} \cap B_{k_2} = \varnothing$ for $k_1 \neq k_2$, which is emphasized by the symbol '$\bigsqcup$'.

For every $k = 1, \dots, n$ and every $i = 1, \dots, k-1$, we have $A_k \cap A_i \in \mathcal{S}$ by Definition 2.5 (ii). Hence, by Definition 2.5 (iii), each set $A_k \setminus (A_k \cap A_i)$ is a finite disjoint union of sets in $\mathcal{S}$. Therefore, B_k, being a finite intersection of such sets, is a finite union of sets in $\mathcal{S}$. Refining this finite union if necessary, there exist $c_k \in \mathbb{N}$ and pairwise disjoint sets $C_{k,1}, \dots, C_{k,c_k} \in \mathcal{S}$ such that

$$B_k = \bigsqcup_{i=1}^{c_k} C_{k,i}$$

Now consider

$$A_k \setminus B_k = A_k \cap \bigcup_{i=1}^{k-1} A_i = \bigcup_{i=1}^{k-1} (A_k \cap A_i)$$

Since each $A_k \cap A_i$ belongs to $\mathcal{S}$, the set $A_k \setminus B_k$ is a finite union of sets in $\mathcal{S}$. Refining again if necessary, there exist $d_k \in \mathbb{N}$ and pairwise disjoint sets $D_{k,1}, \dots, D_{k,d_k} \in \mathcal{S}$ such that

$$A_k \setminus B_k = \bigsqcup_{i=1}^{d_k} D_{k,i}$$

Since

$$A_k = B_k \sqcup (A_k \setminus B_k) = \left(\bigsqcup_{i=1}^{c_k} C_{k,i}\right) \sqcup \left(\bigsqcup_{i=1}^{d_k} D_{k,i}\right)$$

is a disjoint union, additivity yields

$$\mu(A_k) = \sum_{i=1}^{c_k} \mu(C_{k,i}) + \sum_{i=1}^{d_k} \mu(D_{k,i}) \geq \sum_{i=1}^{c_k} \mu(C_{k,i})$$

Moreover, the sets $B_1, \dots, B_n$ are pairwise disjoint and

$$A \subseteq \bigcup_{k=1}^{n} A_k = \bigsqcup_{k=1}^{n} B_k = \bigsqcup_{k=1}^{n} \bigsqcup_{i=1}^{c_k} C_{k,i} \tag{2.1}$$

Hence

$$A = \bigsqcup_{k=1}^{n} \bigsqcup_{i=1}^{c_k} (C_{k,i} \cap A)$$

where the union is disjoint. By additivity, we have

$$\mu(A) = \sum_{k=1}^{n} \sum_{i=1}^{c_k} \mu(C_{k,i} \cap A) \leq \sum_{k=1}^{n} \sum_{i=1}^{c_k} \mu(C_{k,i}) \leq \sum_{k=1}^{n} \mu(A_k)$$

Hence μ is finitely subadditive. □

Theorem 2.5 (Carathéodory Extension)**.** Let $\mathcal{S} \subseteq 2^{\Omega}$ be a semiring and

$$\mu : \mathcal{S} \to [0, \infty]$$

be an additive (i), σ-subadditive (ii) and σ-finite (iii) set function with $\mu(\varnothing) = 0$. Then there is a unique σ-finite measure

$$\tilde{\mu} : \sigma(\mathcal{S}) \to [0, \infty]$$

such that $\tilde{\mu}(E) = \mu(E)$ for all $E \in \mathcal{S}$. ◁

Example 2.5. Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, \dots, n$ be measure spaces (Definition 1.5). As in Example 1.4, we set

$$\Omega = \bigtimes_{i=1}^{n} \Omega_i$$

$$\mathcal{F} = \bigotimes_{i=1}^{n} \mathcal{F}_i := \sigma\left(\left\{\bigtimes_{i=1}^{n} A_i : A_i \in \mathcal{F}_i\right\}\right)$$

If all measures μ_i are σ-finite (iii), then there exists a unique measure $\mu = \bigotimes_{i=1}^{n} \mu_i$ called the *product measure* such that

$$\mu\left(\bigtimes_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mu_i(A_i) \tag{2.2}$$

◀

Using Theorem 2.5, we are now able to prove the following result:

Theorem 2.6. There exists a uniquely determined measure λ^d on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with the property that

$$\lambda^d((\boldsymbol{a}, \boldsymbol{b}]) = \prod_{i=1}^{d} (b_i - a_i)$$

for all $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ with $\boldsymbol{a} < \boldsymbol{b}$. The measure λ^d is called the *Lebesgue-Borel measure*. ◁

Definition 2.7. A set $K \subseteq \mathbb{R}^d$ is called *bounded* if there exist $\boldsymbol{x} \in \mathbb{R}^d$ and $r > 0$ such that

$$K \subseteq B_r(\boldsymbol{x})$$

where $B_r(\boldsymbol{x})$ is the open ball defined in Example 2.1. ◀

Definition 2.8. An *open cover* of a set $K \subseteq \mathbb{R}^d$ is a family $(O_i)_{i \in I}$ with $O_i \in \mathcal{T}$[1] for all $i \in I$ such that

$$K \subseteq \bigcup_{i \in I} O_i$$

A *finite subcover* is a finite subfamily $O_{i_1}, \dots, O_{i_n}$ with

$$K \subseteq \bigcup_{k=1}^{n} O_{i_k}$$

◀

Remark 2.6. In $\mathbb{R}^d$, the following definitions of compact sets are equivalent:

- K is closed (Definition 2.1) and bounded (Definition 2.7)
- every open cover of K admits a finite subcover

This is the Heine–Borel theorem. ◀

[1] $\mathcal{T}$ denotes the canonical topology on $\mathbb{R}^d$, see Example 2.1

Proof (of Theorem 2.6). Let $\mathcal{C}$ be the semiring of half-open rectangles from Example 2.4, and define

$$\mu\big((\boldsymbol{a},\boldsymbol{b}]\big) := \prod_{j=1}^{d}(b_j - a_j)$$

This set function μ is additive, since the volume of a box equals the sum of the volumes of finitely many disjoint boxes forming it. Hence, in order to apply Theorem 2.5, it remains to verify that μ is σ-subadditive and σ-finite.

We first show σ-subadditivity. Let $(\boldsymbol{a},\boldsymbol{b}],\ (\boldsymbol{a}^{(1)},\boldsymbol{b}^{(1)}],\ (\boldsymbol{a}^{(2)},\boldsymbol{b}^{(2)}],\dots \in \mathcal{C}$ with

$$(\boldsymbol{a},\boldsymbol{b}] \subseteq \bigcup_{i=1}^{\infty}(\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)}]$$

We show that

$$\mu\big((\boldsymbol{a},\boldsymbol{b}]\big) \le \sum_{i=1}^{\infty}\mu\big((\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)}]\big) \tag{2.3}$$

To this end, we use a compactness argument to reduce (2.3) to finite subadditivity. Let $\varepsilon > 0$. For every $i \in \mathbb{N}$, choose $\boldsymbol{b}_\varepsilon^{(i)} > \boldsymbol{b}^{(i)}$ (componentwise) such that

$$\mu\big((\boldsymbol{a}^{(i)},\boldsymbol{b}_\varepsilon^{(i)}]\big) \le \mu\big((\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)}]\big) + \varepsilon 2^{-i-1} \tag{2.4}$$

Further choose $\boldsymbol{a}_\varepsilon \in (\boldsymbol{a},\boldsymbol{b})$ such that

$$\mu\big((\boldsymbol{a}_\varepsilon,\boldsymbol{b}]\big) \ge \mu\big((\boldsymbol{a},\boldsymbol{b}]\big) - \frac{\varepsilon}{2} \tag{2.5}$$

Then we have the inclusions

$$[\boldsymbol{a}_\varepsilon,\boldsymbol{b}] \subseteq (\boldsymbol{a},\boldsymbol{b}] \subseteq \bigcup_{i=1}^{\infty}(\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)}] \subseteq \bigcup_{i=1}^{\infty}(\boldsymbol{a}^{(i)},\boldsymbol{b}_\varepsilon^{(i)})$$

Since $[\boldsymbol{a}_\varepsilon,\boldsymbol{b}]$ is compact and the last union is an open cover of it, there exists $N \in \mathbb{N}$ such that

$$(\boldsymbol{a}_\varepsilon,\boldsymbol{b}] \subseteq [\boldsymbol{a}_\varepsilon,\boldsymbol{b}] \subseteq \bigcup_{i=1}^{N}(\boldsymbol{a}^{(i)},\boldsymbol{b}_\varepsilon^{(i)}) \subseteq \bigcup_{i=1}^{N}(\boldsymbol{a}^{(i)},\boldsymbol{b}_\varepsilon^{(i)}]$$

Since μ is finitely subadditive on $\mathcal{C}$ (see Lemma 2.4), we obtain

$$\mu\big((\boldsymbol{a}_\varepsilon,\boldsymbol{b}]\big) \le \sum_{i=1}^{N}\mu\big((\boldsymbol{a}^{(i)},\boldsymbol{b}_\varepsilon^{(i)}]\big) \tag{2.6}$$

Combining (2.5), (2.6) and (2.4), we conclude

$$\mu\big((\boldsymbol{a},\boldsymbol{b}]\big) \le \frac{\varepsilon}{2} + \sum_{i=1}^{N}\Big(\mu\big((\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)}]\big) + \varepsilon 2^{-i-1}\Big)$$

Hence

$$\mu\big((\boldsymbol{a},\boldsymbol{b}]\big) \le \frac{\varepsilon}{2} + \sum_{i=1}^{N}\mu\big((\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)}]\big) + \sum_{i=1}^{N}\varepsilon 2^{-i-1} \le \varepsilon + \sum_{i=1}^{\infty}\mu\big((\boldsymbol{a}^{(i)},\boldsymbol{b}^{(i)}]\big)$$

since $\mu(\cdot) \ge 0$. Letting $\varepsilon \searrow 0$ yields (2.3). Thus μ is σ-subadditive.

It remains to check σ-finiteness. But $\mathbb{R}^d = \bigcup_{n=1}^{\infty}(-n,n]^d$ and $\mu\big((-n,n]^d\big) = (2n)^d < \infty$ for all $n \in \mathbb{N}$. Hence μ is σ-finite.

All assumptions of Theorem 2.5 are therefore satisfied. Consequently, there exists a unique σ-finite measure $\lambda^d : \mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{C}) \to [0,\infty]$ such that $\lambda^d\big((\boldsymbol{a},\boldsymbol{b}]\big) = \mu\big((\boldsymbol{a},\boldsymbol{b}]\big) = \prod_{j=1}^{d}(b_j - a_j)$ for all $\boldsymbol{a},\boldsymbol{b} \in \mathbb{R}^d$ with $\boldsymbol{a} < \boldsymbol{b}$. This is the Lebesgue–Borel measure. □

Exercise 2.7 (Translation invariance on cuboids). Work on $\mathbb{R}^d$ and let λ^d be the Lebesgue–Borel measure. For $\boldsymbol{h} \in \mathbb{R}^d$ and $A \subseteq \mathbb{R}^d$, define

$$A + \boldsymbol{h} \coloneqq \{\boldsymbol{x} + \boldsymbol{h} : \boldsymbol{x} \in A\}$$

(a) Show that for every half-open cuboid $C = (\boldsymbol{a}, \boldsymbol{b}]$ and every $\boldsymbol{h} \in \mathbb{R}^d$, one has

$$C + \boldsymbol{h} = (\boldsymbol{a} + \boldsymbol{h}, \boldsymbol{b} + \boldsymbol{h}]$$

(b) Deduce that

$$\lambda^d(C + \boldsymbol{h}) = \lambda^d(C)$$

for every half-open cuboid C.

(c) Fix $\boldsymbol{h} \in \mathbb{R}^d$ and consider

$$\mu_{\boldsymbol{h}}(A) \coloneqq \lambda^d(A + \boldsymbol{h})$$

for $A \in \mathcal{B}(\mathbb{R}^d)$. Show that $\mu_{\boldsymbol{h}}$ is a measure on $\mathcal{B}(\mathbb{R}^d)$.

(d) Show that $\mu_{\boldsymbol{h}}$ and λ^d coincide on the semiring

$$\mathcal{C} = \{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a} < \boldsymbol{b}\}$$

(e) Deduce from the uniqueness statement in Theorem 2.5 that

$$\lambda^d(A + \boldsymbol{h}) = \lambda^d(A)$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$. ✍

Solution.

(a) Let $C = (\boldsymbol{a}, \boldsymbol{b}]$ and $\boldsymbol{h} \in \mathbb{R}^d$. We show both inclusions.

$\subseteq$: First, let $\boldsymbol{y} \in C + \boldsymbol{h}$. Then there exists $\boldsymbol{x} \in C$ such that $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{h}$. Since $\boldsymbol{x} \in C = (\boldsymbol{a}, \boldsymbol{b}]$, we have $\boldsymbol{a} < \boldsymbol{x} \le \boldsymbol{b}$ componentwise, and hence

$$\boldsymbol{a} + \boldsymbol{h} < \boldsymbol{x} + \boldsymbol{h} = \boldsymbol{y} \le \boldsymbol{b} + \boldsymbol{h}$$

Thus $\boldsymbol{y} \in (\boldsymbol{a} + \boldsymbol{h}, \boldsymbol{b} + \boldsymbol{h}]$.

$\supseteq$: Conversely, let $\boldsymbol{y} \in (\boldsymbol{a} + \boldsymbol{h}, \boldsymbol{b} + \boldsymbol{h}]$. Define $\boldsymbol{x} \coloneqq \boldsymbol{y} - \boldsymbol{h}$. Then

$$\boldsymbol{a} + \boldsymbol{h} < \boldsymbol{y} \le \boldsymbol{b} + \boldsymbol{h} \quad \Longrightarrow \quad \boldsymbol{a} < \boldsymbol{y} - \boldsymbol{h} \le \boldsymbol{b}$$

and therefore $\boldsymbol{x} \in (\boldsymbol{a}, \boldsymbol{b}] = C$. Since $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{h}$, this implies $\boldsymbol{y} \in C + \boldsymbol{h}$.

Combining both inclusions yields the desired equality.

(b) By Theorem 2.6,

$$\lambda^d(C + \boldsymbol{h}) \overset{\text{(a)}}{=} \lambda^d((\boldsymbol{a} + \boldsymbol{h}, \boldsymbol{b} + \boldsymbol{h}]) = \prod_{i=1}^{d} \big((b_i + h_i) - (a_i + h_i)\big) = \prod_{i=1}^{d} (b_i - a_i) = \lambda^d(C)$$

(c) We first note that $A + \boldsymbol{h} \in \mathcal{B}(\mathbb{R}^d)$ for every $A \in \mathcal{B}(\mathbb{R}^d)$. Indeed, the translation map

$$T_{\boldsymbol{h}} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$$
$$\boldsymbol{x} \longmapsto T_{\boldsymbol{h}}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{h}$$

is a homeomorphism, hence maps Borel sets to Borel sets. Thus $\mu_{\boldsymbol{h}}$ is well-defined.

Clearly,

$$\mu_{\boldsymbol{h}}(\varnothing) = \lambda^d(\varnothing + \boldsymbol{h}) = \lambda^d(\varnothing) = 0$$

Now let $(A_n)_{n\in\mathbb{N}}$ be pairwise disjoint sets in $\mathcal{B}(\mathbb{R}^d)$. Then the translated sets $(A_n + \boldsymbol{h})_{n\in\mathbb{N}}$ are pairwise disjoint: if $\boldsymbol{y} \in (A_i + \boldsymbol{h}) \cap (A_j + \boldsymbol{h})$, then

$$\boldsymbol{y} = \boldsymbol{x}_i + \boldsymbol{h} = \boldsymbol{x}_j + \boldsymbol{h}$$

for some $\boldsymbol{x}_i \in A_i$ and $\boldsymbol{x}_j \in A_j$, so $\boldsymbol{x}_i = \boldsymbol{x}_j$. Therefore $A_i \cap A_j \neq \varnothing$, which implies $i = j$. Moreover,

$$\left(\bigcup_{n=1}^{\infty} A_n\right) + \boldsymbol{h} = \bigcup_{n=1}^{\infty} (A_n + \boldsymbol{h}) \tag{2.7}$$

Hence,

$$\mu_{\boldsymbol{h}}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lambda^d\left(\left(\bigcup_{n=1}^{\infty} A_n\right) + \boldsymbol{h}\right) \overset{(2.7)}{=} \lambda^d\left(\bigcup_{n=1}^{\infty}(A_n + \boldsymbol{h})\right)$$
$$\overset{(1.2)}{=} \sum_{n=1}^{\infty} \lambda^d(A_n + \boldsymbol{h}) = \sum_{n=1}^{\infty} \mu_{\boldsymbol{h}}(A_n)$$

Thus $\mu_{\boldsymbol{h}}$ is a measure on $\mathcal{B}(\mathbb{R}^d)$.

(d) Let $C = (\boldsymbol{a}, \boldsymbol{b}] \in \mathcal{C}$. Then by (b),

$$\mu_{\boldsymbol{h}}(C) = \lambda^d(C + \boldsymbol{h}) = \lambda^d(C)$$

Hence $\mu_{\boldsymbol{h}}$ and λ^d coincide on $\mathcal{C}$.

(e) Both $\mu_{\boldsymbol{h}}$ and λ^d are measures on $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{C})$ and, by (d), they agree on $\mathcal{C}$. Moreover, λ^d is σ-finite on $\mathcal{C}$, since

$$\mathbb{R}^d = \bigcup_{n=1}^{\infty}(-n, n]^d \qquad \text{and} \qquad \lambda^d((-n,n]^d) = (2n)^d < \infty$$

By the uniqueness statement in Theorem 2.5, we obtain

$$\mu_{\boldsymbol{h}} = \lambda^d$$

on $\mathcal{B}(\mathbb{R}^d)$. Thus,

$$\lambda^d(A + \boldsymbol{h}) = \mu_{\boldsymbol{h}}(A) = \lambda^d(A)$$

for all $A \in \mathcal{B}(\mathbb{R}^d)$. ◀

Definition 2.9 (Semicontinuity). A function $f : \mathbb{R} \to \mathbb{R}$ is called

- *upper semicontinuous* if for every $x_0 \in \mathbb{R}$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that

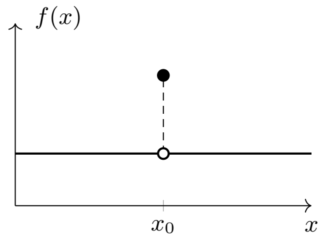
$$|x - x_0| < \delta \implies f(x) < f(x_0) + \varepsilon$$

i.e. it cannot suddenly "jump upward"

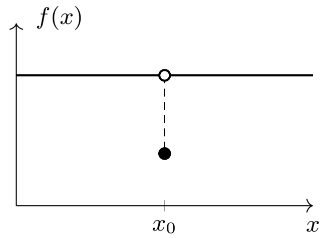
- *lower semicontinuous* if for every $x_0 \in \mathbb{R}$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|x - x_0| < \delta \implies f(x) > f(x_0) - \varepsilon$$

i.e. it cannot suddenly "jump downward"



upper semicontinuous, but not lower semicontinuous at x_0



lower semicontinuous, but not upper semicontinuous at x_0

A real-valued function f is upper (respectively, lower) semicontinuous at a point x_0 if, roughly speaking, the function values for arguments near x_0 are not much higher (respectively, lower) than $f(x_0)$. Equivalently, a function on a domain X is lower semi-continuous if its epigraph $\{(x,t) \in X \times \mathbb{R} : t \geq f(x)\}$ is closed in $X \times \mathbb{R}$, and upper semi-continuous if $-f$ is lower semi-continuous. ◀

Example 2.8 (Lebesgue-Stieltjes measure). Let $\Omega = \mathbb{R}$ and $\mathcal{S} = \{(a,b] : a, b \in \mathbb{R}, a \leq b\}$. $\mathcal{S}$ is a semiring (see Example 2.4) with $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$.

Consider a monotonically increasing and right-continuous function $F : \mathbb{R} \to \mathbb{R}$. Then the set function

$$\begin{aligned} \tilde{\mu}_F : \mathcal{S} &\longrightarrow [0, \infty) \\ (a,b] &\longmapsto \tilde{\mu}_F((a,b]) = F(b) - F(a) \end{aligned} \tag{2.8}$$

is additive, σ-subadditive, σ-finite and satisfies $\tilde{\mu}_F(\varnothing) = 0$.

By Theorem 2.5, we can uniquely extend $\tilde{\mu}_F$ to a σ-finite measure μ_F on $\mathcal{B}(\mathbb{R})$, called the *Lebesgue-Stieltjes measure* with *distribution function* F. ◀

Definition 2.10. A right continuous monotonically increasing function $F : \mathbb{R} \to [0,1]$ with

$$F(-\infty) \coloneqq \lim_{x\to-\infty} F(x) = 0 \qquad \text{and} \qquad F(\infty) \coloneqq \lim_{x\to\infty} F(x) = 1$$

is called a *cumulative distribution function*, *probability distribution function* or simply *distribution function*. If $\mathbb{P}$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then

$$\begin{aligned} F_{\mathbb{P}} : \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto F_{\mathbb{P}}(x) = \mathbb{P}((-\infty, x]) \end{aligned}$$

is called the *distribution function* of $\mathbb{P}$. ◀

Clearly, $F_{\mathbb{P}}$ is right continuous and $F_{\mathbb{P}}(-\infty) = 0$, since $\mathbb{P}$ is upper semicontinuous and finite. Since $\mathbb{P}$ is lower semicontinuous, we have $F_{\mathbb{P}}(\infty) = \mathbb{P}(\mathbb{R}) = 1$. Therefore, $F_{\mathbb{P}}$ is indeed a distribution function if $\mathbb{P}$ is a probability measure. More generally, every finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a Lebesgue-Stieltjes measure for some function F.

Exercise 2.9 (Semicontinuity and distribution functions)**.** Work on $\mathbb{R}$ equipped with its canonical topology. Recall Definition 2.9.

(a) Show that $F_{\mathbb{P}}$ is monotonically increasing.

(b) Show that $F_{\mathbb{P}}$ is upper semicontinuous.

(c) Show that $F_{\mathbb{P}}$ is lower semicontinuous from the right.

(d) Show that $F_{\mathbb{P}}$ is right continuous. ✍

Solution.

(a) Let $x, y \in \mathbb{R}$ with $x \le y$. Then $(-\infty, x] \subseteq (-\infty, y]$. Hence, by Theorem 1.6,

$$F_{\mathbb{P}}(x) = \mathbb{P}((-\infty, x]) \le \mathbb{P}((-\infty, y]) = F_{\mathbb{P}}(y)$$

(b) Fix $x \in \mathbb{R}$ and $\varepsilon > 0$.

- If $y \le x$, then by (a),

$$F_{\mathbb{P}}(y) \le F_{\mathbb{P}}(x) < F_{\mathbb{P}}(x) + \varepsilon$$

- It remains to control the case $y > x$. Let $(\delta_n)_{n\in\mathbb{N}}$ be a sequence with $\delta_n > 0$ and $\delta_n \searrow 0$. Consider

$$C_n \coloneqq (x, x + \delta_n]$$

for $n \in \mathbb{N}$. Then $(C_n)_{n\in\mathbb{N}}$ is a decreasing sequence of events with

$$\bigcap_{n=1}^{\infty} C_n = \varnothing$$

Thus, by continuity of $\mathbb{P}$ from above (Theorem 1.11),

$$\lim_{n\to\infty} \mathbb{P}(C_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} C_n\right) = \mathbb{P}(\varnothing) = 0$$

Hence there exists $N \in \mathbb{N}$ such that $\mathbb{P}((x, x + \delta_N]) < \varepsilon$ for all $n \ge N$. Set $\delta \coloneqq \delta_N$. If $x < y < x + \delta$, then

$$(x, y] \subseteq (x, x + \delta]$$

and therefore

$$\begin{aligned} F_{\mathbb{P}}(y) &= \mathbb{P}((-\infty, y]) \\ &= \mathbb{P}((-\infty, x] \sqcup (x, y]) \\ &\overset{(1.2)}{=} \mathbb{P}((-\infty, x]) + \mathbb{P}((x, y]) \\ &\overset{(1.3)}{\le} \mathbb{P}((-\infty, x]) + \mathbb{P}((x, x + \delta]) \\ &< F_{\mathbb{P}}(x) + \varepsilon \end{aligned}$$

Combining the cases $y \le x$ and $y > x$ yields

$$|y - x| < \delta \quad \Longrightarrow \quad F_{\mathbb{P}}(y) < F_{\mathbb{P}}(x) + \varepsilon$$

Thus $F_{\mathbb{P}}$ is upper semicontinuous.

(c) Fix $x \in \mathbb{R}$ and $\varepsilon > 0$. Since $F_\mathbb{P}$ is monotonically increasing by (a), we have for every $y \geq x$

$$F_\mathbb{P}(y) \geq F_\mathbb{P}(x) > F_\mathbb{P}(x) - \varepsilon$$

Thus we may choose any $\delta > 0$, and then

$$x \leq y < x + \delta \quad \Longrightarrow \quad F_\mathbb{P}(y) > F_\mathbb{P}(x) - \varepsilon$$

Hence $F_\mathbb{P}$ is lower semicontinuous from the right.

(d) Fix $x \in \mathbb{R}$ and $\varepsilon > 0$. By (b), there exists $\delta_{(\mathrm{b})} > 0$ such that

$$|y - x| < \delta_{(\mathrm{b})} \quad \Longrightarrow \quad F_\mathbb{P}(y) < F_\mathbb{P}(x) + \varepsilon$$

By (c), there exists $\delta_{(\mathrm{c})} > 0$ such that

$$x \leq y < x + \delta_{(\mathrm{c})} \quad \Longrightarrow \quad F_\mathbb{P}(y) > F_\mathbb{P}(x) - \varepsilon$$

Set $\delta \coloneqq \min\{\delta_{(\mathrm{b})}, \delta_{(\mathrm{c})}\}$. Then, for every $y \in [x, x + \delta)$, both estimates hold, and hence

$$F_\mathbb{P}(x) - \varepsilon < F_\mathbb{P}(y) < F_\mathbb{P}(x) + \varepsilon$$

Equivalently,

$$|F_\mathbb{P}(y) - F_\mathbb{P}(x)| < \varepsilon$$

for all $y \in [x, x + \delta)$. Thus

$$\lim_{y \searrow x} F_\mathbb{P}(y) = F_\mathbb{P}(x)$$

so $F_\mathbb{P}$ is right continuous.

Alternatively, let $(y_n)_{n \in \mathbb{N}}$ be a sequence with $y_n \searrow x$. Then the events $(-\infty, y_n]$ form a decreasing sequence and

$$\bigcap_{n=1}^{\infty} (-\infty, y_n] = (-\infty, x]$$

Therefore, by continuity of $\mathbb{P}$ from above (Theorem 1.11),

$$\lim_{n \to \infty} F_\mathbb{P}(y_n) = \lim_{n \to \infty} \mathbb{P}((-\infty, y_n]) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} (-\infty, y_n]\right) = \mathbb{P}((-\infty, x]) = F_\mathbb{P}(x)$$

◀

Theorem 2.7. The map $\mathbb{P} \longmapsto F_\mathbb{P}$ is a bijection from the set of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to the set of probability distribution functions. ◁

Exercise 2.10 (Lebesgue-Stieltjes measure for step functions)**.** Let $x_1 < \cdots < x_n$ be real numbers and $p_1, \ldots, p_n \geq 0$ with $\sum_{k=1}^n p_k = 1$. Define

$$F(x) \coloneqq \sum_{k=1}^{n} p_k \mathbf{1}_{[x_k, \infty)}(x) \tag{2.9}$$

for $x \in \mathbb{R}$. Let μ_F be the Lebesgue-Stieltjes measure (Example 2.8) associated with F.

(a) Show that F
  (i) is monotonically increasing
  (ii) is right continuous
  (iii) satisfies $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

(b) Compute $\mu_F((a, b])$ explicitly in terms of the points x_k and the weights p_k.

(c) Show that $\mu_F(\{x_k\}) = p_k$ for $k = 1, \ldots, n$.

(d) Show that if $A \in \mathcal{B}(\mathbb{R})$, then

$$\mu_F(A) = \sum_{k=1}^{n} p_k \mathbf{1}_A(x_k)$$

(e) Conclude that every finitely supported probability measure on $\mathbb{R}$, that is, every probability measure of the form

$$\nu = \sum_{k=1}^{n} p_k \delta_{x_k}$$

for some $x_1 < \cdots < x_n$ and $p_1, \ldots, p_n \geq 0$ with $\sum_{k=1}^n p_k = 1$, is a Lebesgue-Stieltjes measure (Example 2.8). ✍

Solution.

(a) (i) Let $x \le y$. Then

$$\mathbf{1}_{[x_k,\infty)}(x) \le \mathbf{1}_{[x_k,\infty)}(y)$$

for every $k \in \{1,\dots,n\}$. Since $p_k \ge 0$ for all k, summing over k yields $F(x) \le F(y)$. Hence F is monotonically increasing.

$y_1 \ge y_2 \ge y_3 \ge \cdots$, $\lim_{n\to\infty} y_n = x$ i.e. $\forall n \in \mathbb{N}: \ y_{n+1} \le y_n$, $\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n \ge N: \ |y_n - x| < \varepsilon$

(ii) Each function $\mathbf{1}_{[x_k,\infty)}$ is right continuous. Indeed, if $y_m \searrow x$, then: if $x < x_k$, eventually $y_m < x_k$, so $\mathbf{1}_{[x_k,\infty)}(y_m) = 0 = \mathbf{1}_{[x_k,\infty)}(x)$; if $x \ge x_k$, then $y_m \ge x_k$ for all m, so $\mathbf{1}_{[x_k,\infty)}(y_m) = \mathbf{1}_{[x_k,\infty)}(x) = 1$. Thus

$$\lim_{m\to\infty} \mathbf{1}_{[x_k,\infty)}(y_m) = \mathbf{1}_{[x_k,\infty)}(x)$$

A finite sum of right continuous functions is right continuous, so F is right continuous.

(iii) If $x \to -\infty$, then for every k we eventually have $x < x_k$, hence $\mathbf{1}_{[x_k,\infty)}(x) = 0$. Therefore $F(x) = 0$ eventually, and so $\lim_{x\to-\infty} F(x) = 0$. If $x \to \infty$, then for every k we eventually have $x \ge x_k$, hence $\mathbf{1}_{[x_k,\infty)}(x) = 1$. Therefore $F(x) = \sum_{k=1}^n p_k = 1$ eventually, and so $\lim_{x\to\infty} F(x) = 1$.

(b) Using the definition of F, we obtain

$$\begin{aligned}\mu_F((a,b]) &\overset{(2.8)}{=} F(b) - F(a) \overset{(2.9)}{=} \sum_{k=1}^n p_k \mathbf{1}_{[x_k,\infty)}(b) - \sum_{k=1}^n p_k \mathbf{1}_{[x_k,\infty)}(a) \\ &= \sum_{k=1}^n p_k \big(\mathbf{1}_{[x_k,\infty)}(b) - \mathbf{1}_{[x_k,\infty)}(a)\big) = \sum_{k=1}^n p_k \mathbf{1}_{(a,b]}(x_k) = \sum_{k:a<x_k\le b} p_k\end{aligned}$$

(c) Fix $k \in \{1,\dots,n\}$. If we choose $\varepsilon > 0$ such that

$$(x_k - \varepsilon, x_k] \cap \{x_1,\dots,x_n\} = \{x_k\}$$

then by part (b),

$$\mu_F((x_k-\varepsilon, x_k]) \overset{\text{(b)}}{=} \sum_{j:x_k-\varepsilon<x_j\le x_k} p_j = p_k$$

Moreover, F is constant on

$$(x_k-\varepsilon, x_k) = \bigcup_{\frac{1}{\varepsilon}<m\in\mathbb{N}} \left(x_k-\varepsilon, x_k - \frac{1}{m}\right]$$

since this interval contains no point x_j. Hence

$$\mu_F((x_k-\varepsilon,x_k)) \overset{(1.6)}{=} \lim_{m\to\infty} \mu_F\left(\left(x_k-\varepsilon, x_k-\frac{1}{m}\right]\right) = 0$$

Since $(x_k-\varepsilon,x_k] = (x_k-\varepsilon,x_k) \sqcup \{x_k\}$, additivity yields

$$\mu_F((x_k-\varepsilon,x_k]) = \mu_F((x_k-\varepsilon,x_k)) + \mu_F(\{x_k\}) = \mu_F(\{x_k\})$$

Consequently, $\mu_F(\{x_k\}) = p_k$.

(d) Define a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by

$$\mu := \sum_{k=1}^n p_k \delta_{x_k}$$

where $\delta_{x_k}(A) = \mathbf{1}_A(x_k)$ for every $A \in \mathcal{B}(\mathbb{R})$. Then for every $A \in \mathcal{B}(\mathbb{R})$,

$$\mu(A) = \sum_{k:x_k\in A} p_k$$

Moreover, for every $a \le b$,

$$\mu((a,b]) = \sum_{k:a<x_k\le b} p_k$$

By part (b), $\mu((a,b]) = \mu_F((a,b])$ for all $a \le b$. Since both μ and μ_F are finite Borel measures on $\mathbb{R}$ and the family $\mathcal{S} := \{(a,b] : a,b\in\mathbb{R},\ a\le b\}$ generates $\mathcal{B}(\mathbb{R})$, Theorem 2.5 implies that $\mu = \mu_F$. Therefore,

$$\mu_F(A) = \mu(A) = \sum_{k:x_k\in A} p_k = \sum_{k=1}^n p_k \mathbf{1}_A(x_k)$$

for every $A \in \mathcal{B}(\mathbb{R})$.

(e) Define $F(x) := \sum_{k=1}^{n} p_k \mathbf{1}_{[x_k,\infty)}(x)$. By part (d), the Lebesgue–Stieltjes measure μ_F associated with F satisfies

$$\mu_F(A) = \sum_{k:x_k \in A} p_k = \sum_{k=1}^{n} p_k \mathbf{1}_A(x_k) = \sum_{k=1}^{n} p_k \delta_{x_k}(A) = \nu(A)$$

for all $A \in \mathcal{B}(\mathbb{R})$. Hence $\nu = \mu_F$. Thus every finitely supported probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a Lebesgue–Stieltjes measure. ◀

2.3 Null sets

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, we might expect that every subset of an event E with $\mathbb{P}(E) = 0$ also has probability zero. In general, however, subsets of such events need not be measurable.

Definition 2.11. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

(i) A set $B \in \mathcal{F}$ with $\mu(B) = 0$ is called a μ-null set. By

$$\mathcal{N}_\mu := \{N \in 2^\Omega : B \in \mathcal{F},\ N \subseteq B,\ \mu(B) = 0\}$$

we denote the family of all subsets of μ-null sets.

(ii) Let $P(\omega)$ be a property of $\omega \in \Omega$. We say that P holds μ-almost everywhere (a.e.) if there exists a null set N such that $P(\omega)$ holds for every $\omega \in \Omega \setminus N$.

(iii) If $\mu = \mathbb{P}$ is a probability measure, we say that P holds $\mathbb{P}$-almost surely (a.s.).

(iv) The measure space $(\Omega, \mathcal{F}, \mu)$ is called complete if $\mathcal{F}$ contains all subsets of μ-null sets. ◀

Theorem 2.8. Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space and define

$$\mathcal{G} := \{A \cup N : A \in \mathcal{F},\ N \in \mathcal{N}_\mu\}$$

Then $\mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{N}_\mu)$. In particular, $\mathcal{G}$ is a σ-algebra. ◁

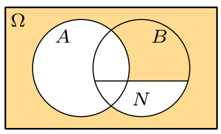
Proof. First note that if $N \in \mathcal{N}_\mu$, then by Definition 2.11 there exists a measurable μ-null set $B \in \mathcal{F}$ such that $N \subseteq B$.

We show that $\mathcal{G}$ is a σ-algebra (Definition 1.3):

(i) Since $\emptyset \in \mathcal{N}_\mu$ and $\Omega \in \mathcal{F}$, we have $\Omega = \Omega \cup \emptyset \in \mathcal{G}$. ✓

(ii) Now let $G = A \cup N \in \mathcal{G}$ with $A \in \mathcal{F}$ and $N \in \mathcal{N}_\mu$. Choose $B \in \mathcal{F}$ with $\mu(B) = 0$ and $N \subseteq B$. Then

$$G^\complement = \underbrace{(A \cup B)^\complement}_{\in \mathcal{F}} \cup \underbrace{\big(B \setminus (A \cup N)\big)}_{\substack{\subseteq B \\ \in \mathcal{N}_\mu}}$$



Therefore $G^\complement \in \mathcal{G}$. ✓

(iii) Consider $(G_i)_{i \in \mathbb{N}}$ with $G_i = A_i \cup N_i \in \mathcal{G}$ for $i \in \mathbb{N}$ with $A_i \in \mathcal{F}$ and $N_i \in \mathcal{N}_\mu$. Choose measurable μ-null sets $B_i \in \mathcal{F}$ such that $N_i \subseteq B_i$. Then

$$\bigcup_{i=1}^{\infty} G_i = \bigcup_{i=1}^{\infty} (A_i \cup N_i) \stackrel{2}{=} \left(\bigcup_{i=1}^{\infty} A_i\right) \cup \left(\bigcup_{i=1}^{\infty} N_i\right)$$

By Definition 1.3, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. Since $N_i \subseteq B_i$ for all i, we have

$$\bigcup_{i=1}^{\infty} N_i \subseteq \bigcup_{i=1}^{\infty} B_i$$

and by countable subadditivity from Theorem 1.9,

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) \leq \sum_{i=1}^{\infty} \mu(B_i) = 0$$

implying $\bigcup_{i=1}^{\infty} N_i \in \mathcal{N}_\mu$. Thus $\bigcup_{i=1}^{\infty} G_i \in \mathcal{G}$. ✓

[2] $x \in \bigcup_{i=1}^{\infty} (A_i \cup N_i) \Leftrightarrow \exists i \in \mathbb{N} : x \in A_i \cup N_i \Leftrightarrow \exists i \in \mathbb{N} : (x \in A_i \vee x \in N_i) \Leftrightarrow x \in \bigcup_{i=1}^{\infty} A_i \vee x \in \bigcup_{i=1}^{\infty} N_i \Leftrightarrow x \in \left(\bigcup_{i=1}^{\infty} A_i\right) \cup \left(\bigcup_{i=1}^{\infty} N_i\right)$

Hence $\mathcal{G}$ is a σ-algebra.

Since $A = A \cup \emptyset \in \mathcal{G}$ for every $A \in \mathcal{F}$ and $N = \emptyset \cup N \in \mathcal{G}$ for every $N \in \mathcal{N}_\mu$, we have

$$\mathcal{F} \cup \mathcal{N}_\mu \subseteq \mathcal{G}$$

As $\mathcal{G}$ is a σ-algebra, Theorem 1.3 yields

$$\sigma(\mathcal{F} \cup \mathcal{N}_\mu) \subseteq \mathcal{G}$$

Conversely, if $G \in \mathcal{G}$, then $G = A \cup N$ with $A \in \mathcal{F}$ and $N \in \mathcal{N}_\mu$. Since both A and N belong to $\mathcal{F} \cup \mathcal{N}_\mu$, they belong to $\sigma(\mathcal{F} \cup \mathcal{N}_\mu)$ by Theorem 1.3. As this is a σ-algebra, it is closed under unions, hence $G \in \sigma(\mathcal{F} \cup \mathcal{N}_\mu)$. Therefore

$$\mathcal{G} \subseteq \sigma(\mathcal{F} \cup \mathcal{N}_\mu)$$

Combining both inclusions gives the claim. □

We have the following fact, which states that every σ-finite measure space admits a completion.

Fact 2.9. Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space. Then there exist a unique smallest σ-algebra $\mathcal{F}^* \supset \mathcal{F}$ and an extension μ^* of μ to $\mathcal{F}^*$ such that $(\Omega, \mathcal{F}^*, \mu^*)$ is complete. The measure space $(\Omega, \mathcal{F}^*, \mu^*)$ is called the completion of $(\Omega, \mathcal{F}, \mu)$. Moreover,

$$\mathcal{F}^* = \sigma\left(\mathcal{F} \cup \mathcal{N}_\mu\right) \quad \text{and} \quad \mu^*(A \cup N) = \mu(A)$$

for any $A \in \mathcal{F}$ and $N \in \mathcal{N}_\mu$. ◁

Example 2.11. Let λ^d be the Lebesgue-Borel measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then λ^d can be uniquely extended to a measure λ^* on

$$\mathcal{B}^*(\mathbb{R}^d) = \sigma(\mathcal{B}(\mathbb{R}^d) \cup \mathcal{N})$$

where $\mathcal{N}$ is the family of all subsets of Lebesgue-Borel null sets. The σ-algebra $\mathcal{B}^*(\mathbb{R}^d)$ is called the σ-algebra of Lebesgue measurable sets. In what follows, we call λ^d the Lebesgue-Borel measure and λ^* the Lebesgue measure. ◀

3 Random variables

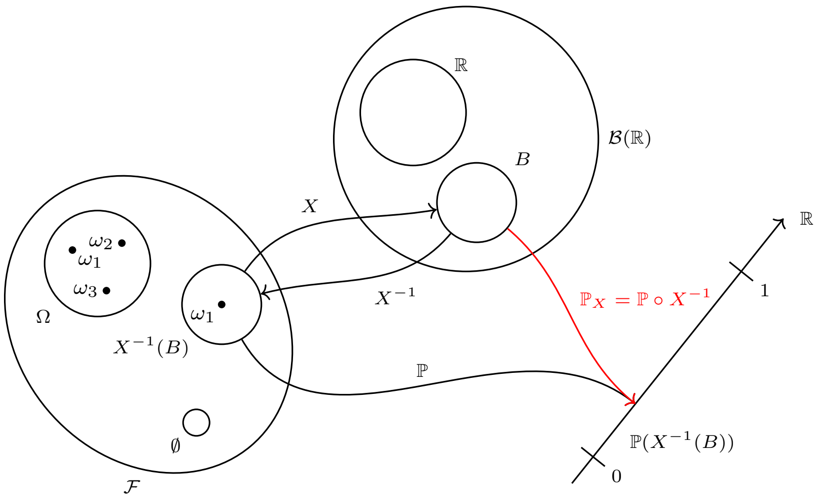
3.1 Measurable functions

Definition 3.1. Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces (see Definition 1.3). A function $f : \Omega \to \Omega'$ is called *$(\mathcal{F}, \mathcal{F}')$-measurable* or simply *measurable* if

$$f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F}$$

for every $B \in \mathcal{F}'$, i.e. the preimage of every measurable set is measurable. ◀

Definition 3.2. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $X : \Omega \to \mathbb{R}$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable, we call X a *random variable* (see diagram). If $\boldsymbol{X} : \Omega \to \mathbb{R}^d$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$-measurable, we call $\boldsymbol{X}$ a *random vector*. ◀



The concept of random variables ensures that we can assign probabilities to sets of the form

$$\{\omega \in \Omega : X(\omega) \in B\} = X^{-1}(B)$$

where $B \in \mathcal{B}(\mathbb{R})$ is a Borel set. The probability of the above set is given by $\mathbb{P}(X^{-1}(B))$.

Notation 3.3. As shorthand notation, we often write $\{X \in B\}$ for $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$, and $\mathbb{P}(X \in B)$ for $\mathbb{P}(X^{-1}(B))$. ◀

Example 3.1. If $\mathcal{F}$ is the trivial σ-algebra $\{\varnothing, \Omega\}$, then a function $X : \Omega \to \Omega'$ is measurable if and only if $X^{-1}(B) \in \{\varnothing, \Omega\}$ for every $B \in \mathcal{F}'$.

In particular, if $\Omega' = \mathbb{R}$ and $\mathcal{F}' = \mathcal{B}(\mathbb{R})$, then the only measurable functions $X : \Omega \to \mathbb{R}$ are the constant functions: if X were not constant, there would be $\omega_1, \omega_2 \in \Omega$ with $X(\omega_1) \neq X(\omega_2)$, and then the Borel set $\{X(\omega_1)\}$ would have preimage neither $\varnothing$ nor Ω, contradicting measurability.

If $\mathcal{F}$ is the power set 2^Ω, then every function $X : \Omega \to \Omega'$ is measurable. ◀

Theorem 3.1. Let $(\Omega, \mathcal{F})$ be a measurable space (Definition 1.3) and $A \subseteq \Omega$. The indicator function

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable if and only if $A \in \mathcal{F}$. ◁

Proof. Consider $B \in \mathcal{B}(\mathbb{R})$. We have

$$\mathbf{1}_A^{-1}(B) = \begin{cases} \varnothing & \text{if } 0, 1 \notin B \\ A & \text{if } 0 \notin B, 1 \in B \\ A^\complement & \text{if } 0 \in B, 1 \notin B \\ \Omega & \text{if } 0, 1 \in B \end{cases}$$

If $A \in \mathcal{F}$, then the four possible preimages are all in $\mathcal{F}$, so $\mathbf{1}_A$ is measurable. If $A \notin \mathcal{F}$, then the preimage of $B = \{1\}$ is A, which is not in $\mathcal{F}$, so $\mathbf{1}_A$ is not measurable. □

Example 3.2. Consider the indicator function $\mathbf{1}_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ of the rational numbers $\mathbb{Q} \subset \mathbb{R}$. Since $\mathbb{Q}$ is countable, we have $\mathbb{Q} \in \mathcal{B}(\mathbb{R})$. Hence $\mathbf{1}_{\mathbb{Q}}^{-1}(\{1\}) = \mathbb{Q}$ and $\mathbf{1}_{\mathbb{Q}}^{-1}(\{0\}) = \mathbb{R} \setminus \mathbb{Q}$ are Borel sets, so $\mathbf{1}_{\mathbb{Q}}$ is measurable. ◀

Fact 3.2 (Properties of preimage)**.** Let $f : \Omega \to \Omega'$ and $g : \Omega' \to \Omega''$ be functions. Define $h := g \circ f$. Then

(i) $f^{-1}(A^{\complement}) = (f^{-1}(A))^{\complement}$ for any $A \subseteq \Omega'$

(ii) Let $(A_i)_{i \in I}$ be a family of subsets of Ω'. Then

$$f^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f^{-1}(A_i) \qquad \text{and} \qquad f^{-1}\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f^{-1}(A_i)$$

(iii) $h^{-1}(B) = f^{-1}(g^{-1}(B))$ for any $B \subseteq \Omega''$ ◁

Proof. (i) $x \in f^{-1}(A^{\complement}) \Leftrightarrow f(x) \in A^{\complement} \Leftrightarrow f(x) \notin A \Leftrightarrow x \notin f^{-1}(A) \Leftrightarrow x \in (f^{-1}(A))^{\complement}$

(ii) $x \in f^{-1}\left(\bigcup_{i \in I} A_i\right) \Leftrightarrow f(x) \in \bigcup_{i \in I} A_i \Leftrightarrow \exists i \in I : f(x) \in A_i \Leftrightarrow \exists i \in I : x \in f^{-1}(A_i) \Leftrightarrow x \in \bigcup_{i \in I} f^{-1}(A_i)$
and
$x \in f^{-1}\left(\bigcap_{i \in I} A_i\right) \Leftrightarrow f(x) \in \bigcap_{i \in I} A_i \Leftrightarrow \forall i \in I : f(x) \in A_i \Leftrightarrow \forall i \in I : x \in f^{-1}(A_i) \Leftrightarrow x \in \bigcap_{i \in I} f^{-1}(A_i)$

(iii) $x \in h^{-1}(B) \Leftrightarrow h(x) \in B \Leftrightarrow g(f(x)) \in B \Leftrightarrow f(x) \in g^{-1}(B) \Leftrightarrow x \in f^{-1}(g^{-1}(B))$ □

3.2 Construction of measurable functions

Theorem 3.3. Let $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$, $(\Omega'', \mathcal{F}'')$ be measurable spaces (Definition 1.3). Assume $f : \Omega \to \Omega'$ is $(\mathcal{F}, \mathcal{F}')$-measurable and $g : \Omega' \to \Omega''$ is $(\mathcal{F}', \mathcal{F}'')$-measurable. Define $h := g \circ f$. Then $h : \Omega \to \Omega''$ is $(\mathcal{F}, \mathcal{F}'')$-measurable. ◁

Proof. Let $B \in \mathcal{F}''$. Since g is $(\mathcal{F}', \mathcal{F}'')$-measurable, we have $g^{-1}(B) \in \mathcal{F}'$. Since f is $(\mathcal{F}, \mathcal{F}')$-measurable, we have $f^{-1}(g^{-1}(B)) \in \mathcal{F}$. Hence, by Fact 3.2.(iii), $h^{-1}(B) \in \mathcal{F}$. □

Theorem 3.4 provides an effective way to check measurability of a function (Definition 3.1).

Theorem 3.4. Let $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$ be measurable spaces (Definition 1.3) and $f : \Omega \to \Omega'$ be a function. Let $\mathcal{A}$ be a generator of $\mathcal{F}'$, i.e. $\sigma(\mathcal{A}) = \mathcal{F}'$ (Theorem 1.3). Then f is $(\mathcal{F}, \mathcal{F}')$-measurable if and only if $f^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{A}$.

Proof. '⇒': If f is $(\mathcal{F}, \mathcal{F}')$-measurable, then $f^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{A} \subset \mathcal{F}'$.

'⇐': Assume $f^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{A}$. Consider

$$\mathcal{C} := \{B \subseteq \Omega' : f^{-1}(B) \in \mathcal{F}\}$$

i.e. the collection of subsets of Ω' whose preimages are measurable. By construction, $\mathcal{A} \subseteq \mathcal{C}$. It is therefore enough to show that $\mathcal{C}$ is a σ-algebra (Definition 1.3), since by Theorem 1.3 we then have $\mathcal{F}' = \sigma(\mathcal{A}) \subseteq \mathcal{C}$.

- $\Omega' \in \mathcal{C}$, since $f^{-1}(\Omega') = \Omega \in \mathcal{F}$. ✓
- Consider $B \in \mathcal{C}$. By definition, $f^{-1}(B) \in \mathcal{F}$. Since $\mathcal{F}$ is a σ-algebra, we have $(f^{-1}(B))^{\complement} \in \mathcal{F}$. Therefore, by Fact 3.2.(i), $f^{-1}(B^{\complement}) \in \mathcal{F}$. Hence $B^{\complement} \in \mathcal{C}$. ✓
- Now let $(B_i)_{i=1}^{\infty} \subseteq \mathcal{C}$. Then $f^{-1}(B_i) \in \mathcal{F}$ for every $i \in \mathbb{N}$. Since $\mathcal{F}$ is a σ-algebra, we have $\bigcup_{i=1}^{\infty} f^{-1}(B_i) \in \mathcal{F}$. Therefore, by Fact 3.2.(ii), $f^{-1}(\bigcup_{i=1}^{\infty} B_i) \in \mathcal{F}$. Hence $\bigcup_{i=1}^{\infty} B_i \in \mathcal{C}$. ✓

Thus, $\mathcal{C}$ is a σ-algebra. As mentioned, this implies that f is $(\mathcal{F}, \mathcal{F}')$-measurable. □

Definition 3.4 (Image measure)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $(\Omega', \mathcal{F}')$ be a measurable space and let $T : \Omega \to \Omega'$ be measurable. The *image measure* of μ under T is the measure

$$\mu_T : \mathcal{F}' \to [0, \infty]$$

defined by

$$\mu_T(A) := \mu(T^{-1}(A)) \tag{3.1}$$

for all $A \in \mathcal{F}'$.

If $X : \Omega \to \mathbb{R}$ is a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the image measure $\mathbb{P}_X$ is called the *distribution of X*. ◀

Recall the definition of (topologically) continuous functions from Definition 2.3. As the Borel σ-algebra is generated by a topology, we immediately have the following result on the measurability of continuous functions.

Corollary 3.5. A continuous function $f : \Omega \to \Omega'$ (Definition 2.3) between topological spaces (Ω, τ) and (Ω', τ') is $(\mathcal{B}(\Omega), \mathcal{B}(\Omega'))$-measurable, where $\mathcal{B}(\Omega) = \sigma(\tau)$ and $\mathcal{B}(\Omega') = \sigma(\tau')$. ◁

Proof. Since f is continuous (Definition 2.3), we have $f^{-1}(U) \in \tau \subseteq \mathcal{B}(\Omega)$ for every open set $U \in \tau'$. But τ' generates the Borel σ-algebra on Ω', i.e. $\mathcal{B}(\Omega') = \sigma(\tau')$ (see Definition 2.2). Hence, by Theorem 3.4, it follows that f is $(\mathcal{B}(\Omega), \mathcal{B}(\Omega'))$-measurable. □

Example 3.3. Any continuous function $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ is $(\mathcal{B}(\mathbb{R}^{d_1}), \mathcal{B}(\mathbb{R}^{d_2}))$-measurable. ◀

Theorem 3.6. Let $(\Omega, \mathcal{F})$ be a measurable space and let $f, g : \Omega \to \mathbb{R}$. If f and g are $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable, then so are $f + g$, $f - g$, $f \cdot g$, $c \cdot f$ for every $c \in \mathbb{R}$. If, in addition, $g(\omega) \neq 0$ for all $\omega \in \Omega$, then so is $f/g : \Omega \to \mathbb{R}$. ◁

Proof. Define $\alpha : \Omega \to \mathbb{R}^2$, $\alpha(\omega) := [f(\omega), g(\omega)]^\top$. For any open rectangle $(a, b) \times (c, d) \subseteq \mathbb{R}^2$, we have

$$\alpha^{-1}\big((a, b) \times (c, d)\big) = f^{-1}\big((a, b)\big) \cap g^{-1}\big((c, d)\big) \in \mathcal{F}$$

since (by the measurability of f and g) both preimages on the right-hand side belong to $\mathcal{F}$, and $\mathcal{F}$ is closed under finite intersections. The family of open rectangles generates $\mathcal{B}(\mathbb{R}^2)$, hence by Theorem 3.4, α is measurable.

The maps $(x, y) \mapsto x + y$, $x - y$, xy, $c \cdot x$, x/y are continuous if we consider x/y only on $\{(x, y) : y \neq 0\}$ and, hence, measurable by Corollary 3.5. Thus, by measurability of α, Fact 3.2.(iii) and Theorem 3.3, the functions $f + g$, $f - g$, $f \cdot g$, $c \cdot f$ and f/g are measurable. □

Remark 3.4. It is convenient to include the special symbols ∞ and $-\infty$ and to work in the extended real number system $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. The operations $\infty - \infty$, $(-\infty) + \infty$, ∞/∞, $\infty/(-\infty)$, $(-\infty)/\infty$, $(-\infty)/(-\infty)$ are not defined. However, by convention, we define $0 \cdot \infty$ and $0 \cdot (-\infty)$ to be equal to 0. ◀

Theorem 3.7. Let $(\Omega, \mathcal{F})$ be a measurable space and let $f_i : \Omega \to \overline{\mathbb{R}}$, $i \in \mathbb{N}$, be a sequence of $(\mathcal{F}, \mathcal{B}(\overline{\mathbb{R}}))$-measurable functions. Then,

$$\inf_{i \in \mathbb{N}} f_i \quad \text{and} \quad \sup_{i \in \mathbb{N}} f_i$$

are measurable. ◁

Proof. By Theorem 3.4, it suffices to check the preimages of the generator $\mathcal{A} = \{[-\infty, r] : r \in \mathbb{R}\}$ of $\mathcal{B}(\overline{\mathbb{R}})$. Consider the function $g(\omega) := \sup_{i \in \mathbb{N}} f_i(\omega)$. Let $r \in \mathbb{R}$. From the definition of the supremum, we obtain

$$g^{-1}([-\infty, r]) = \{\omega \in \Omega : \sup_{i \in \mathbb{N}} f_i(\omega) \leq r\} = \bigcap_{i=1}^{\infty} \{\omega \in \Omega : f_i(\omega) \leq r\} = \bigcap_{i=1}^{\infty} f_i^{-1}([-\infty, r])$$

Since each f_i is measurable and $[-\infty, r] \in \mathcal{B}(\overline{\mathbb{R}})$, it follows that $f_i^{-1}([-\infty, r]) \in \mathcal{F}$ for every $i \in \mathbb{N}$. Since the right-hand side is a countable intersection of measurable sets, the left-hand side is also measurable. Hence $g^{-1}([-\infty, r]) \in \mathcal{F}$ for every $r \in \mathbb{R}$. Therefore, by Theorem 3.4, the supremum $\sup_{i \in \mathbb{N}} f_i$ is measurable.

Similarly, for the infimum, we check the preimages of the generator $\mathcal{A} = \{[r, \infty] : r \in \mathbb{R}\}$ of $\mathcal{B}(\overline{\mathbb{R}})$. Now consider the function $g(\omega) := \inf_{i \in \mathbb{N}} f_i(\omega)$. Let $r \in \mathbb{R}$. From the definition of the infimum, we obtain

$$g^{-1}([r, \infty]) = \{\omega \in \Omega : \inf_{i \in \mathbb{N}} f_i(\omega) \geq r\} = \bigcap_{i=1}^{\infty} \{\omega \in \Omega : f_i(\omega) \geq r\} = \bigcap_{i=1}^{\infty} f_i^{-1}([r, \infty])$$

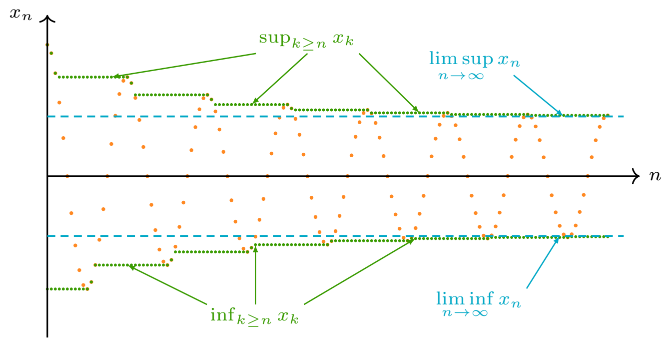
Since each f_i is measurable and $[r, \infty] \in \mathcal{B}(\overline{\mathbb{R}})$, it follows that $f_i^{-1}([r, \infty]) \in \mathcal{F}$ for every $i \in \mathbb{N}$. Since the right-hand side is a countable intersection of measurable sets, the left-hand side is also measurable. Hence $g^{-1}([r, \infty]) \in \mathcal{F}$ for every $r \in \mathbb{R}$. Therefore, by Theorem 3.4, the infimum $\inf_{i \in \mathbb{N}} f_i$ is measurable.

Alternatively, using Theorem 3.6 we could also direcly infer that $\inf_{i \in \mathbb{N}} f_i(\omega)$ is measurable, since $\inf_{i \in \mathbb{N}} f_i(\omega) = -\sup_{i \in \mathbb{N}}(-f_i(\omega))$. □

Recall that for a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$, the limit inferior and limit superior may be written as

$$\liminf_{n\to\infty} x_n = \sup_{n\in\mathbb{N}} \inf_{k\geq n} x_k \qquad \text{and} \qquad \limsup_{n\to\infty} x_n = \inf_{n\in\mathbb{N}} \sup_{k\geq n} x_k \tag{3.2}$$

respectively, since the sequence of infima is monotonically increasing and the sequence of suprema is monotonically decreasing.



Therefore, we may characterize them as follows:

- $\liminf_{n\to\infty} x_n$ is the largest $b_{\text{lower}} \in \mathbb{R}$ such that for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $x_n > b_{\text{lower}} - \varepsilon$ for all $n \geq N$
- $\limsup_{n\to\infty} x_n$ is the smallest $b_{\text{upper}} \in \mathbb{R}$ such that for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $x_n < b_{\text{upper}} + \varepsilon$ for all $n \geq N$
- for any sequence $(x_n)_{n\in\mathbb{N}}$, we have $\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n$, with equality when $\lim_{n\to\infty} x_n$ exists

Theorem 3.8. Let $f_i : \Omega \to \overline{\mathbb{R}}$, $i \in \mathbb{N}$, be a sequence of measurable functions. Then, the functions

$$\liminf_{i\to\infty} f_i \quad \text{and} \quad \limsup_{i\to\infty} f_i$$

are measurable. Furthermore, if $\lim_{i\to\infty} f_i(\omega)$ exists for all $\omega \in \Omega$, then $\lim_{i\to\infty} f_i$ is measurable. ◁

Proof. From Theorem 3.7 we know that the supremum and infimum of a sequence of measurable functions are measurable. By the previous characterization of the limit inferior and limit superior, it follows that they are measurable. If the pointwise limit of f_i exists, then the limit inferior and limit superior coincide, and hence the limit is measurable as well. □

Exercise 3.5 (Indicators, limsup, and liminf of sets). Let $(\Omega, \mathcal{F})$ be a measurable space and let $(A_n)_{n\in\mathbb{N}}$ be a sequence of sets in $\mathcal{F}$.

(i) Show that

$$\mathbf{1}_{\bigcup_{n=1}^\infty A_n} = \sup_{n\in\mathbb{N}} \mathbf{1}_{A_n}, \qquad \mathbf{1}_{\bigcap_{n=1}^\infty A_n} = \inf_{n\in\mathbb{N}} \mathbf{1}_{A_n}$$

(ii) Show that

$$\mathbf{1}_{\limsup_{n\to\infty} A_n} = \limsup_{n\to\infty} \mathbf{1}_{A_n}, \qquad \mathbf{1}_{\liminf_{n\to\infty} A_n} = \liminf_{n\to\infty} \mathbf{1}_{A_n}$$

(iii) Show that

$$\limsup_{n\to\infty} A_n, \qquad \liminf_{n\to\infty} A_n$$

belong to $\mathcal{F}$. ✍

Solution.

(i) We prove both identities pointwise. Fix $\omega \in \Omega$.

- sup:
  - If $\omega \in \bigcup_{n=1}^\infty A_n$, then $\omega \in A_n$ for some $n \in \mathbb{N}$. Hence $\mathbf{1}_{A_n}(\omega) = 1$ for some $n \in \mathbb{N}$, so

$$\sup_{n\in\mathbb{N}} \mathbf{1}_{A_n}(\omega) = 1 = \mathbf{1}_{\bigcup_{n=1}^\infty A_n}(\omega)$$

- If $\omega \notin \bigcup_{n=1}^{\infty} A_n$, then $\omega \notin A_n$ for all $n \in \mathbb{N}$. Hence $\mathbf{1}_{A_n}(\omega) = 0$ for all $n \in \mathbb{N}$, so
$$\sup_{n\in\mathbb{N}} \mathbf{1}_{A_n}(\omega) = 0 = \mathbf{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega)$$

Thus for every $\omega \in \Omega$,
$$\mathbf{1}_{\bigcup_{n=1}^{\infty} A_n}(\omega) = \sup_{n\in\mathbb{N}} \mathbf{1}_{A_n}(\omega)$$

- inf:
  - If $\omega \in \bigcap_{n=1}^{\infty} A_n$, then $\omega \in A_n$ for all $n \in \mathbb{N}$. Hence $\mathbf{1}_{A_n}(\omega) = 1$ for all $n \in \mathbb{N}$, so
$$\inf_{n\in\mathbb{N}} \mathbf{1}_{A_n}(\omega) = 1 = \mathbf{1}_{\bigcap_{n=1}^{\infty} A_n}(\omega)$$
  - If $\omega \notin \bigcap_{n=1}^{\infty} A_n$, then $\omega \notin A_n$ for some $n \in \mathbb{N}$. Hence $\mathbf{1}_{A_n}(\omega) = 0$ for some $n \in \mathbb{N}$, so
$$\inf_{n\in\mathbb{N}} \mathbf{1}_{A_n}(\omega) = 0 = \mathbf{1}_{\bigcap_{n=1}^{\infty} A_n}(\omega)$$

  Thus for every $\omega \in \Omega$,
$$\mathbf{1}_{\bigcap_{n=1}^{\infty} A_n}(\omega) = \inf_{n\in\mathbb{N}} \mathbf{1}_{A_n}(\omega)$$

Since $\omega \in \Omega$ was arbitrary, both identities follow.

(ii) We again prove both identities pointwise. Fix $\omega \in \Omega$.

- limsup:
  - If $\omega \in \limsup_{n\to\infty} A_n$, then $\omega \in A_n$ for infinitely many $n \in \mathbb{N}$. Hence $\mathbf{1}_{A_n}(\omega) = 1$ for infinitely many $n \in \mathbb{N}$, and therefore
$$\limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega) = 1 = \mathbf{1}_{\limsup_{n\to\infty} A_n}(\omega)$$
  - If $\omega \notin \limsup_{n\to\infty} A_n$, then ω belongs to only finitely many A_n. Hence $\mathbf{1}_{A_n}(\omega) = 0$ for all sufficiently large n, and therefore
$$\limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega) = 0 = \mathbf{1}_{\limsup_{n\to\infty} A_n}(\omega)$$

  Thus
$$\mathbf{1}_{\limsup_{n\to\infty} A_n}(\omega) = \limsup_{n\to\infty} \mathbf{1}_{A_n}(\omega)$$

- liminf:
  - If $\omega \in \liminf_{n\to\infty} A_n$, then $\omega \in A_n$ for all sufficiently large n. Hence $\mathbf{1}_{A_n}(\omega) = 1$ for all sufficiently large n, and therefore
$$\liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega) = 1 = \mathbf{1}_{\liminf_{n\to\infty} A_n}(\omega)$$
  - If $\omega \notin \liminf_{n\to\infty} A_n$, then $\omega \notin A_n$ for infinitely many $n \in \mathbb{N}$. Hence $\mathbf{1}_{A_n}(\omega) = 0$ for infinitely many $n \in \mathbb{N}$, and therefore
$$\liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega) = 0 = \mathbf{1}_{\liminf_{n\to\infty} A_n}(\omega)$$

  Thus
$$\mathbf{1}_{\liminf_{n\to\infty} A_n}(\omega) = \liminf_{n\to\infty} \mathbf{1}_{A_n}(\omega)$$

Since $\omega \in \Omega$ was arbitrary, both identities follow.

(iii) By Theorem 3.1, each $\mathbf{1}_{A_n}$ is measurable. Hence, by Theorem 3.8, the functions
$$\limsup_{n\to\infty} \mathbf{1}_{A_n}, \qquad \liminf_{n\to\infty} \mathbf{1}_{A_n}$$
are measurable. By (ii), this means that
$$\mathbf{1}_{\limsup_{n\to\infty} A_n}, \qquad \mathbf{1}_{\liminf_{n\to\infty} A_n}$$
are measurable. Moreover, since $\{1\} \in \mathcal{B}(\mathbb{R})$, we have
$$\limsup_{n\to\infty} A_n = \mathbf{1}^{-1}_{\limsup_{n\to\infty} A_n}(\{1\}), \qquad \liminf_{n\to\infty} A_n = \mathbf{1}^{-1}_{\liminf_{n\to\infty} A_n}(\{1\})$$
so, by Definition 3.1, both sets belong to $\mathcal{F}$. ◀

Proposition 3.9. Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \overline{\mathbb{R}}$ be measurable. For $n \in \mathbb{N}$, define the truncation

$$f_n(\omega) := \begin{cases} -n & \text{if } f(\omega) < -n \\ f(\omega) & \text{if } -n \leq f(\omega) \leq n \\ n & \text{if } f(\omega) > n \end{cases}$$

(a) Show that $f_n : \Omega \to \mathbb{R}$ is measurable for every $n \in \mathbb{N}$.

(b) Show that $|f_n(\omega)| \leq n$ for all $\omega \in \Omega$.

(c) Show that if $f(\omega) \in \mathbb{R}$, then $f_n(\omega) \to f(\omega)$ as $n \to \infty$.

(d) Describe the behavior of $f_n(\omega)$ when $f(\omega) = \infty$ and when $f(\omega) = -\infty$. ◁

Proof. (a) We employ Theorem 3.4 to check the preimages of the generator

$$\mathcal{A} = \{[a, b] : a, b \in \mathbb{R},\ a \leq b\}$$

of $\mathcal{B}(\mathbb{R})$. Let $a, b \in \mathbb{R}$ with $a \leq b$. Since $f_n(\omega) \in [-n, n]$ for every $\omega \in \Omega$, we have

$$f_n^{-1}([a, b]) = \begin{cases} \varnothing & \text{if } b < -n \text{ or } a > n \\ \Omega & \text{if } a \leq -n \text{ and } b \geq n \\ f^{-1}([-\infty, b]) & \text{if } a \leq -n \leq b < n \\ f^{-1}([a, b]) & \text{if } -n < a \leq b < n \\ f^{-1}([a, \infty]) & \text{if } -n < a \leq n \leq b \end{cases}$$

In every case, $f_n^{-1}([a, b]) \in \mathcal{F}$ because f is $(\mathcal{F}, \mathcal{B}(\overline{\mathbb{R}}))$-measurable. Hence f_n is measurable.

(b) By definition of f_n, we always have $-n \leq f_n(\omega) \leq n$. Therefore $|f_n(\omega)| \leq n$ for all $\omega \in \Omega$.

(c) Assume that $f(\omega) \in \mathbb{R}$. Choose $N \in \mathbb{N}$ with $N \geq |f(\omega)|$, e.g. $N = \lceil |f(\omega)| \rceil$. Then for every $n \geq N$ we have

$$-n \leq f(\omega) \leq n$$

so by the definition of the truncation, $f_n(\omega) = f(\omega)$ for every $n \geq N$. Hence $f_n(\omega) \to f(\omega)$ as $n \to \infty$.

(d) If $f(\omega) = \infty$, then $f(\omega) > n$ $\forall n \in \mathbb{N}$, hence $f_n(\omega) = n$ $\forall n$. Thus $f_n(\omega) \to \infty$.

If $f(\omega) = -\infty$, then $f(\omega) < -n$ $\forall n \in \mathbb{N}$, hence $f_n(\omega) = -n$ $\forall n$. Thus $f_n(\omega) \to -\infty$. □

Finally, we note that for functions of several variables, separate measurability (i.e. while keeping all but one variable fixed) does in general not imply joint measurability. A nice counterexample is the following:

Example 3.6. Let $E \subset [0, 1]$ be such that $E \notin \mathcal{B}([0, 1])$. Define $f : [0, 1]^2 \to \mathbb{R}$ by

$$f(x, y) := \begin{cases} 1 & \text{if } x = y \text{ and } x \in E \\ 0 & \text{otherwise} \end{cases}$$

or equivalently, $f(x, y) = \mathbf{1}_{\{(x,x):x \in E\}}(x, y)$.

We show first that f is separately measurable, i.e.

- for every fixed $x \in [0, 1]$ the function

$$f(x, \cdot) : [0, 1] \to \mathbb{R}, \qquad y \mapsto f(x, y)$$

  is $(\mathcal{B}([0, 1]), \mathcal{B}(\mathbb{R}))$-measurable

- for every fixed $y \in [0, 1]$ the function

$$f(\cdot, y) : [0, 1] \to \mathbb{R}, \qquad x \mapsto f(x, y)$$

  is $(\mathcal{B}([0, 1]), \mathcal{B}(\mathbb{R}))$-measurable

- Fix $x \in [0,1]$ and let $B \in \mathcal{B}(\mathbb{R})$. Then

$$\{y \in [0,1] : f(x,y) \in B\} = \begin{cases} \begin{cases} [0,1] & \text{if } 0 \in B \\ \varnothing & \text{if } 0 \notin B \end{cases} & \text{if } x \notin E \\ \\ \begin{cases} [0,1] & \text{if } 0,1 \in B \\ [0,1] \setminus \{x\} & \text{if } 0 \in B,\ 1 \notin B \\ \{x\} & \text{if } 0 \notin B,\ 1 \in B \\ \varnothing & \text{if } 0,1 \notin B \end{cases} & \text{if } x \in E \end{cases}$$

In every case, this set belongs to $\mathcal{B}([0,1])$. Hence $f(x,\cdot) : [0,1] \to \mathbb{R}$ is $(\mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable for every fixed $x \in [0,1]$.

- Fix $y \in [0,1]$ and let $B \in \mathcal{B}(\mathbb{R})$. Then

$$\{x \in [0,1] : f(x,y) \in B\} = \begin{cases} \begin{cases} [0,1] & \text{if } 0 \in B \\ \varnothing & \text{if } 0 \notin B \end{cases} & \text{if } y \notin E \\ \\ \begin{cases} [0,1] & \text{if } 0,1 \in B \\ [0,1] \setminus \{y\} & \text{if } 0 \in B,\ 1 \notin B \\ \{y\} & \text{if } 0 \notin B,\ 1 \in B \\ \varnothing & \text{if } 0,1 \notin B \end{cases} & \text{if } y \in E \end{cases}$$

In every case, this set belongs to $\mathcal{B}([0,1])$. Hence $f(\cdot,y) : [0,1] \to \mathbb{R}$ is $(\mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable for every fixed $y \in [0,1]$.

However, f is not jointly measurable, i.e. not $(\mathcal{B}([0,1]) \otimes \mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable.

Indeed, consider the diagonal map

$$\begin{aligned} d : [0,1] &\longrightarrow [0,1]^2 \\ x &\longmapsto d(x) := (x,x) \end{aligned}$$

The function d is continuous, hence $(\mathcal{B}([0,1]), \mathcal{B}([0,1]^2))$-measurable by Corollary 3.5. Moreover, since we have

$$\mathcal{B}([0,1]^2) = \mathcal{B}([0,1]) \otimes \mathcal{B}([0,1])$$

d is also $(\mathcal{B}([0,1]), \mathcal{B}([0,1]) \otimes \mathcal{B}([0,1]))$-measurable.

If f were jointly measurable, i.e. $(\mathcal{B}([0,1]) \otimes \mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable, then by Theorem 3.3, the composition

$$f \circ d : [0,1] \to \mathbb{R}$$

would be $(\mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable. But

$$(f \circ d)(x) = f(d(x)) = f(x,x) = \mathbf{1}_E(x)$$

Hence $\mathbf{1}_E : [0,1] \to \mathbb{R}$ would be $(\mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable. By Theorem 3.1, this would imply $E \in \mathcal{B}([0,1])$, contradicting the choice of E.

Therefore f is separately $(\mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable in each variable, but not jointly $(\mathcal{B}([0,1]) \otimes \mathcal{B}([0,1]), \mathcal{B}(\mathbb{R}))$-measurable. ◀

4 Independence

4.1 Independent Random Variables

Let A be an event with probability $\mathbb{P}(A)$. If we know that another event B has occurred, we update our information on A by replacing $\mathbb{P}(A)$ with the *conditional probability*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

provided $\mathbb{P}(B) > 0$. Note that an event may occur and still have probability 0, so this assumption is necessary.

Independence of A and B means that this update has no effect, i.e. $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ or equivalently $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Hence, the occurence of B does not change the likelihood of A.

Definition 4.1 (Independence of events)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Two events (see Definition 1.1) $A, B \in \mathcal{F}$ are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. ◀

Example 4.1.

- always independent: Ω and $\varnothing$ are independent of any event $A \in \mathcal{F}$, since

$$\mathbb{P}(\Omega \cap A) = \mathbb{P}(A) = \mathbb{P}(\Omega)\mathbb{P}(A) \quad \text{and} \quad \mathbb{P}(\varnothing \cap A) = 0 = \mathbb{P}(\varnothing)\mathbb{P}(A)$$

- never independent: if $0 < \mathbb{P}(A) < 1$, then A and $A^{\complement}$ are not independent, since

$$\mathbb{P}(A \cap A^{\complement}) = 0 \neq \mathbb{P}(A)\mathbb{P}(A^{\complement})$$ ◀

Definition 4.2 (Independence of random variables)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X, Y : \Omega \to \mathbb{R}$ be random variables (see Definition 3.2). We say that X and Y are *independent* if the events $\{X \in A\}$ and $\{Y \in B\}$ are independent for all $A, B \in \mathcal{B}(\mathbb{R})$, i.e., if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

for all $A, B \in \mathcal{B}(\mathbb{R})$. ◀

It is often convenient to rephrase independence in terms of generated σ-algebras.

Definition 4.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable. We call

$$\boxed{\sigma(X) := \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\} \subseteq \mathcal{F}}$$

the *σ-algebra generated by X*. It encodes the information carried by X. ◀

Definition 4.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{F}_1, \mathcal{F}_2 \subseteq \mathcal{F}$ be two σ-algebras. We say that $\mathcal{F}_1$ and $\mathcal{F}_2$ are *independent* if

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$$

for all $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$. ◀

Theorem 4.1. Two random variables X and Y are independent if and only if the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent. ◁

The independence of random variables is already guaranteed, if they are independent on a semiring (Definition 2.5) generating the Borel σ-algebra (Definition 2.2).

Theorem 4.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X, Y : \Omega \to \mathbb{R}$ be random variables. Suppose that the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by a semiring $\mathcal{S} \subseteq 2^{\mathbb{R}}$, i.e. $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$. If

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

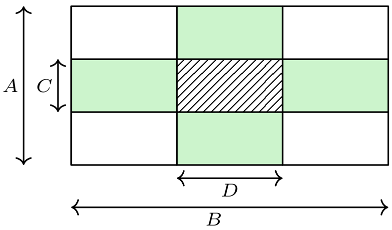
for all $A, B \in \mathcal{S}$, then X and Y are independent. ◁

Proof. We start by showing that the product $\mathcal{S} \times \mathcal{S} := \{S_1 \times S_2 : S_1, S_2 \in \mathcal{S}\}$ is again a semiring (Definition 2.5):

- $\varnothing = \varnothing \times \varnothing \in \mathcal{S} \times \mathcal{S}$. ✓

- If $A_1 \times A_2, B_1 \times B_2 \in \mathcal{S} \times \mathcal{S}$, then
$$(A_1 \times A_2) \cap (B_1 \times B_2) \stackrel{(0.5)}{=} (A_1 \cap B_1) \times (A_2 \cap B_2)$$
and since $\mathcal{S}$ is a semiring, $A_1 \cap B_1 \in \mathcal{S}$ and $A_2 \cap B_2 \in \mathcal{S}$. Hence $(A_1 \times A_2) \cap (B_1 \times B_2) \in \mathcal{S} \times \mathcal{S}$. ✓
- Let $A \times B, C \times D \in \mathcal{S} \times \mathcal{S}$ with $C \times D \subseteq A \times B$. If $C = \varnothing$ or $D = \varnothing$, then $C \times D = \varnothing$, hence
$$(A \times B) \setminus (C \times D) = A \times B \in \mathcal{S} \times \mathcal{S}.$$
Otherwise, choose $a_0 \in C$ and $b_0 \in D$. Then for every $a \in C$ we have $(a, b_0) \in C \times D \subseteq A \times B$, so $a \in A$. Hence $C \subseteq A$. Similarly, for every $b \in D$ we have $(a_0, b) \in C \times D \subseteq A \times B$, so $b \in B$. Hence $D \subseteq B$.

  Since $\mathcal{S}$ is a semiring, there exist pairwise disjoint sets $A_1, \dots, A_m \in \mathcal{S}$ and $B_1, \dots, B_n \in \mathcal{S}$ such that
$$A \setminus C = \bigsqcup_{i=1}^{m} A_i \qquad \text{and} \qquad B \setminus D = \bigsqcup_{j=1}^{n} B_j$$



  Therefore,
$$\begin{aligned}(A \times B) \setminus (C \times D) &= ((A \setminus C) \times (B \setminus D)) \sqcup (C \times (B \setminus D)) \sqcup ((A \setminus C) \times D) \\ &= \left(\bigsqcup_{i=1}^{m} \bigsqcup_{j=1}^{n} (A_i \times B_j)\right) \sqcup \left(\bigsqcup_{j=1}^{n} (C \times B_j)\right) \sqcup \left(\bigsqcup_{i=1}^{m} (A_i \times D)\right)\end{aligned}$$
which is a finite disjoint union of sets in $\mathcal{S} \times \mathcal{S}$. ✓

Thus $\mathcal{S} \times \mathcal{S}$ is a semiring.

Moreover, we have $\sigma(\mathcal{S} \times \mathcal{S}) = \sigma(\mathcal{S}) \otimes \sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$, where $\otimes$ denotes the product σ-algebra (see Example 1.4):

$\subseteq$: Every set $S_1 \times S_2$ with $S_1, S_2 \in \mathcal{S}$ belongs to $\sigma(\mathcal{S}) \otimes \sigma(\mathcal{S})$, since $\mathcal{S} \subseteq \sigma(\mathcal{S})$. Hence, by the minimality of the generated σ-algebra (Theorem 1.3), $\sigma(\mathcal{S} \times \mathcal{S}) \subseteq \sigma(\mathcal{S}) \otimes \sigma(\mathcal{S})$.

$\supseteq$: For the converse inclusion, it is enough to show that $A \times B \in \sigma(\mathcal{S} \times \mathcal{S})$ for all $A, B \in \sigma(\mathcal{S})$, i.e.
$$\{A \times B : A, B \in \sigma(\mathcal{S})\} \subseteq \sigma(\mathcal{S} \times \mathcal{S})$$
because $\sigma(\mathcal{S}) \otimes \sigma(\mathcal{S})$ is generated by such rectangles (Example 1.4).

Fix $B \in \mathcal{S}$ and define
$$\mathcal{D}_B := \{A \subseteq \mathbb{R} : A \times B \in \sigma(\mathcal{S} \times \mathcal{S})\}$$
We show that $\mathcal{D}_B$ is a σ-algebra:

- First, $\varnothing \in \mathcal{D}_B$, since $\varnothing \times B = \varnothing \in \sigma(\mathcal{S} \times \mathcal{S})$. ✓
- Next, let $A \in \mathcal{D}_B$. By the countable covering assumption from the σ-finiteness part, choose $S_m \in \mathcal{S}$ such that $\mathbb{R} = \bigcup_{m=1}^{\infty} S_m$. Then
$$\mathbb{R} \times B = \left(\bigcup_{m=1}^{\infty} S_m\right) \times B = \bigcup_{m=1}^{\infty} (S_m \times B) \in \sigma(\mathcal{S} \times \mathcal{S})$$
because $S_m \times B \in \mathcal{S} \times \mathcal{S}$ for all $m \in \mathbb{N}$. Since $A \times B \in \sigma(\mathcal{S} \times \mathcal{S})$, we obtain
$$A^{\complement} \times B = (\mathbb{R} \setminus A) \times B = (\mathbb{R} \times B) \setminus (A \times B) \in \sigma(\mathcal{S} \times \mathcal{S})$$
and hence $A^{\complement} \in \mathcal{D}_B$. ✓

- Finally, if $A_n \in \mathcal{D}_B$ for all $n \in \mathbb{N}$, then

$$\left(\bigcup_{n=1}^{\infty} A_n\right) \times B = \bigcup_{n=1}^{\infty} (A_n \times B) \in \sigma(\mathcal{S} \times \mathcal{S})$$

so $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}_B$. ✓

Thus $\mathcal{D}_B$ is a σ-algebra.

Moreover, $\mathcal{S} \subseteq \mathcal{D}_B$, because $A \times B \in \mathcal{S} \times \mathcal{S} \subseteq \sigma(\mathcal{S} \times \mathcal{S})$ for every $A \in \mathcal{S}$.

Therefore, by the minimality of $\sigma(\mathcal{S})$ (Theorem 1.3), we have $\sigma(\mathcal{S}) \subseteq \mathcal{D}_B$.

That means

$$A \times B \in \sigma(\mathcal{S} \times \mathcal{S}) \qquad (\circledast)$$

for all $A \in \sigma(\mathcal{S})$ and all $B \in \mathcal{S}$.

Now fix $A \in \sigma(\mathcal{S})$ and define

$$\mathcal{D}_A := \{B \subseteq \mathbb{R} : A \times B \in \sigma(\mathcal{S} \times \mathcal{S})\}$$

which is again a σ-algebra by the same argument as for $\mathcal{D}_B$.

By $\circledast$, $\mathcal{S} \subseteq \mathcal{D}_A$: if $B \in \mathcal{S}$, then $\circledast$, applied to the fixed $A \in \sigma(\mathcal{S})$, gives $A \times B \in \sigma(\mathcal{S} \times \mathcal{S})$, hence $B \in \mathcal{D}_A$.

Thus, again by the minimality of $\sigma(\mathcal{S})$ (Theorem 1.3), we get $\sigma(\mathcal{S}) \subseteq \mathcal{D}_A$.

Therefore

$$A \times B \in \sigma(\mathcal{S} \times \mathcal{S})$$

for all $A, B \in \sigma(\mathcal{S})$.

Hence every generating rectangle of $\sigma(\mathcal{S}) \otimes \sigma(\mathcal{S})$ belongs to $\sigma(\mathcal{S} \times \mathcal{S})$, and so

$$\sigma(\mathcal{S}) \otimes \sigma(\mathcal{S}) \subseteq \sigma(\mathcal{S} \times \mathcal{S})$$

Since by assumption $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$, we conclude that

$$\sigma(\mathcal{S} \times \mathcal{S}) = \sigma(\mathcal{S}) \otimes \sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$$

as claimed.

We define the set functions $\mu, \mu_{\text{prod}} : \mathcal{S} \times \mathcal{S} \to [0, 1]$ by

$$\mu(A \times B) := \mathbb{P}(X \in A, Y \in B) \qquad \mu_{\text{prod}}(A \times B) := \mathbb{P}_X(A)\mathbb{P}_Y(B)$$

where the image measure $\mathbb{P}_Z$ is defined as

$$\mathbb{P}_Z(C) := (\mathbb{P} \circ Z^{-1})(C) = \mathbb{P}(Z^{-1}(C)) = \mathbb{P}(Z \in C)$$

for any random variable $Z : \Omega \to \mathbb{R}$ and any $C \in \mathcal{B}(\mathbb{R})$ (see Definition 3.4).

By assumption we have $\mu(A \times B) = \mu_{\text{prod}}(A \times B)$ for all $A, B \in \mathcal{S}$. Both set functions μ and μ_{prod} satisfy the assumptions of the Carathéodory extension Theorem 2.5:

- $\mu(\varnothing) = \mu_{\text{prod}}(\varnothing) = 0$. ✓
- (i): Let $A_1 \times B_1, \ldots, A_n \times B_n$ be pairwise disjoint sets in $\mathcal{S} \times \mathcal{S}$ such that their union is in $\mathcal{S} \times \mathcal{S}$. We can write

$$\bigsqcup_{k=1}^{n} (A_k \times B_k) = A \times B$$

since we know that the union on the left must be in $\mathcal{S} \times \mathcal{S}$. Then

$$\begin{aligned}\{\omega \in \Omega : X(\omega) \in A, Y(\omega) \in B\} &= \{\omega \in \Omega : X(\omega) \in A \wedge Y(\omega) \in B\} \\ &= \{\omega \in \Omega : (X(\omega), Y(\omega)) \in A \times B\} \\ &= \left\{\omega \in \Omega : (X(\omega), Y(\omega)) \in \bigsqcup_{k=1}^{n} (A_k \times B_k)\right\} \\ &= \left\{\omega \in \Omega : \bigvee_{k=1}^{n} ((X(\omega), Y(\omega)) \in A_k \times B_k)\right\}\end{aligned}$$

$$= \bigsqcup_{k=1}^{n} \{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_k \times B_k\}$$

where the last union is indeed disjoint, since, for $i \neq j$,

$$\begin{aligned}
&\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_i \times B_i\} \cap \{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_j \times B_j\} \\
&= \{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_i \times B_i \wedge (X(\omega), Y(\omega)) \in A_j \times B_j\} \\
&= \{\omega \in \Omega : (X(\omega), Y(\omega)) \in (A_i \times B_i) \cap (A_j \times B_j)\} \\
&= \{\omega \in \Omega : (X(\omega), Y(\omega)) \in \varnothing\} \\
&= \varnothing
\end{aligned}$$

Hence,

$$\begin{aligned}
\mu(A \times B) &= \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A, Y(\omega) \in B\}\right) \\
&= \mathbb{P}\left(\bigsqcup_{k=1}^{n} \{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_k \times B_k\}\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in A_k \times B_k\}\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A_k, Y(\omega) \in B_k\}\right) \\
&= \sum_{k=1}^{n} \mu(A_k \times B_k)
\end{aligned}$$

For μ_{prod}, let $\nu := \mathbb{P}_X \otimes \mathbb{P}_Y$ be the product measure (see Example 2.5) on $\sigma(\mathcal{S} \times \mathcal{S})$. On rectangles $A \times B \in \mathcal{S} \times \mathcal{S}$, we have by definition

$$\nu(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B) = \mu_{\text{prod}}(A \times B)$$

Since ν is a measure (Definition 1.5), it is σ-additive, hence in particular finitely additive, hence

$$\begin{aligned}
\mu_{\text{prod}}(A \times B) &= \nu(A \times B) \\
&= \nu\left(\bigsqcup_{k=1}^{n} (A_k \times B_k)\right) \\
&= \sum_{k=1}^{n} \nu(A_k \times B_k) \\
&= \sum_{k=1}^{n} \mu_{\text{prod}}(A_k \times B_k)
\end{aligned}$$

Thus μ_{prod} is additive as well. ✓

- (ii): Let $A \times B \in \mathcal{S} \times \mathcal{S}$ and $(A_i \times B_i)_{i \in \mathbb{N}}$ with $A_i \times B_i \in \mathcal{S} \times \mathcal{S}$ for all i be such that

$$A \times B \subseteq \bigcup_{k=1}^{\infty} (A_k \times B_k)$$

For μ, this implies

$$\begin{aligned}
\{\omega : X(\omega) \in A, Y(\omega) \in B\} &= \{\omega : (X(\omega), Y(\omega)) \in A \times B\} \\
&\subseteq \left\{\omega : (X(\omega), Y(\omega)) \in \bigcup_{k=1}^{\infty} (A_k \times B_k)\right\} \\
&= \left\{\omega : \bigvee_{k=1}^{\infty} ((X(\omega), Y(\omega)) \in A_k \times B_k)\right\} \\
&= \bigcup_{k=1}^{\infty} \{\omega : (X(\omega), Y(\omega)) \in A_k \times B_k\} \\
&= \bigcup_{k=1}^{\infty} \{\omega : X(\omega) \in A_k, Y(\omega) \in B_k\}
\end{aligned}$$

Hence, by the σ-subadditivity of the probability measure $\mathbb{P}$,

$$\begin{aligned}\mu(A \times B) &= \mathbb{P}(X \in A, Y \in B) \\ &\overset{(1.3)}{\leq} \mathbb{P}\left(\bigcup_{k=1}^{\infty}\{X \in A_k, Y \in B_k\}\right) \\ &\overset{(1.5)}{\leq} \sum_{k=1}^{\infty} \mathbb{P}(X \in A_k, Y \in B_k) \\ &= \sum_{k=1}^{\infty} \mu(A_k \times B_k)\end{aligned}$$

Thus μ is σ-subadditive.

For μ_{prod}, we again use $\nu := \mathbb{P}_X \otimes \mathbb{P}_Y$. Since ν is a measure and $\mu_{\text{prod}} = \nu$ on rectangles in $\mathcal{S} \times \mathcal{S}$, we obtain

$$\begin{aligned}\mu_{\text{prod}}(A \times B) &= \nu(A \times B) \\ &\overset{(1.3)}{\leq} \nu\left(\bigcup_{k=1}^{\infty}(A_k \times B_k)\right) \\ &\overset{(1.5)}{\leq} \sum_{k=1}^{\infty} \nu(A_k \times B_k) \\ &= \sum_{k=1}^{\infty} \mu_{\text{prod}}(A_k \times B_k)\end{aligned}$$

Thus μ_{prod} is σ-subadditive as well. ✓

- (iii): For simplicity[3], assume the semiring $\mathcal{S}$ covers $\mathbb{R}$, i.e.

$$\mathbb{R} = \bigcup_{m=1}^{\infty} S_m$$

for some $(S_m)_{m \in \mathbb{N}}$ with $S_m \in \mathcal{S}$ for all m. Hence

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \left(\bigcup_{m=1}^{\infty} S_m\right) \times \left(\bigcup_{n=1}^{\infty} S_n\right) = \bigcup_{m=1}^{\infty}\bigcup_{n=1}^{\infty}(S_m \times S_n)$$

and since $\mathbb{N} \times \mathbb{N}$ is countable, this is a countable cover of $\mathbb{R}^2$ by sets in $\mathcal{S} \times \mathcal{S}$.

Moreover, for every $m, n \in \mathbb{N}$,

$$0 \leq \mu(S_m \times S_n) = \mathbb{P}(X \in S_m, Y \in S_n) \leq 1$$

and

$$0 \leq \mu_{\text{prod}}(S_m \times S_n) = \mathbb{P}_X(S_m)\mathbb{P}_Y(S_n) \leq 1$$

Thus each covering set has finite μ- and μ_{prod}-mass. Therefore both μ and μ_{prod} are σ-finite. ✓

Therefore, by Theorem 2.5, the extensions $\tilde{\mu}, \tilde{\mu}_{\text{prod}}$ of μ, μ_{prod} to $\sigma(\mathcal{S} \times \mathcal{S})$ are unique.

We conclude

$$\mathbb{P}(X \in A, Y \in B) = \tilde{\mu}(A \times B) = \tilde{\mu}_{\text{prod}}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B)$$

for all $A, B \in \mathcal{B}(\mathbb{R})$, because $A \times B \in \mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{S} \times \mathcal{S})$ and both extensions agree on all of $\sigma(\mathcal{S} \times \mathcal{S})$ by uniqueness. Hence X and Y are independent. □

Corollary 4.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X, Y : \Omega \to \mathbb{R}$ be random variables. We define $F_{X,Y}(x, y) := \mathbb{P}(X \leq x, Y \leq y)$, $F_X(x) := \mathbb{P}(X \leq x)$, $F_Y(y) := \mathbb{P}(Y \leq y)$. Then X and Y are independent if and only if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

for all $x, y \in \mathbb{R}$. ◁

[3] for our proof strategy via semirings, it is necessary

Example 4.2 (Independent Gaussian random variables)**.** It is well known that two zero-mean Gaussian random variables X and Y with zero correlation are independent. Without loss of generality, let us assume that X and Y have zero mean. By "zero correlation", we mean that $\mathbb{E}[XY] = 0$. We have

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y) \,\mathrm{d}x \,\mathrm{d}y$$

with the density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right)$$

This density can be factorized into the product of two Gaussian densities

$$f_{X,Y}(x,y) = \left(\frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/(2\sigma_x^2)}\right) \left(\frac{1}{\sqrt{2\pi}\sigma_y} e^{-y^2/(2\sigma_y^2)}\right) =: f_X(x) f_Y(y)$$

Therefore, we have

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y) \,\mathrm{d}x \,\mathrm{d}y = \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(x) f_Y(y) \,\mathrm{d}x \,\mathrm{d}y \\ &= \int_{-\infty}^{x} f_X(x) \,\mathrm{d}x \int_{-\infty}^{y} f_Y(y) \,\mathrm{d}y = F_X(x) F_Y(y) \end{aligned}$$

which shows by Corollary 4.3 that X and Y are independent.

While uncorrelatedness is necessary for independence, it is a much weaker condition in general. When the joint probability distribution is not Gaussian, two random variables can be uncorrelated while still not being independent. ◀

An important property of independence is that it is stable under measurable transformations.

Theorem 4.4 (Independence under measurable transformations)**.** Let $X, Y : \Omega \to \mathbb{R}$ be independent random variables (Definition 3.2). Let $f, g : \mathbb{R} \to \mathbb{R}$ be $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$-measurable functions (Definition 3.1). Then $f(X)$ and $g(Y)$ are independent random variables. ◁

Proof. By Theorem 3.3, the functions $f(X)$ and $g(Y)$ are measurable, hence random variables. To prove that they are independent, it suffices, according to Definition 4.2, to show that for any two sets $A, B \in \mathcal{B}(\mathbb{R})$, the events $\{f(X) \in A\}$ and $\{g(Y) \in B\}$ are independent.

Since f and g are measurable, $f^{-1}(A)$ and $g^{-1}(B)$ are Borel sets. We note that, by Fact 3.2.(iii),

$$\{h(Z) \in C\} = (h \circ Z)^{-1}(C) = Z^{-1}(h^{-1}(C)) = \{Z \in h^{-1}(C)\}$$

and hence $\{f(X) \in A\} = \{X \in f^{-1}(A)\}$ and $\{g(Y) \in B\} = \{Y \in g^{-1}(B)\}$. By the independence of X and Y, the events $\{X \in f^{-1}(A)\}$ and $\{Y \in g^{-1}(B)\}$ are independent. Thus $\{f(X) \in A\}$ and $\{g(Y) \in B\}$ are independent. □

Exercise 4.3 (Independence under derived observables)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X, Y : \Omega \to \mathbb{R}$ be independent random variables, i.e. $X \perp Y$ (Definition 4.2).

(a) Let $Z : \Omega \to \mathbb{R}$ be a random variable and $h : \mathbb{R} \to \mathbb{R}$ be a measurable function (Definition 3.1). Show that

$$\sigma(h(Z)) \subseteq \sigma(Z)$$

(b) Let $f, g : \mathbb{R} \to \mathbb{R}$ be measurable functions. Deduce from the independence of $\sigma(X)$ and $\sigma(Y)$ that $f(X)$ and $g(Y)$ are independent.

(c) Let $A, B \in \mathcal{B}(\mathbb{R})$. Show that the Bernoulli random variables

$$U := \mathbf{1}_{\{X \in A\}}, \qquad V := \mathbf{1}_{\{Y \in B\}}$$

are independent. ✍

Solution.

(a) Let $C \in \sigma(h(Z))$. Then, by Definition 4.3, there exists $D \in \mathcal{B}(\mathbb{R})$ such that

$$C = (h(Z))^{-1}(D)$$

By Fact 3.2.(iii),

$$(h(Z))^{-1}(D) = (h \circ Z)^{-1}(D) = Z^{-1}(h^{-1}(D))$$

Since h is measurable,

$$E := h^{-1}(D) \in \mathcal{B}(\mathbb{R})$$

Thus, we can write

$$C = Z^{-1}(h^{-1}(D)) = Z^{-1}(E)$$

with $E \in \mathcal{B}(\mathbb{R})$. Hence $C \in \sigma(Z)$. Since $C \in \sigma(h(Z))$ was arbitrary, we conclude that

$$\sigma(h(Z)) \subseteq \sigma(Z)$$

(b) Since $X \perp Y$, by Theorem 4.1, $\sigma(X)$ and $\sigma(Y)$ are independent in the sense of Definition 4.4, i.e.

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$$

for all $A_1 \in \sigma(X)$ and $A_2 \in \sigma(Y)$. By (a),

$$\sigma(f(X)) \subseteq \sigma(X), \qquad \sigma(g(Y)) \subseteq \sigma(Y)$$

and thus the same holds for all $A_1 \in \sigma(f(X))$ and $A_2 \in \sigma(g(Y))$. Hence, $\sigma(f(X))$ and $\sigma(g(Y))$ are independent in the sense of Definition 4.4, and thus, again by Theorem 4.1, $f(X)$ and $g(Y)$ are independent.

(c) Denote

$$f := \mathbf{1}_A, \qquad g := \mathbf{1}_B$$

Since $A, B \in \mathcal{B}(\mathbb{R})$, the functions $f, g : \mathbb{R} \to \mathbb{R}$ are measurable by Theorem 3.1. Moreover,

$$U = \mathbf{1}_{\{X \in A\}} = \mathbf{1}_A(X) = f(X), \qquad V = \mathbf{1}_{\{Y \in B\}} = \mathbf{1}_B(Y) = g(Y)$$

Indeed, for every $\omega \in \Omega$,

$$U(\omega) = \mathbf{1}_{\{X \in A\}}(\omega) = \mathbf{1}_A(X(\omega)), \qquad V(\omega) = \mathbf{1}_{\{Y \in B\}}(\omega) = \mathbf{1}_B(Y(\omega))$$

Since $X \perp Y$, by (b), $f(X)$ and $g(Y)$ are independent. Hence, $U \perp V$. ◀

Remark 4.4. Theorem 4.4 particularly applies to continuous and piecewise continuous functions and directly extends to random vectors. ◀

We close this section by considering independence of infinitely many random variables.

Definition 4.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{I}$ be an arbitrary index set.

(i) A family of events $(A_i)_{i \in \mathcal{I}}$, where $A_i \in \mathcal{F}$ is *independent* if

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \tag{4.1}$$

for all finite subsets $J \subseteq \mathcal{I}$.

(ii) A family $(\mathcal{A}_i)_{i \in \mathcal{I}}$, where $\mathcal{A}_i \subseteq \mathcal{F}$, is *independent* if the family of events $(A_i)_{i \in \mathcal{I}}$ is independent for every choice of $A_i \in \mathcal{A}_i$. ◀

Remark 4.5. An independent family of events $(A_i)_{i \in \mathcal{I}} \subseteq \mathcal{F}$ in particular is pairwise independent, i.e. $A_i \perp A_j$ for $i \neq j$. The contrary is usually not true. ◀

Example 4.6. We roll a fair die twice and consider the following events:

- A_1: the first roll shows an odd number
- A_2: the second roll shows an odd number
- A_3: the sum of the dice is odd, i.e. $A_3 = A_1 \triangle A_2$

Then, $A_1 \perp A_2$, $A_1 \perp A_3$, $A_2 \perp A_3$, i.e. the events are pairwise independent. However,

$$0 = \mathbb{P}(\varnothing) = \mathbb{P}(A_1 \cap A_2 \cap A_3) \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

i.e. the family A_1, A_2, A_3 is not independent in the sense of Definition 4.5 (i). ◀

Definition 4.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{I}$ be an arbitrary index set. The family $(X_i)_{i\in\mathcal{I}}$ of random variables is *independent* if the family of $(\sigma(X_i))_{i\in\mathcal{I}}$ is independent. ◀

Finally, we remark that it is indeed sufficient to consider only finite intersections of events to define independence. Let $(A_i)_{i\in\mathbb{N}}$ be a sequence of events such that any finite subfamily is independent, and let

$$B_k := \bigcap_{i=1}^{k} A_i$$

Then $(B_k)_{k\in\mathbb{N}}$ is a decreasing sequence with $\bigcap_{i=1}^{\infty} A_i = \bigcap_{k=1}^{\infty} B_k$. By continuity of $\mathbb{P}$ from above (Theorem 1.11), we have

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcap_{k=1}^{\infty} B_k\right) = \lim_{k\to\infty} \mathbb{P}(B_k)$$

Since $A_1, \ldots, A_k$ are independent, we obtain

$$\mathbb{P}(B_k) = \mathbb{P}\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} \mathbb{P}(A_i)$$

and hence

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{k\to\infty} \mathbb{P}(B_k) = \lim_{k\to\infty} \prod_{i=1}^{k} \mathbb{P}(A_i) = \prod_{i=1}^{\infty} \mathbb{P}(A_i)$$

4.2 Borel-Cantelli lemmas

describe when events occur infinitely often.

Theorem 4.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(A_n)_{n\in\mathbb{N}}$, $A_n \in \mathcal{F}$, be a sequence of events. Then

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty \quad \Longrightarrow \quad \mathbb{P}\left(\limsup_{i\to\infty} A_i\right) = 0$$

◁

Proof. We have

$$\limsup_{i\to\infty} A_i = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j = \bigcap_{k=1}^{\infty} B_k$$

with $B_k := \bigcup_{j=k}^{\infty} A_j$ for $k \in \mathbb{N}$. Since $(B_k)_{k\in\mathbb{N}}$ is a decreasing sequence of events, we have

$$\mathbb{P}\left(\limsup_{i\to\infty} A_i\right) = \mathbb{P}\left(\bigcap_{k=1}^{\infty} B_k\right) = \lim_{k\to\infty} \mathbb{P}(B_k)$$

by continuity of $\mathbb{P}$ from above (Theorem 1.11). On the other hand, by the σ-subadditivity of $\mathbb{P}$ (Theorem 1.9), we have

$$\mathbb{P}(B_k) = \mathbb{P}\left(\bigcup_{j=k}^{\infty} A_j\right) \leq \sum_{j=k}^{\infty} \mathbb{P}(A_j)$$

and since the series $\sum_{i=1}^{\infty} \mathbb{P}(A_i)$ converges, the tail $\sum_{j=k}^{\infty} \mathbb{P}(A_j)$ must go to zero as $k \to \infty$. This implies $\mathbb{P}(B_k) \to 0$ as $k \to \infty$, which yields $\mathbb{P}(\limsup_{i\to\infty} A_i) = 0$. □

Theorem 4.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(A_n)_{n\in\mathbb{N}}$, $A_n \in \mathcal{F}$, be a sequence of events that are independent in the sense of Definition 4.5 (i). Then

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty \quad \Longrightarrow \quad \mathbb{P}\left(\limsup_{i\to\infty} A_i\right) = 1$$

◁

Proof. We consider

$$\left(\limsup_{i\to\infty} A_i\right)^{\complement} = \left(\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty} A_j\right)^{\complement} \overset{(0.7,0.8)}{=} \bigcup_{k=1}^{\infty}\bigcap_{j=k}^{\infty} A_j^{\complement}$$

and show that

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty}\bigcap_{j=k}^{\infty} A_j^{\complement}\right) = 0$$

By Theorem 1.9, it suffices to prove that $\mathbb{P}(\bigcap_{j=k}^{\infty} A_j^{\complement}) = 0$ for all $k \in \mathbb{N}$.

We use the inequality $1 - x \leq e^{-x}$, which holds for all $x \in \mathbb{R}$. Fix $k \in \mathbb{N}$ and let $k < m \in \mathbb{N}$. We have

$$\mathbb{P}\left(\bigcap_{j=k}^{m} A_j^{\complement}\right) \overset{(4.1)}{=} \prod_{j=k}^{m} \mathbb{P}\left(A_j^{\complement}\right) = \prod_{j=k}^{m}(1 - \mathbb{P}(A_j)) \leq \prod_{j=k}^{m} e^{-\mathbb{P}(A_j)} = e^{-\sum_{j=k}^{m}\mathbb{P}(A_j)}$$

where we used the fact that independence is preserved under taking complements.

Since $\sum_{i=1}^{\infty}\mathbb{P}(A_i) = \infty$, the tail $\sum_{j=k}^{m}\mathbb{P}(A_j)$ must go to infinity as $m \to \infty$ and hence

$$e^{-\sum_{j=k}^{m}\mathbb{P}(A_j)} \to 0$$

as $m \to \infty$. By continuity of $\mathbb{P}$ from above (Theorem 1.11), we have

$$\mathbb{P}\left(\bigcap_{j=k}^{\infty} A_j^{\complement}\right) = \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{j=k}^{m} A_j^{\complement}\right) \leq \lim_{m\to\infty} e^{-\sum_{j=k}^{m}\mathbb{P}(A_j)} = 0$$

for all $k \in \mathbb{N}$. □

Example 4.7 (Monkey typing Shakespeare)**.** Let $\mathcal{A}$ be a finite alphabet and $w = a_1 \cdots a_n$ a fixed word of length n with $a_i \in \mathcal{A}$ for all i. Consider an infinite random string $(X_i)_{i\in\mathbb{N}}$ where the X_i are independent and uniformly distributed on $\mathcal{A}$. Then, with probability 1, the word w appears infinitely often as a consecutive block in the infinite string.

To see this, partition the infinite string into infinitely many disjoint blocks of length n. For $k \in \mathbb{N}$, let A_k be the event that the k-th block (i.e. the positions $(k-1)n+1, \ldots, kn$) is exactly equal to the word w. Since the blocks do not overlap, the events $(A_k)_{k\geq 1}$ are independent. Moreover,

$$\mathbb{P}(A_k) = \prod_{i=1}^{n}\mathbb{P}(X_{(k-1)n+i} = a_i) = \left(\frac{1}{|\mathcal{A}|}\right)^n > 0$$

for every $k \in \mathbb{N}$. Hence

$$\sum_{k=1}^{\infty}\mathbb{P}(A_k) = \sum_{k=1}^{\infty}\left(\frac{1}{|\mathcal{A}|}\right)^n = \infty$$

By Theorem 4.6, this implies

$$\mathbb{P}\left(\limsup_{k\to\infty} A_k\right) = 1$$

Thus infinitely many of the blocks are equal to w with probability 1. In particular, the word w appears at least once in the infinite string in n consecutive positions with probability 1.

Let w be the complete works of Shakespeare and consider a monkey randomly typing on a typewriter one character at a time. Theorem 4.6 then states that, given infinite time, the monkey will reproduce the works of Shakespeare infinitely often with probability 1. ◀

Combining Theorem 4.5 and Theorem 4.6 yields Kolmogorov's zero-one law for independent events.

Theorem 4.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(A_n)_{n\in\mathbb{N}}$ be a sequence of independent events. Then

$$\mathbb{P}\left(\limsup_{i\to\infty} A_i\right) = \begin{cases} 1 & \text{if } \sum_{i=1}^{\infty}\mathbb{P}(A_i) = \infty \\ 0 & \text{if } \sum_{i=1}^{\infty}\mathbb{P}(A_i) < \infty \end{cases}$$ ◁

Similarly to Definition 4.3, we can define the σ-algebra generated by a collection of random variables:

Definition 4.7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{X_i : i \in I\}$ be a collection of random variables, where I is an arbitrary, possibly countably or uncountably infinite index set. We call

$$\boxed{\sigma(\{X_i : i \in I\}) \coloneqq \sigma\big(\{X_i^{-1}(B) : i \in I, B \in \mathcal{B}(\mathbb{R})\}\big) \subseteq \mathcal{F}}$$

the *σ-algebra generated by $\{X_i : i \in I\}$*, sometimes abbreviated $\sigma(X_i : i \in I)$. ◀

Remark 4.8. Unlike for single random variables (Definition 4.3), the outer $\sigma(\cdot)$ is necessary here because the collection

$$\{X_i^{-1}(B) : i \in I,\ \ B \in \mathcal{B}(\mathbb{R})\}$$

is generally not a σ-algebra itself: it contains events depending on one random variable at a time, but not necessarily events involving several random variables simultaneously. In contrast, for a single random variable X, the collection

$$\{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$$

is already a σ-algebra, since, by Fact 3.2, preimages preserve the set operations required for a σ-algebra (Definition 1.3). ◀

Exercise 4.9 (Borel–Cantelli and almost sure behavior of maxima)**.** Let $(X_n)_{n\in\mathbb{N}}$ be independent real-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Assume that

$$\mathbb{P}(X_n > n) = \frac{1}{n^2}$$

for all $n \in \mathbb{N}$. Define

$$M_n \coloneqq \max\{X_1, \dots, X_n\}$$

for $n \in \mathbb{N}$.

(a) Show that the event

$$A \coloneqq \{\omega \in \Omega : X_n(\omega) > n \text{ i.o.}\}$$

has probability zero.

(b) Show that, with probability one, there exists $N(\omega)$ such that

$$X_n(\omega) \le n$$

for all $n \ge N(\omega)$.

(c) Show that, $\mathbb{P}$-almost surely,

$$\frac{M_n}{n} \le 1$$

eventually.

(d) Interpret this result in terms of the growth of the sequence $(X_n)_{n\in\mathbb{N}}$. ✍

Solution.

(a) Define the events

$$A_n \coloneqq \{\omega \in \Omega : X_n(\omega) > n\}$$

for $n \in \mathbb{N}$. Then

$$A = \limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

By assumption,

$$\mathbb{P}(A_n) = \mathbb{P}(X_n > n) = \frac{1}{n^2}$$

Hence

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \zeta(2) = \frac{\pi^2}{6} < \infty$$

where $\zeta(x) \coloneqq \sum_{n=1}^{\infty} n^{-x}$ is the Riemann zeta function, defined for $x > 1$. Thus, by Theorem 4.5,

$$\mathbb{P}(A) = \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0$$

(b) Observe that

$$\begin{aligned} A^{\complement} &= \{\omega \in \Omega : X_n(\omega) > n \text{ only finitely often}\} \\ &= \{\omega \in \Omega : X_n(\omega) \le n \text{ eventually}\} \\ &= \{\omega \in \Omega : \exists N(\omega) \in \mathbb{N} \text{ such that } X_n(\omega) \le n \text{ for all } n \ge N(\omega)\} \end{aligned}$$

From (a), we know that $\mathbb{P}(A) = 0$, hence

$$\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A) = 1$$

Therefore, for almost every $\omega \in \Omega$, there exists $N(\omega)$ such that for all $n \ge N(\omega)$,

$$X_n(\omega) \le n$$

(c) Fix $\omega \in A^{\complement}$. By (b), there exists $N(\omega) \in \mathbb{N}$ such that for all $n \ge N(\omega)$,

$$X_n(\omega) \le n$$

Since the random variables are real-valued, i.e. $-\infty < X_i(\omega) < \infty$ for all $i \in \mathbb{N}$, define

$$C(\omega) := \max\{0, X_1(\omega), \ldots, X_{N(\omega)-1}(\omega)\}$$

and

$$m(\omega) := \max\{N(\omega), \lceil C(\omega) \rceil\}$$

Then, for every $n \ge m(\omega)$, we have

- $X_i(\omega) \le C(\omega) \le n$ for all $i \in \{1, \ldots, N(\omega) - 1\}$
- $X_i(\omega) \le i \le n$ for all $i \in \{N(\omega), \ldots, n\}$

Hence

$$M_n(\omega) = \max_{1 \le i \le n} X_i(\omega) = \max\{\underbrace{X_1(\omega), \ldots, X_{N(\omega)-1}(\omega)}_{\le C(\omega) \le n}, \underbrace{X_{N(\omega)}(\omega), \ldots, X_n(\omega)}_{\le n}\} \le n$$

for all $n \ge m(\omega)$. Therefore

$$\frac{M_n(\omega)}{n} \le 1$$

for all $n \ge m(\omega)$. Since $\mathbb{P}(A^{\complement}) = 1$, we conclude that

$$\frac{M_n}{n} \le 1$$

eventually $\mathbb{P}$-almost surely.

(d) The running maximum M_n grows no faster than linearly, almost surely. More precisely, with probability one, there exists $m(\omega) \in \mathbb{N}$ such that

$$M_n(\omega) \le n$$

for all $n \ge m(\omega)$. Equivalently, the sequence $(X_n)_{n \in \mathbb{N}}$ exceeds the linear threshold n only finitely often. ◀

Exercise 4.10 (Infinitely many successes). Let $(X_n)_{n \in \mathbb{N}}$ be independent Bernoulli random variables with

$$\mathbb{P}(X_n = 1) = p \qquad \mathbb{P}(X_n = 0) = 1 - p \qquad p \in (0, 1)$$

Consider the event

$$A := \{\omega \in \Omega : X_n(\omega) = 1 \text{ for infinitely many } n \in \mathbb{N}\}$$

(a) Show that A belongs to the tail σ-algebra

$$\mathcal{T} := \bigcap_{N=1}^{\infty} \sigma(\{X_n : n \ge N\})$$

(b) Deduce that $\mathbb{P}(A) \in \{0, 1\}$

(c) Show that $\mathbb{P}(A) = 1$ ✍

Solution.

(a) Define $A_n := \{\omega \in \Omega : X_n(\omega) = 1\}$ for $n \in \mathbb{N}$. We have

$$A = \limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \bigcap_{k=N}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

for any fixed $N \in \mathbb{N}$, since removing finitely many initial events does not change the event that A_k occurs for infinitely many k.

Next, we define

$$\mathcal{T}_N := \sigma(\{X_n : n \geq N\})$$

for $N \in \mathbb{N}$. Obviously, $A_n \in \mathcal{T}_N$ for all $n \geq N$. Since $\mathcal{T}_N$ is closed under countable unions, we have

$$\bigcup_{n=k}^{\infty} A_n \in \mathcal{T}_N$$

for all $k \geq N$. Hence, for any fixed $N \in \mathbb{N}$, we also have

$$A = \bigcap_{k=N}^{\infty} \bigcup_{n=k}^{\infty} A_n \in \mathcal{T}_N$$

since $\mathcal{T}_N$ is closed under countable intersections. Since this holds for all $N \in \mathbb{N}$, we have

$$A \in \bigcap_{N=1}^{\infty} \mathcal{T}_N = \mathcal{T}$$

(b) Since the X_n are independent random variables, the $A_n = X_n^{-1}(\{1\}) \in \sigma(X_n)$ are independent events. Since

$$A = \limsup_{n\to\infty} A_n$$

we have by Theorem 4.7 that $\mathbb{P}(A) \in \{0, 1\}$.

(c) We show that the value is in fact 1. Consider the complement

$$A^{\complement} = \{\omega \in \Omega : X_n(\omega) = 0 \text{ eventually}\} = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{X_n = 0\}$$

For fixed $N \in \mathbb{N}$, define

$$B_N := \bigcap_{n=N}^{\infty} \{X_n = 0\}$$

and, for $m \geq N$,

$$B_{N,m} := \bigcap_{n=N}^{m} \{X_n = 0\}$$

Then $(B_{N,m})_{m\geq N}$ is a decreasing sequence of events with

$$B_N = \bigcap_{m=N}^{\infty} B_{N,m}$$

By continuity of $\mathbb{P}$ from above (Theorem 1.11), we obtain

$$\mathbb{P}(B_N) = \lim_{m\to\infty} \mathbb{P}(B_{N,m})$$

Using the independence of the random variables X_n, we get

$$\mathbb{P}(B_{N,m}) = \mathbb{P}\left(\bigcap_{n=N}^{m} \{X_n = 0\}\right) \overset{(4.1)}{=} \prod_{n=N}^{m} \mathbb{P}(X_n = 0) = (1-p)^{m-N+1}$$

Since $0 < 1 - p < 1$, it follows that

$$\mathbb{P}(B_N) = \lim_{m\to\infty} (1-p)^{m-N+1} = 0$$

for every $N \in \mathbb{N}$. Hence, by countable subadditivity (Theorem 1.9),

$$\mathbb{P}(A^{\complement}) = \mathbb{P}\left(\bigcup_{N=1}^{\infty} B_N\right) \overset{(1.5)}{\leq} \sum_{N=1}^{\infty} \mathbb{P}(B_N) = 0$$

Thus $\mathbb{P}(A^{\complement}) = 0$, and therefore $\mathbb{P}(A) = 1$.

Alternatively, this follows immediately from Theorem 4.6. The events $A_n = \{X_n = 1\}$ are independent and

$$\mathbb{P}(A_n) = \mathbb{P}(X_n = 1) = p \in (0,1)$$

for all $n \in \mathbb{N}$. Thus,

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} p = \infty$$

and hence, by Theorem 4.6, we obtain $\mathbb{P}(A) = 1$. ◀

5 The integral

5.1 Construction of the integral

The construction a general measure integral is performed in the following steps:

1. Definition 5.1: Construction for indicator functions
2. Definition 5.2: Construction for non-negative simple functions
3. Definition 5.3: Construction for non-negative measurable functions
4. Definition 5.4: Construction for integrable functions

Throughout this section, let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

Definition 5.1 (Step 1)**.** Let $f \coloneqq \mathbf{1}_A$ with $A \in \mathcal{F}$. We define

$$\int f \, \mathrm{d}\mu \coloneqq \mu(A) \tag{5.1}$$

◀

Next, we consider non-negative simple functions

$$f \coloneqq \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i} \tag{5.2}$$

where $\alpha_1, \ldots, \alpha_m \geq 0$ and $A_1, \ldots, A_m \in \mathcal{F}$ with $\Omega = \bigcup_{i=1}^{m} A_i$. Without loss of generality, we assume $A_i \cap A_j = \varnothing$ for $i \neq j$. As the representation of f in (5.2) is not unique, we require the following lemma, to be able to define an integral in a sensible way.

Lemma 5.1. Let f be a non-negative simple function and suppose

$$f = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i} = \sum_{j=1}^{n} \beta_j \mathbf{1}_{B_j}$$

Then, there holds

$$\sum_{i=1}^{m} \alpha_i \mu(A_i) = \sum_{j=1}^{n} \beta_j \mu(B_j)$$

◁

Proof. By definition of f, there holds

$$\mu(A_i) = \sum_{j=1}^{n} \mu(A_i \cap B_j) \qquad i = 1, \ldots, m$$

as well as

$$\mu(B_j) = \sum_{i=1}^{m} \mu(A_i \cap B_j) \qquad j = 1, \ldots, n$$

Furthermore, we observe that $\alpha_i = \beta_j$, whenever $\mu(A_i \cap B_j) > 0$. Hence, we immediately obtain

$$\sum_{i=1}^{m} \alpha_i \mu(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \mu(A_i \cap B_j) = \sum_{j=1}^{n} \sum_{i=1}^{m} \beta_j \mu(A_i \cap B_j) = \sum_{j=1}^{n} \beta_j \mu(B_j)$$

□

Definition 5.2 (Step 2)**.** Let $f = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i}$ be a non-negative simple function. We define

$$\int f \, \mathrm{d}\mu \coloneqq \sum_{i=1}^{m} \alpha_i \mu(A_i) \tag{5.3}$$

◀

The integral derived in the previous definition has the following elementary properties.

Lemma 5.2. Let f, g be non-negative simple functions and let $a \in [0, \infty)$. There holds

(i) $\int (f + g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$

(ii) $\int (af) \, \mathrm{d}\mu = a \int f \, \mathrm{d}\mu$

(iii) $f \leq g$ (pointwise) implies $\int f \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu$ ◁

Proof. (i) Let $f = \sum_{i=1}^m \alpha_i \mathbf{1}_{A_i}$, $g = \sum_{j=1}^n \beta_j \mathbf{1}_{B_j}$. Since $A_1, \dots, A_m$ as well as $B_1, \dots, B_n$ are disjoint partitions of Ω, we can write

$$f = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \mathbf{1}_{A_i \cap B_j} \qquad g = \sum_{i=1}^m \sum_{j=1}^n \beta_j \mathbf{1}_{A_i \cap B_j}$$

Hence, we have the representation

$$f + g = \sum_{i=1}^m \sum_{j=1}^n (\alpha_i + \beta_j) \mathbf{1}_{A_i \cap B_j}$$

implying

$$\int (f+g)\,\mathrm{d}\mu = \sum_{i=1}^m \sum_{j=1}^n (\alpha_i + \beta_j)\mu(A_i \cap B_j) = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \mu(A_i \cap B_j) + \sum_{i=1}^m \sum_{j=1}^n \beta_j \mu(A_i \cap B_j) = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$$

(ii) Follows directly from the definition of the integral and the linearity of the sum. Indeed, if $f = \sum_{i=1}^m \alpha_i \mathbf{1}_{A_i}$, then $af = a\sum_{i=1}^m \alpha_i \mathbf{1}_{A_i} = \sum_{i=1}^m a\alpha_i \mathbf{1}_{A_i}$, and hence

$$\int af\,\mathrm{d}\mu = \sum_{i=1}^m a\alpha_i \mu(A_i) = a\sum_{i=1}^m \alpha_i \mu(A_i) = a\int f\,\mathrm{d}\mu$$

(iii) Let f, g be represented as in (i). Then, there holds $\alpha_i \le \beta_j$ whenever $\mu(A_i \cap B_j) > 0$. We obtain

$$\int f\,\mathrm{d}\mu = \sum_{i=1}^m \sum_{j=1}^n \alpha_i \mu(A_i \cap B_j) \le \sum_{i=1}^m \sum_{j=1}^n \beta_j \mu(A_i \cap B_j) = \int g\,\mathrm{d}\mu$$

as claimed. □

To define the integral for non-negative measurable functions, we require the following approximation property.

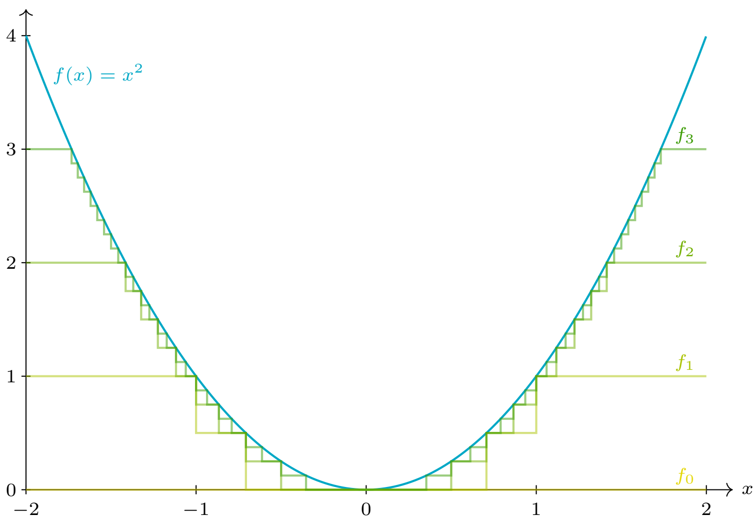
Lemma 5.3. Let $f : \Omega \to [0, \infty]$ be measurable (Definition 3.1). Then

(i) there exists a monotonically increasing sequence $(f_n)_{n\in\mathbb{N}}$ of non-negative simple functions such that $f_n \to f$

(ii) there exist measurable sets $A_1, A_2, \dots \in \mathcal{F}$ and numbers $\alpha_1, \alpha_2, \dots \ge 0$ such that

$$f = \sum_{n=1}^\infty \alpha_n \mathbf{1}_{A_n}$$

◁



Proof. (i) For $n \in \mathbb{N}_0$, define $f_n = \min\{2^{-n}\lfloor 2^n f \rfloor, n\}$. The function f_n is measurable and assumes at most $n2^n + 1$ (finitely many) different values; $0, 2^{-n}, 2 \cdot 2^{-n}, 3 \cdot 2^{-n}, \dots, n$. Hence it is a simple function. Clearly, $f_n(\omega) \le f(\omega)$ for all $\omega \in \Omega$. Moreover, the sequence f_n is monotonically increasing and converging to f.

(ii) Let f_n be as in (i). Since the sequence $(f_n)_{n\in\mathbb{N}_0}$ is monotonically increasing, the functions $f_n - f_{n-1}$ are non-negative for every $n \in \mathbb{N}$. Moreover, since f_n and f_{n-1} are simple functions, also $f_n - f_{n-1}$ is a non-negative simple function. Set

$$B_{n,i} := \{\omega \in \Omega : f_n(\omega) - f_{n-1}(\omega) = i2^{-n}\} \qquad \text{and} \qquad \beta_{n,i} := i2^{-n}$$

for $n \in \mathbb{N}$ and $i = 1, \dots, 2^n$. Since $f_n - f_{n-1}$ is measurable, all sets $B_{n,i}$ are measurable. Furthermore, the sets $B_{n,i}$ describe the non-zero level sets of $f_n - f_{n-1}$. For every fixed $n \in \mathbb{N}$, the refinement from f_{n-1} to f_n can increase the value by at most 1, and the possible increments are multiples of 2^{-n}. So the possible non-zero values of $f_n - f_{n-1}$ are contained in $2^{-n}, 2 \cdot 2^{-n}, 3 \cdot 2^{-n}, \dots, 1$. The zero level is omitted, since it does not contribute to the sum. Hence

$$f_n - f_{n-1} = \sum_{i=1}^{2^n} \beta_{n,i} \mathbf{1}_{B_{n,i}}$$

for every $n \in \mathbb{N}$. By changing the numeration $(n,i) \mapsto m$, we get $(\alpha_m)_{m\in\mathbb{N}}$ and $(A_m)_{m\in\mathbb{N}}$ such that

$$f \overset{(0.1)}{=} f_0 + \sum_{n=1}^{\infty} (f_n - f_{n-1}) = \sum_{n=1}^{\infty} \sum_{i=1}^{2^n} \beta_{n,i} \mathbf{1}_{B_{n,i}} = \sum_{m=1}^{\infty} \alpha_m \mathbf{1}_{A_m}$$

□

Similarly to the case of simple functions (Lemma 5.1), we need to show that the integral is independent of the approximating sequence.

Lemma 5.4. Let $(u_n)_{n\in\mathbb{N}}$ and $(v_m)_{m\in\mathbb{N}}$ be monotonically increasing sequences of non-negative simple functions. Then, if $\sup_{n\in\mathbb{N}} u_n = \sup_{m\in\mathbb{N}} v_m$, there holds

$$\sup_{n\in\mathbb{N}} \int u_n \,\mathrm{d}\mu = \sup_{m\in\mathbb{N}} \int v_m \,\mathrm{d}\mu$$

◁

Proof. Let $v_m = \sum_{j=1}^{k_m} \alpha_{m,j} \mathbf{1}_{A_{m,j}}$ and define the sets $B_{n,m}^s := \{u_n \geq s v_m\} \in \mathcal{F}$, $s \in \mathbb{R}$. Since u_n and v_m are measurable, we have $B_{n,m}^s \in \mathcal{F}$. Moreover, since $(u_n)_{n\in\mathbb{N}}$ is monotonically increasing, the sets $B_{n,m}^s$ are monotonically increasing in n as well. Indeed, if $\omega \in B_{n,m}^s$, then $u_n(\omega) \geq s v_m(\omega)$ and $u_{n+1}(\omega) \geq u_n(\omega)$, hence $\omega \in B_{n+1,m}^s$.

For $s \in (0,1)$, we have due to $\sup_{n\in\mathbb{N}} u_n \geq v_m$ that $\bigcup_{n=1}^{\infty} B_{n,m}^s = \Omega$. Consequently, for every $j = 1, \dots, k_m$, we have

$$A_{m,j} \cap \bigcup_{n=1}^{\infty} B_{n,m}^s = A_{m,j}$$

and therefore, since $A_{m,j} \cap B_{n,m}^s$ is increasing in n, continuity from below (Theorem 1.10) gives

$$\lim_{n\to\infty} \mu(A_{m,j} \cap B_{n,m}^s) \overset{(1.6)}{=} \mu\left(\bigcup_{n=1}^{\infty} (A_{m,j} \cap B_{n,m}^s)\right) = \mu\left(A_{m,j} \cap \bigcup_{n=1}^{\infty} B_{n,m}^s\right) = \mu(A_{m,j}) \tag{5.4}$$

By construction, we have $u_n \geq s v_m \mathbf{1}_{B_{n,m}^s}$, since, on $B_{n,m}^s$ this is exactly the inequality $u_n \geq s v_m$, and outside $B_{n,m}^s$ the right hand side is zero. Moreover, $s v_m \mathbf{1}_{B_{n,m}^s}$ is a non-negative simple function (5.2), since

$$s v_m \mathbf{1}_{B_{n,m}^s} = s \sum_{j=1}^{k_m} \alpha_{m,j} \mathbf{1}_{A_{m,j}} \mathbf{1}_{B_{n,m}^s} = \sum_{j=1}^{k_m} s \alpha_{m,j} \mathbf{1}_{A_{m,j} \cap B_{n,m}^s}$$

Hence, by Lemma 5.2,

$$\int u_n \,\mathrm{d}\mu \overset{\text{(iii)}}{\geq} \int s v_m \mathbf{1}_{B_{n,m}^s} \,\mathrm{d}\mu \overset{\text{(ii)}}{=} s \int v_m \mathbf{1}_{B_{n,m}^s} \,\mathrm{d}\mu \overset{(5.2)}{=} s \sum_{j=1}^{k_m} \alpha_{m,j} \mu(A_{m,j} \cap B_{n,m}^s)$$

Taking the supremum over n yields

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \int u_n \,\mathrm{d}\mu &\geq s \sup_{n\in\mathbb{N}} \sum_{j=1}^{k_m} \alpha_{m,j} \mu(A_{m,j} \cap B_{n,m}^s) = s \lim_{n\to\infty} \sum_{j=1}^{k_m} \alpha_{m,j} \mu(A_{m,j} \cap B_{n,m}^s) \\
&= s \sum_{j=1}^{k_m} \alpha_{m,j} \lim_{n\to\infty} \mu(A_{m,j} \cap B_{n,m}^s) \overset{(5.4)}{=} s \sum_{j=1}^{k_m} \alpha_{m,j} \mu(A_{m,j}) = s \int v_m \,\mathrm{d}\mu
\end{aligned}$$

where we used the fact that the sequence inside the supremum is increasing in n (hence the limit and the supremum are equal), and since the sum is finite, we could pass the limit through the sum. Letting $s \nearrow 1$ and taking the supremum over $m \in \mathbb{N}$ yields the claim. The opposite inequality is obtained analogously, by exchanging the roles of $(u_n)_{n\in\mathbb{N}}$ and $(v_m)_{m\in\mathbb{N}}$. □

Definition 5.3 (Step 3). Let $f : \Omega \to [0, \infty]$ be measurable and let $(f_n)_{n\in\mathbb{N}}$ be a monotonically increasing sequence of non-negative simple functions with $f = \sup_{n\in\mathbb{N}} f_n$. We define

$$\int f \,\mathrm{d}\mu := \sup_{n\in\mathbb{N}} \int f_n \,\mathrm{d}\mu \tag{5.5}$$

◀

The properties from Step 2 directly transfer to Step 3.

Lemma 5.5. Let $f, g : \Omega \to [0, \infty]$ be measurable and let $a \in [0, \infty)$. There holds

(i) $\int (f + g) \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\mu$

(ii) $\int (af) \,\mathrm{d}\mu = a \int f \,\mathrm{d}\mu$

(iii) $f \leq g$ (pointwise) implies $\int f \,\mathrm{d}\mu \leq \int g \,\mathrm{d}\mu$ ◁

By splitting a given measurable function $f : \Omega \to [-\infty, \infty]$ into its positive part

$$f^+ := \sup\{f, 0\}$$

and its negative part

$$f^- := \sup\{-f, 0\}$$

we arrive at the final step.

Definition 5.4 (Step 4). A measurable function $f : \Omega \to [-\infty, \infty]$ is called μ-integrable if

$$\int |f| \,\mathrm{d}\mu = \int f^+ \,\mathrm{d}\mu + \int f^- \,\mathrm{d}\mu < \infty$$

If either of the terms on the right hand side is bounded, we define the μ-integral of f according to

$$\int f \,\mathrm{d}\mu := \int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu \in [-\infty, \infty] \tag{5.6}$$

Finally, for $A \in \mathcal{F}$, we define the integral of f over A by

$$\boxed{\int_A f \,\mathrm{d}\mu := \int f \cdot \mathbf{1}_A \,\mathrm{d}\mu} \tag{5.7}$$

◀

Lemma 5.6. For integrable functions f, g, there holds

(i) $f \leq g$ μ-almost everywhere implies $\int f \,\mathrm{d}\mu \leq \int g \,\mathrm{d}\mu$

(ii) $f = g$ μ-almost everywhere implies $\int f \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu$

(iii) $\int |f| \,\mathrm{d}\mu < \infty$ implies $|f| < \infty$ μ-almost everywhere ◁

Example 5.1.

1. For $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, we denote the Lebesgue integral by

$$\int_{\mathbb{R}^d} f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} := \int_{\mathbb{R}^d} f \,\mathrm{d}\lambda^d \tag{5.8}$$

   More generally, for $A \in \mathcal{B}(\mathbb{R}^d)$, we write

$$\int_A f(\boldsymbol{x}) \,\mathrm{d}\boldsymbol{x} := \int_A f \,\mathrm{d}\lambda^d \tag{5.9}$$

   In case $d = 1$ and $A = (a, b)$ or $A = [a, b]$, we also write

$$\int_a^b f(x) \,\mathrm{d}x := \int_A f \,\mathrm{d}\lambda^1 \tag{5.10}$$

2. For $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_F)$ with the Stieltjes measure μ_F (Example 2.8), we denote the Stieltjes integral by

$$\int_\Omega g(x) \,\mathrm{d}F(x) := \int_\Omega g \,\mathrm{d}\mu_F$$

   If $F \in C^1(\mathbb{R})$ with $F' = f$, there holds

$$\int_\Omega g(x) \,\mathrm{d}F(x) = \int_\Omega g(x) f(x) \,\mathrm{d}x$$

3. If Ω is countable, $\mathcal{F} = 2^\Omega$ and $\mu = \sum_{\omega\in\Omega} \delta_\omega$ is the counting measure (Definition 1.6), then there holds

$$\int f \,\mathrm{d}\mu = \sum_{\omega\in\Omega} f(\omega) \tag{5.11}$$

◀

Definition 5.5. $\mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ denotes the set of all real-valued μ-integrable functions. ◀

Lemma 5.7. The set $\mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ together with the usual addition and scalar multiplication of functions is a real vector space. Moreover, the mapping $f \mapsto \int f \, d\mu$ is a continuous linear form on $\mathcal{L}^1(\Omega, \mathcal{F}, \mu)$. ◁

For random variables, the linear form defined earlier amounts to the expectation.

Definition 5.6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. The expectation of X is defined as

$$\boxed{\mathbb{E}[X] := \int_\Omega X \, d\mathbb{P}} \tag{5.12}$$

◀

Example 5.2. We consider infinitely many rolls of a fair die. Let $D := \{1, 2, 3, 4, 5, 6\}$ with σ-algebra 2^D and probability measure $\mathbb{P}_D$ given by $\mathbb{P}_D(\{k\}) = \frac{1}{6}$ for $k \in D$. Set $\Omega := D^{\mathbb{N}}$ and for each $i \in \mathbb{N}$ define the coordinate maps $X_i : \Omega \to D$,

$$\omega \mapsto X_i(\omega) := \omega_i = \sum_{k \in D} k \, \mathbf{1}_{X_i^{-1}(\{k\})}(\omega)$$

We define the σ-algebra

$$\mathcal{F} := \sigma\left(X_i^{-1}(A) : i \in \mathbb{N}, \; A \in 2^D\right)$$

which is the smallest σ-algebra on Ω which makes all coordinate maps X_i measurable. On the algebra of cylinder sets

$$\mathcal{C} := \left\{\bigcap_{i=1}^r X_i^{-1}(A_i) : r \in \mathbb{N}, \; A_i \in 2^D\right\}$$

$\bigtimes_{i=1}^r A_i \times \bigtimes_{i=r+1}^\infty D = \{\omega \in \Omega : X_1(\omega) \in A_1, \ldots, X_r(\omega) \in A_r\}$ is a cylinder

we define $\mathbb{P}$ by

$$\mathbb{P}\left(\left(\bigtimes_{i=1}^r A_i\right) \times \left(\bigtimes_{i=r+1}^\infty D\right)\right) := \prod_{i=1}^r \mathbb{P}_D(A_i) \tag{5.13}$$

for $r \in \mathbb{N}$ and $A_1, \ldots, A_r \in 2^D$. Then, by Theorem 2.5, $\mathbb{P}$ extends uniquely to a probability measure on $\mathcal{F}$, which we call the *infinite product measure* and write $\mathcal{F} = \bigotimes_{i \in \mathbb{N}} 2^D$, $\mathbb{P} = \mathbb{P}_D^{\otimes \mathbb{N}}$.

Now, we note that each X_m has the same distribution $\mathbb{P}_D$, and

$\bigtimes_{i=1}^{m-1} D \times \{k\} \times \bigtimes_{i=m+1}^\infty D \in \mathcal{C}$ (also a cylinder)

$$\mathbb{E}[X_m] \overset{(5.12)}{=} \int X_m \, d\mathbb{P} \overset{(5.3)}{=} \sum_{k \in D} k \, \mathbb{P}(X_m^{-1}(\{k\})) \overset{(5.13)}{=} \sum_{k \in D} k \, \mathbb{P}_D(\{k\}) = \frac{1}{6} \sum_{k=1}^6 k = 3.5$$

which is the expected value of a single die roll. Now consider $S_n := \frac{1}{n} \sum_{i=1}^n X_i$, the average result of the first n rolls. By linearity of expectation, there holds

$$\mathbb{E}[S_n] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \overset{\text{Lemma 5.7}}{=} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = 3.5$$

◀

Example 5.3 (Expected waiting time for the first 6). Consider a general die with $\mathbb{P}_D(\{6\}) = p > 0$. Let

$$T(\omega) := \min\{n \in \mathbb{N} : X_n(\omega) = 6\}$$

be the waiting time for the first 6. We have

$$\{T = k\} = \left\{\omega \in \Omega : \bigwedge_{i=1}^{k-1} X_i(\omega) \neq 6 \wedge X_k(\omega) = 6\right\} = \bigcap_{i=1}^{k-1} \{X_i \neq 6\} \cap \{X_k = 6\}$$

and therefore, by independence of the coordinate maps,

$$\mathbb{P}(T = k) = \left(\prod_{i=1}^{k-1} \mathbb{P}(X_i \neq 6)\right) \cdot \mathbb{P}(X_k = 6) = (1 - p)^{k-1} p$$

Now, we can compute the expected waiting time for the occurrence of the first 6. There holds

$$\mathbb{E}[T] \overset{(5.12)}{=} \int_\Omega T \, d\mathbb{P} = \sum_{k \in \mathbb{N}} \int_{B_k} T \, d\mathbb{P}$$

since the sets $B_k = \{T = k\}$ partition Ω up to the $\mathbb{P}$-null set $\{T = \infty\}$. Due to $T|_{B_k} = k$, we arrive at

$$\mathbb{E}[T] = \sum_{k \in \mathbb{N}} \int_{B_k} T \, d\mathbb{P} = \sum_{k \in \mathbb{N}} k \, \mathbb{P}(B_k) = \sum_{k \in \mathbb{N}} kp(1 - p)^{k-1} = \frac{1}{p}$$

◀

Proposition 5.8 (Integration with respect to finitely many Dirac measures)**.** Let $(E,\mathcal{E})$ be a measurable space, let $x_1,\dots,x_n \in E$, and let $a_1,\dots,a_n \geq 0$. Set

$$\mu := \sum_{i=1}^{n} a_i \delta_{x_i}$$

where $\delta_{x_i}(A) = \mathbf{1}_A(x_i)$ for $A \in \mathcal{E}$. If $f : E \to \overline{\mathbb{R}}$ is μ-integrable, then

$$\int_E f \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i f(x_i) \tag{5.14}$$

◁

Proof. We follow the construction steps of the integral.

1. Let $f = \mathbf{1}_A$ with $A \in \mathcal{E}$. Then, by the definition of μ, we have

$$\mu(A) = \sum_{i=1}^{n} a_i \delta_{x_i}(A) = \sum_{i=1}^{n} a_i \mathbf{1}_A(x_i) \tag{5.15}$$

   Hence,

$$\int_E \mathbf{1}_A \, \mathrm{d}\mu \overset{(5.1)}{=} \mu(A) \overset{(5.15)}{=} \sum_{i=1}^{n} a_i \mathbf{1}_A(x_i)$$

2. Let $s : E \to [0,\infty)$ be a non-negative simple function of the form $s = \sum_{j=1}^{m} \alpha_j \mathbf{1}_{A_j}$. Then,

$$\begin{aligned}
\int_E s \, \mathrm{d}\mu &\overset{(5.3)}{=} \sum_{j=1}^{m} \alpha_j \mu(A_j) \\
&\overset{(5.15)}{=} \sum_{j=1}^{m} \alpha_j \sum_{i=1}^{n} a_i \mathbf{1}_{A_j}(x_i) \\
&= \sum_{i=1}^{n} a_i \sum_{j=1}^{m} \alpha_j \mathbf{1}_{A_j}(x_i) = \sum_{i=1}^{n} a_i s(x_i)
\end{aligned}$$

3. Let $g : E \to [0,\infty]$ be measurable. By Lemma 5.3, there exists a monotonically increasing sequence $(s_k)_{k\in\mathbb{N}}$ of non-negative simple functions such that $s_k \nearrow g$ pointwise. By the previous step,

$$\begin{aligned}
\int_E g \, \mathrm{d}\mu &\overset{(5.5)}{=} \sup_{k\in\mathbb{N}} \int_E s_k \, \mathrm{d}\mu \\
&= \sup_{k\in\mathbb{N}} \sum_{i=1}^{n} a_i s_k(x_i) \\
&= \sum_{i=1}^{n} a_i \sup_{k\in\mathbb{N}} s_k(x_i) = \sum_{i=1}^{n} a_i g(x_i)
\end{aligned}$$

   where in the last step we used that the sum is finite and $s_k(x_i) \nearrow g(x_i)$ for every $i = 1,\dots,n$.

4. Let $f : E \to \overline{\mathbb{R}}$ be μ-integrable. By the previous step applied to f^+ and f^-, we have

$$\int_E f^+ \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i f^+(x_i) \qquad \text{and} \qquad \int_E f^- \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i f^-(x_i)$$

   Therefore,

$$\begin{aligned}
\int_E f \, \mathrm{d}\mu &\overset{(5.6)}{=} \int_E f^+ \, \mathrm{d}\mu - \int_E f^- \, \mathrm{d}\mu \\
&= \sum_{i=1}^{n} a_i f^+(x_i) - \sum_{i=1}^{n} a_i f^-(x_i) \\
&= \sum_{i=1}^{n} a_i \big(f^+(x_i) - f^-(x_i)\big) = \sum_{i=1}^{n} a_i f(x_i)
\end{aligned}$$

   This proves (5.14). □

Proposition 5.9 (Integration with respect to countably many Dirac measures)**.** Let $(E, \mathcal{E})$ be a measurable space, let $(x_i)_{i\in\mathbb{N}}$ be a sequence in E, and let $(a_i)_{i\in\mathbb{N}}$ be a sequence in $[0, \infty)$. Set

$$\mu := \sum_{i\in\mathbb{N}} a_i \delta_{x_i}$$

where $\delta_{x_i}(A) = \mathbf{1}_A(x_i)$ for $A \in \mathcal{E}$. If $f : E \to \overline{\mathbb{R}}$ is μ-integrable, then

$$\int_E f \,\mathrm{d}\mu = \sum_{i\in\mathbb{N}} a_i f(x_i) \tag{5.16}$$

◁

Proof. We follow the construction steps of the integral.

1. Let $f = \mathbf{1}_A$ with $A \in \mathcal{E}$. Then, by the definition of μ, we have

$$\mu(A) = \sum_{i\in\mathbb{N}} a_i \delta_{x_i}(A) = \sum_{i\in\mathbb{N}} a_i \mathbf{1}_A(x_i) \tag{5.17}$$

Hence,

$$\int_E \mathbf{1}_A \,\mathrm{d}\mu \overset{(5.1)}{=} \mu(A) \overset{(5.17)}{=} \sum_{i\in\mathbb{N}} a_i \mathbf{1}_A(x_i)$$

2. Let $s : E \to [0, \infty)$ be a non-negative simple function of the form $s = \sum_{j=1}^m \alpha_j \mathbf{1}_{A_j}$. Then,

$$\begin{aligned}
\int_E s \,\mathrm{d}\mu &\overset{(5.3)}{=} \sum_{j=1}^m \alpha_j \mu(A_j) \\
&\overset{(5.17)}{=} \sum_{j=1}^m \alpha_j \sum_{i\in\mathbb{N}} a_i \mathbf{1}_{A_j}(x_i) \\
&= \sum_{i\in\mathbb{N}} a_i \sum_{j=1}^m \alpha_j \mathbf{1}_{A_j}(x_i) = \sum_{i\in\mathbb{N}} a_i s(x_i)
\end{aligned}$$

where the interchange of the sums is allowed because all terms are non-negative and the sum over $j = 1, \ldots, m$ is finite.

3. Let $g : E \to [0, \infty]$ be measurable. By Lemma 5.3, there exists a monotonically increasing sequence $(s_k)_{k\in\mathbb{N}}$ of non-negative simple functions such that $s_k \nearrow g$ pointwise. By the previous step,

$$\begin{aligned}
\int_E g \,\mathrm{d}\mu &\overset{(5.5)}{=} \sup_{k\in\mathbb{N}} \int_E s_k \,\mathrm{d}\mu \\
&= \sup_{k\in\mathbb{N}} \sum_{i\in\mathbb{N}} a_i s_k(x_i) \\
&= \sup_{k\in\mathbb{N}} \sup_{N\in\mathbb{N}} \sum_{i=1}^N a_i s_k(x_i) \\
&= \sup_{N\in\mathbb{N}} \sup_{k\in\mathbb{N}} \sum_{i=1}^N a_i s_k(x_i) \\
&= \sup_{N\in\mathbb{N}} \sum_{i=1}^N a_i \sup_{k\in\mathbb{N}} s_k(x_i) \\
&= \sup_{N\in\mathbb{N}} \sum_{i=1}^N a_i g(x_i) = \sum_{i\in\mathbb{N}} a_i g(x_i)
\end{aligned}$$

where we used that, for non-negative series, the infinite sum is the supremum of its partial sums, that the supremum over the product index set $\mathbb{N} \times \mathbb{N}$ is independent of the order in which the two suprema are taken (i.e. suprema commute), and that the sum over $i = 1, \ldots, N$ is finite.

4. Let $f : E \to \overline{\mathbb{R}}$ be μ-integrable. Applying the previous step to $|f|$ yields

$$\sum_{i\in\mathbb{N}} a_i |f(x_i)| = \int_E |f| \,\mathrm{d}\mu < \infty$$

Hence $\sum_{i\in\mathbb{N}} a_i f(x_i)$ is absolutely convergent.

By the previous step applied to f^+ and f^-, we have

$$\int_E f^+ \, \mathrm{d}\mu = \sum_{i \in \mathbb{N}} a_i f^+(x_i) \qquad \text{and} \qquad \int_E f^- \, \mathrm{d}\mu = \sum_{i \in \mathbb{N}} a_i f^-(x_i)$$

and both sums are finite. Therefore,

$$\begin{aligned} \int_E f \, \mathrm{d}\mu &\overset{(5.6)}{=} \int_E f^+ \, \mathrm{d}\mu - \int_E f^- \, \mathrm{d}\mu \\ &= \sum_{i \in \mathbb{N}} a_i f^+(x_i) - \sum_{i \in \mathbb{N}} a_i f^-(x_i) \\ &= \sum_{i \in \mathbb{N}} a_i \big(f^+(x_i) - f^-(x_i)\big) = \sum_{i \in \mathbb{N}} a_i f(x_i) \end{aligned}$$

where the penultimate equality is justified by absolute convergence. This proves (5.16). □

5.2 Properties of the Integral

Theorem 5.10 (Change of variables)**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(\Omega', \mathcal{F}')$ be a measurable space. Further let $T : \Omega \to \Omega'$ be measurable. Then, there holds for every measurable mapping $f : \Omega' \to \overline{\mathbb{R}}$ that

(i)
$$\int_{\Omega'} f^{\pm} \, \mathrm{d}\mu_T = \int_{\Omega} f^{\pm} \circ T \, \mathrm{d}\mu \tag{5.18}$$

(ii) f is μ_T-integrable if and only if $f \circ T$ is μ-integrable. ◁

Proof. Claim (ii) immediately follows from (i). The proof of (i) follows the construction steps of the integral.

First, let $f = \mathbf{1}_A$, $A \in \mathcal{F}'$. Then, there holds

$$\int f \, \mathrm{d}\mu_T \overset{(5.1)}{=} \mu_T(A) \overset{(3.1)}{=} \mu(T^{-1}(A)) \overset{(5.1)}{=} \int \mathbf{1}_{T^{-1}(A)} \, \mathrm{d}\mu = \int \mathbf{1}_A \circ T \, \mathrm{d}\mu = \int f \circ T \, \mathrm{d}\mu$$

Second, the left-hand side and the right-hand side of (i) are linear in f. Therefore, the claim also holds for non-negative simple functions.

Finally, approximating $f \geq 0$ by a monotone increasing sequence and taking limits proves the claim for any non-negative measurable function. □

Corollary 5.11. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable. Further, let $f : \mathbb{R} \to \overline{\mathbb{R}}$ be measurable, such that either $f(X)$ is integrable or non-negative. Then, there holds

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{P}_X \tag{5.19}$$

Especially, there holds $\mathbb{E}[X] = \int_{\mathbb{R}} t \, \mathrm{d}\mathbb{P}_X$ and $\mathbb{E}[X^2] = \int_{\mathbb{R}} t^2 \, \mathrm{d}\mathbb{P}_X$. ◁

Proof.

$$\mathbb{E}[f(X)] \overset{(5.12)}{=} \int_{\Omega} f(X) \, \mathrm{d}\mathbb{P} = \int_{\Omega} f \circ X \, \mathrm{d}\mathbb{P} \overset{(i)}{=} \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{P}_X$$
□

Remark 5.4. Corollary 5.11 tells us that the statistics of X are determined by the distribution/image measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The underlying probability space is not relevant. Especially, if $X : \Omega \to \mathbb{N}$, there holds $\mathbb{E}[X] = \sum_{k \in \mathbb{N}} k \, \mathbb{P}(X = k)$ and $\mathbb{E}[X^2] = \sum_{k \in \mathbb{N}} k^2 \, \mathbb{P}(X = k)$. ◀

Definition 5.7. Let $X : \Omega \to \mathbb{R}$ be a random variable. We say that X admits a Lebesgue density if its distribution $\mathbb{P}_X$ admits a density with respect to Lebesgue measure in the sense of Definition 5.8, i.e. if there exists a measurable function $f_X : \mathbb{R} \to [0, \infty]$ such that

$$\mathbb{P}_X(A) = \int_A f_X \, \mathrm{d}\lambda^1 \overset{(5.9)}{=} \int_A f_X(x) \, \mathrm{d}x \tag{5.20}$$

for all $A \in \mathcal{B}(\mathbb{R})$. In this case, we say that f_X is a Lebesgue density of X. Such a density is unique only up to λ^1-almost everywhere equality. ◀

Remark 5.5. The usual formulas for expectations in the discrete and continuous settings are recovered directly from Corollary 5.11.

If $\mathbb{P}_X$ admits a density f_X with respect to Lebesgue measure, then by Theorem 5.15, applied with $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda^1)$, $\nu = \mathbb{P}_X$, $f = f_X$, we have,

$$\begin{aligned}
\mathbb{E}[X] &\overset{(5.12)}{=} \int_\Omega X \,\mathrm{d}\mathbb{P} \\
&= \int_\Omega \mathrm{id}_\mathbb{R} \circ X \,\mathrm{d}\mathbb{P} \\
&\overset{(5.18)}{=} \int_\mathbb{R} \mathrm{id}_\mathbb{R} \,\mathrm{d}\mathbb{P}_X \\
&\overset{(5.23)}{=} \int_\mathbb{R} \mathrm{id}_\mathbb{R} \cdot f_X \,\mathrm{d}\lambda^1 \\
&\overset{(5.8)}{=} \int_\mathbb{R} \big(\mathrm{id}_\mathbb{R} \cdot f_X\big)(x) \,\mathrm{d}x \\
&= \int_\mathbb{R} x \cdot f_X(x) \,\mathrm{d}x
\end{aligned}$$

where the second-last equality is just the notation from Example 5.1.

There are three equivalent ways of treating an $\mathbb{N}$-valued random variable.

First, one may regard X as a real-valued random variable $X : \Omega \to \mathbb{R}$ whose image is contained in $\mathbb{N}$. Then $\mathbb{P}_X$ is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and satisfies $\mathbb{P}_X(\mathbb{N}) = 1$. Thus $\mathbb{P}_X$ is concentrated on $\mathbb{N}$, and therefore $\mathrm{id}_\mathbb{R} = \sum_{k\in\mathbb{N}} k\mathbf{1}_{\{k\}}$ $\mathbb{P}_X$-almost everywhere, i.e. the RHS is the function $x \mapsto x\mathbf{1}_\mathbb{N}(x)$ from $\mathbb{R}$ to $\mathbb{R}$. Hence

$$\begin{aligned}
\mathbb{E}[X] &\overset{\text{Definition 5.6}}{=} \int_\Omega X \,\mathrm{d}\mathbb{P} \\
&\overset{\text{Theorem 5.10}}{=} \int_\mathbb{R} \mathrm{id}_\mathbb{R} \,\mathrm{d}\mathbb{P}_X \\
&\overset{\text{Lemma 5.6 (ii)}}{=} \int_\mathbb{R} \sum_{k\in\mathbb{N}} k\mathbf{1}_{\{k\}} \,\mathrm{d}\mathbb{P}_X \\
&\overset{\text{Corollary 5.13}}{=} \sum_{k\in\mathbb{N}} \int_\mathbb{R} k\mathbf{1}_{\{k\}} \,\mathrm{d}\mathbb{P}_X \\
&\overset{\text{Lemma 5.5 (ii)}}{=} \sum_{k\in\mathbb{N}} k \int_\mathbb{R} \mathbf{1}_{\{k\}} \,\mathrm{d}\mathbb{P}_X \\
&\overset{\text{Definition 5.1}}{=} \sum_{k\in\mathbb{N}} k\, \mathbb{P}_X(\{k\}) \\
&\overset{\text{Definition 3.4}}{=} \sum_{k\in\mathbb{N}} k\, \mathbb{P}(X^{-1}(\{k\})) \\
&= \sum_{k\in\mathbb{N}} k\, \mathbb{P}(\{\omega \in \Omega : X(\omega) = k\}) \\
&= \sum_{k\in\mathbb{N}} k\, \mathbb{P}(X = k)
\end{aligned}$$

Equivalently, one may regard X directly as a random variable $X : \Omega \to \mathbb{N}$. Then $\mathbb{P}_X$ is a probability measure on $(\mathbb{N}, 2^\mathbb{N})$. In this convention, $\mathrm{id}_\mathbb{N} = \sum_{k\in\mathbb{N}} k\mathbf{1}_{\{k\}}$ holds pointwise (i.e. everywhere, not just almost everywhere), and therefore the same computation is written as

$$\begin{aligned}
\mathbb{E}[X] &\overset{\text{Definition 5.6}}{=} \int_\Omega X \,\mathrm{d}\mathbb{P} \\
&\overset{\text{Theorem 5.10}}{=} \int_\mathbb{N} \mathrm{id}_\mathbb{N} \,\mathrm{d}\mathbb{P}_X \\
&= \int_\mathbb{N} \sum_{k\in\mathbb{N}} k\mathbf{1}_{\{k\}} \,\mathrm{d}\mathbb{P}_X \\
&\overset{\text{Corollary 5.13}}{=} \sum_{k\in\mathbb{N}} \int_\mathbb{N} k\mathbf{1}_{\{k\}} \,\mathrm{d}\mathbb{P}_X \\
&\overset{\text{Lemma 5.5 (ii)}}{=} \sum_{k\in\mathbb{N}} k \int_\mathbb{N} \mathbf{1}_{\{k\}} \,\mathrm{d}\mathbb{P}_X
\end{aligned}$$

$$
\begin{aligned}
&\overset{\text{Definition 5.1}}{=} \sum_{k\in\mathbb{N}} k\,\mathbb{P}_X(\{k\})\\
&\overset{\text{Definition 3.4}}{=} \sum_{k\in\mathbb{N}} k\,\mathbb{P}(X^{-1}(\{k\}))\\
&= \sum_{k\in\mathbb{N}} k\,\mathbb{P}(\{\omega\in\Omega : X(\omega)=k\})\\
&= \sum_{k\in\mathbb{N}} k\,\mathbb{P}(X=k)
\end{aligned}
$$

A third, equivalent viewpoint is to write the distribution itself as a countable sum of Dirac measures. Indeed, for every $A \in 2^{\mathbb{N}}$, we have

$$
\begin{aligned}
\mathbb{P}_X(A) &\overset{\text{Definition 3.4}}{=} \mathbb{P}(X^{-1}(A))\\
&= \mathbb{P}(X \in A)\\
&= \sum_{k\in A} \mathbb{P}(X=k)\\
&= \sum_{k\in\mathbb{N}} \mathbb{P}(X=k)\delta_k(A)
\end{aligned}
$$

and therefore

$$
\mathbb{P}_X = \sum_{k\in\mathbb{N}} \mathbb{P}(X=k)\delta_k
$$

as measures on $(\mathbb{N}, 2^{\mathbb{N}})$. Thus, using Proposition 5.9, one can write

$$
\begin{aligned}
\mathbb{E}[X] &\overset{\text{Definition 5.6}}{=} \int_\Omega X\,\mathrm{d}\mathbb{P}\\
&\overset{\text{Theorem 5.10}}{=} \int_{\mathbb{N}} \mathrm{id}_{\mathbb{N}}\,\mathrm{d}\mathbb{P}_X\\
&= \int_{\mathbb{N}} \mathrm{id}_{\mathbb{N}}\,\mathrm{d}\left(\sum_{k\in\mathbb{N}} \mathbb{P}(X=k)\delta_k\right)\\
&\overset{\text{Proposition 5.9}}{=} \sum_{k\in\mathbb{N}} \mathbb{P}(X=k)\,\mathrm{id}_{\mathbb{N}}(k)\\
&= \sum_{k\in\mathbb{N}} k\,\mathbb{P}(X=k)
\end{aligned}
$$

Here, the same formula is obtained by rewriting the measure $\mathbb{P}_X$, rather than by rewriting the function $\mathrm{id}_{\mathbb{N}}$. ◀

We have the following important result for product measures (Example 2.5).

Fact 5.12 (Tonelli–Fubini)**.** Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ-finite measure spaces (Definition 1.5) and let $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ be $(\mathcal{F}_1 \otimes \mathcal{F}_2, \mathcal{B}(\overline{\mathbb{R}}))$-measurable (Definition 3.1).

(i) If $f \geq 0$, then $f(x,\cdot) : y \mapsto f(x,y)$ is $(\mathcal{F}_2, \mathcal{B}(\overline{\mathbb{R}}))$-measurable for every $x \in \Omega_1$.
Furthermore, $f_1 : x \mapsto \int_{\Omega_2} f(x,y)\,\mathrm{d}\mu_2$ is $(\mathcal{F}_1, \mathcal{B}(\overline{\mathbb{R}}))$-measurable.

(ii) If $f \geq 0$, then $f(\cdot,y) : x \mapsto f(x,y)$ is $(\mathcal{F}_1, \mathcal{B}(\overline{\mathbb{R}}))$-measurable for every $y \in \Omega_2$.
Furthermore, $f_2 : y \mapsto \int_{\Omega_1} f(x,y)\,\mathrm{d}\mu_1$ is $(\mathcal{F}_2, \mathcal{B}(\overline{\mathbb{R}}))$-measurable.

(iii) If $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$, then $f(x,\cdot) \in \mathcal{L}^1(\mu_2)$ μ_1-almost everywhere and $f(\cdot,y) \in \mathcal{L}^1(\mu_1)$ μ_2-almost everywhere.

(iv) If $f \geq 0$ or $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$ (Definition 5.5), then there holds

$$
\int_{\Omega_1}\int_{\Omega_2} f\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \int_{\Omega_1\times\Omega_2} f\,\mathrm{d}(\mu_1\otimes\mu_2) = \int_{\Omega_2}\int_{\Omega_1} f\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2 \tag{5.21}
$$

◁

Remark 5.6. (i), (ii) together with (iv) in the case $f \geq 0$ are usually referred to as Tonelli's theorem and allow the value ∞ for the integrals. (iii) and (iv) in the case $f \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$ constitute Fubini's theorem and yield finite and equal iterated and product integrals. ◀

Corollary 5.13. Let $(\Omega, \mathcal{F}, \eta)$ be a measure space and let $f_1, f_2, \ldots : \Omega \to \overline{\mathbb{R}}$ be measurable such that $f_n \geq 0$ for all $n \in \mathbb{N}$ or $\sum_{n\in\mathbb{N}} \int_\Omega |f_n| \,\mathrm{d}\eta < \infty$. Then, there holds

$$\int_\Omega \sum_{n\in\mathbb{N}} f_n \,\mathrm{d}\eta = \sum_{n\in\mathbb{N}} \int_\Omega f_n \,\mathrm{d}\eta \tag{5.22}$$

◁

Proof. We apply Fact 5.12. Set $\mu_1 \coloneqq \eta$, $\mu_2 \coloneqq \sum_{n\in\mathbb{N}} \delta_n$ (counting measure on $\mathbb{N}$, see Example 1.7) and define

$$\begin{aligned} F : \Omega \times \mathbb{N} &\longrightarrow \overline{\mathbb{R}} \\ (\omega, n) &\longmapsto F(\omega, n) \coloneqq f_n(\omega) \end{aligned}$$

which is $(\mathcal{F} \otimes 2^{\mathbb{N}}, \mathcal{B}(\overline{\mathbb{R}}))$-measurable. Indeed, for $B \in \mathcal{B}(\overline{\mathbb{R}})$, we have $f_n^{-1}(B) \in \mathcal{F}$ by measurability of f_n, and $\{n\} \in 2^{\mathbb{N}}$. Hence

$$f_n^{-1}(B) \times \{n\} \in \{A_1 \times A_2 : A_1 \in \mathcal{F},\ A_2 \in 2^{\mathbb{N}}\} \subseteq \underbrace{\sigma(\{A_1 \times A_2 : A_1 \in \mathcal{F},\ A_2 \in 2^{\mathbb{N}}\})}_{\mathcal{F}\otimes 2^{\mathbb{N}}}$$

for every $n \in \mathbb{N}$, where we used the definition of the product σ-algebra (Example 1.4). Since

$$\begin{aligned} F^{-1}(B) &= \{(\omega, k) \in \Omega \times \mathbb{N} : F(\omega, k) \in B\} \\ &= \{(\omega, k) \in \Omega \times \mathbb{N} : f_k(\omega) \in B\} \\ &= \{(\omega, k) \in \Omega \times \mathbb{N} : f_k(\omega) \in B \wedge \texttt{true}\} \\ &= \left\{(\omega, k) \in \Omega \times \mathbb{N} : \left(\bigvee_{n=1}^{k-1} \texttt{false}\right) \vee (f_k(\omega) \in B \wedge \texttt{true}) \vee \left(\bigvee_{n=k+1}^{\infty} \texttt{false}\right)\right\} \\ &= \left\{(\omega, k) \in \Omega \times \mathbb{N} : \bigvee_{n\in\mathbb{N}} (f_n(\omega) \in B \wedge k = n)\right\} \\ &= \bigcup_{n\in\mathbb{N}} \{(\omega, k) \in \Omega \times \mathbb{N} : f_n(\omega) \in B \wedge k = n\} \\ &= \bigcup_{n\in\mathbb{N}} \big\{(\omega, k) \in \Omega \times \mathbb{N} : \omega \in f_n^{-1}(B) \wedge k \in \{n\}\big\} \\ &= \bigcup_{n\in\mathbb{N}} \big\{(\omega, k) : \omega \in f_n^{-1}(B) \wedge k \in \{n\}\big\} \\ &= \bigcup_{n\in\mathbb{N}} \big(f_n^{-1}(B) \times \{n\}\big) \end{aligned}$$

and $\mathcal{F} \otimes 2^{\mathbb{N}}$ is closed under countable unions, it follows that $F^{-1}(B) \in \mathcal{F} \otimes 2^{\mathbb{N}}$ for every $B \in \mathcal{B}(\overline{\mathbb{R}})$.

If $f_n \geq 0$ for all $n \in \mathbb{N}$, then $F \geq 0$.

In the second case, by applying Tonelli to (the non-negative function) $|F|$ and using the assumption,

$$\int_{\Omega\times\mathbb{N}} |F| \,\mathrm{d}(\eta \otimes \mu_2) \overset{\text{(iv)}}{=} \int_{\mathbb{N}} \int_\Omega |F| \,\mathrm{d}\eta \,\mathrm{d}\mu_2 \overset{(5.11)}{=} \sum_{n\in\mathbb{N}} \int_\Omega |f_n| \,\mathrm{d}\eta < \infty$$

so $F \in \mathcal{L}^1(\eta \otimes \mu_2)$ in that case (Definition 5.5).

Thus,

$$\int_\Omega \sum_{n\in\mathbb{N}} f_n \,\mathrm{d}\eta \overset{(5.11)}{=} \int_\Omega \int_{\mathbb{N}} F \,\mathrm{d}\mu_2 \,\mathrm{d}\eta \overset{\text{(iv)}}{=} \int_{\mathbb{N}} \int_\Omega F \,\mathrm{d}\eta \,\mathrm{d}\mu_2 \overset{(5.11)}{=} \sum_{n\in\mathbb{N}} \int_\Omega f_n \,\mathrm{d}\eta$$

□

Corollary 5.14. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X_1, X_2, \ldots$ random variables such that $X_n \geq 0$ for all $n \in \mathbb{N}$ or $\sum_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < \infty$. Then, there holds

$$\mathbb{E}\left[\sum_{n\in\mathbb{N}} X_n\right] = \sum_{n\in\mathbb{N}} \mathbb{E}[X_n]$$

◁

Proof. Follows from Corollary 5.13 and the definition of expectation (5.12). □

For the remainder of this section, we introduce measures that admit a density.

Definition 5.8. Let μ and ν be measures on the measurable space $(\Omega, \mathcal{F})$. We say that ν admits a density with respect to μ if there exists a measurable function $f : \Omega \to [0, \infty]$ such that

$$\nu(A) = \int_A f \, \mathrm{d}\mu$$

for all $A \in \mathcal{F}$. In this case, we say that f is a density of ν with respect to μ. ◀

Theorem 5.15. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f : \Omega \to \overline{\mathbb{R}}$ be measurable and non-negative. Then, the mapping

$$\begin{aligned} \nu : \mathcal{F} &\longrightarrow [0, \infty] \\ A &\longmapsto \nu(A) := \int_A f \, \mathrm{d}\mu \end{aligned}$$

defines a measure (Definition 1.5). In particular, f is a density of ν with respect to μ in the sense of Definition 5.8. For any measurable and non-negative function $g : \Omega \to \overline{\mathbb{R}}$, there holds

$$\boxed{\int_\Omega g \, \mathrm{d}\nu = \int_\Omega g \, f \, \mathrm{d}\mu} \tag{5.23}$$

◁

Proof. Let $A_1, A_2, \ldots \in \mathcal{F}$ be disjoint. Then, there holds

$$\begin{aligned} \nu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \int_{\bigcup_{n\in\mathbb{N}} A_n} f \, \mathrm{d}\mu &\overset{(5.7)}{=} \int_\Omega f \mathbf{1}_{\bigcup_{n\in\mathbb{N}} A_n} \, \mathrm{d}\mu \\ &= \int_\Omega f \sum_{n\in\mathbb{N}} \mathbf{1}_{A_n} \, \mathrm{d}\mu \\ &= \int_\Omega \sum_{n\in\mathbb{N}} f \mathbf{1}_{A_n} \, \mathrm{d}\mu \\ &\overset{(5.22)}{=} \sum_{n\in\mathbb{N}} \int_\Omega f \mathbf{1}_{A_n} \, \mathrm{d}\mu \overset{(5.7)}{=} \sum_{n\in\mathbb{N}} \int_{A_n} f \, \mathrm{d}\mu = \sum_{n\in\mathbb{N}} \nu(A_n) \end{aligned}$$

which shows that ν is a measure.

To prove (5.23), we follow the steps 1, 2, 3 in the construction of the integral:

1. First, let $g = \mathbf{1}_A$ for some $A \in \mathcal{F}$. By the definition of the integral for indicator functions (Definition 5.1),

$$\int_\Omega g \, \mathrm{d}\nu = \int_\Omega \mathbf{1}_A \, \mathrm{d}\nu \overset{(5.1)}{=} \nu(A) = \int_A f \, \mathrm{d}\mu \overset{(5.7)}{=} \int_\Omega f \mathbf{1}_A \, \mathrm{d}\mu = \int_\Omega g f \, \mathrm{d}\mu$$

   where the central equality is just the definition of ν. Thus, the claim holds for indicator functions.

2. Next, let g be a non-negative simple function $g = \sum_{i=1}^m \alpha_i \mathbf{1}_{A_i}$ with $\alpha_i \geq 0$ and $A_i \in \mathcal{F}$. Using the definition of the integral for non-negative simple functions (Definition 5.2) and step 1., we obtain

$$\begin{aligned} \int_\Omega g \, \mathrm{d}\nu \overset{(5.3)}{=} \sum_{i=1}^m \alpha_i \nu(A_i) &\overset{(5.1)}{=} \sum_{i=1}^m \alpha_i \int_\Omega \mathbf{1}_{A_i} \, \mathrm{d}\nu \\ &\overset{1.}{=} \sum_{i=1}^m \alpha_i \int_\Omega f \mathbf{1}_{A_i} \, \mathrm{d}\mu = \int_\Omega f \sum_{i=1}^m \alpha_i \mathbf{1}_{A_i} \, \mathrm{d}\mu = \int_\Omega f g \, \mathrm{d}\mu \end{aligned}$$

   where we used Lemma 5.5 in the penultimate step.

3. Finally, let $g : \Omega \to \overline{\mathbb{R}}$ be measurable and non-negative. By Lemma 5.3, there exists a monotonically increasing sequence $(g_n)_{n\in\mathbb{N}}$ of non-negative simple functions such that $g_n \to g$. Since $g_n \nearrow g$ and $f \geq 0$, we also have $g_n f \nearrow g f$. Therefore, by the definition of the integral for non-negative measurable functions (Definition 5.3) and step 2., we obtain

$$\int_\Omega g \, \mathrm{d}\nu \overset{(5.5)}{=} \sup_{n\in\mathbb{N}} \int_\Omega g_n \, \mathrm{d}\nu \overset{2.}{=} \sup_{n\in\mathbb{N}} \int_\Omega g_n f \, \mathrm{d}\mu \overset{(5.5)}{=} \int_\Omega g f \, \mathrm{d}\mu$$

Hence, the integral identity (5.23) holds for all non-negative measurable g. □

Definition 5.9. Let μ, ν be measures on the measurable space $(\Omega, \mathcal{F})$. The measure ν is called *absolutely continuous* with respect to μ, we write $\nu \ll \mu$, if

$$\mu(A) = 0 \implies \nu(A) = 0$$

for all $A \in \mathcal{F}$. ◀

We close this section by the following statement known as Radon–Nikodym theorem in literature.

Fact 5.16 (Radon–Nikodym theorem). Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space and let $\nu : \mathcal{F} \to [0, \infty]$ be a measure. Then ν is absolutely continuous with respect to μ in the sense of Definition 5.9 if and only if ν admits a density f with respect to μ in the sense of Definition 5.8. Especially, this density f is μ-almost everywhere uniquely determined, i.e. $\nu = f_1\mu = f_2\mu \implies \mu(\{f_1 \neq f_2\}) = 0$. ◁

5.3 L^p-spaces

Throughout this section, we assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. The first result is known as Jensen's inequality.

Theorem 5.17 (Jensen's inequality). Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex, i.e.

$$\lambda\phi(x) + (1 - \lambda)\phi(y) \geq \phi(\lambda x + (1 - \lambda)y)$$

for all $x, y \in \mathbb{R}$ and all $\lambda \in [0, 1]$. If $X : \Omega \to \mathbb{R}$ is a random variable such that X and $\phi(X)$ are integrable, then there holds

$$\boxed{\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]} \tag{5.24}$$

◁

Proof. Since ϕ is convex, there exists for each $x_0 \in \mathbb{R}$ a line $x \mapsto \ell(x) = ax + b$ such that $\ell(x) \leq \phi(x)$ and $\ell(x_0) = \phi(x_0)$. Choose $x_0 = \mathbb{E}[X]$. Then, we infer

$$\phi(\mathbb{E}[X]) = \ell(\mathbb{E}[X]) = a\,\mathbb{E}[X] + b = \mathbb{E}[aX + b] = \mathbb{E}[\ell(X)] \leq \mathbb{E}[\phi(X)]$$

by the linearity (Lemma 5.7) and monotonicity (Lemma 5.6) of the integral and the fact that $b = \mathbb{E}[b]$. □

Example 5.7. Let $X : \Omega \to \mathbb{R}$ be an integrable random variable. Then, there holds, for example,

1. $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
2. $e^{\pm\mathbb{E}[X]} \leq \mathbb{E}[e^{\pm X}]$
3. $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$ and, more generally, $|\mathbb{E}[X]|^p \leq \mathbb{E}[|X|^p]$, $1 \leq p < \infty$ ◀

Definition 5.10. Let $X : \Omega \to \mathbb{R}$ be a random variable with $\mathbb{E}[X^2] < \infty$. Then, the *variance* of X is defined as

$$\mathbb{V}\mathrm{ar}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \in [0, \infty) \tag{5.25}$$

The corresponding *standard deviation* is given by $\sigma(X) := \sqrt{\mathbb{V}\mathrm{ar}[X]}$. ◀

Remark 5.8. There holds $\mathbb{V}\mathrm{ar}[X] = 0$ if $X = \mathbb{E}[X]$ $\mathbb{P}$-almost surely. Moreover, it is easy to see that

$$\mathbb{V}\mathrm{ar}[aX + b] = a^2\,\mathbb{V}\mathrm{ar}[X]$$

for all $a, b \in \mathbb{R}$. ◀

In analogy to the variance, we can also consider higher order moments of random variables.

Definition 5.11. We define the L^p-norm according to

$$\|X\|_p := \begin{cases} \mathbb{E}[|X|^p]^{1/p} & 1 \leq p < \infty \\ \inf\{s \in [0, \infty] : |X| \leq s \ \mathbb{P}\text{-almost surely}\} & p = \infty \end{cases} \tag{5.26}$$

Corresponding to the L^p-norm, we introduce the sets

$$\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) := \{X : \Omega \to \mathbb{R} : \|X\|_p < \infty\}$$

Especially, there holds $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) \subseteq \mathcal{L}^q(\Omega, \mathcal{F}, \mathbb{P})$ for $p \geq q$. ◀

Remark 5.9. The inclusion $\mathcal{L}^p(\Omega, \mathcal{F}, \mu) \subseteq \mathcal{L}^q(\Omega, \mathcal{F}, \mu)$ is usually wrong for general measure spaces. Further, we observe that $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ if and only if $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{V}\mathrm{ar}[X] < \infty$. ◀

Fact 5.18. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. There holds for any $p, q \in [1, \infty]$ that

(i) $\|XY\|_1 \leq \|X\|_p \|Y\|_q$ whenever $1/p + 1/q = 1$ *(Hölder's inequality)*

(ii) $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$ whenever $X + Y$ is a.s. well defined *(Minkowski's inequality)* ◁

Remark 5.10. As a consequence of (ii), the L^p-norms are semi-norms. They become norms under the equivalence relation $X \sim Y$ if and only if $\mathbb{P}(\{X \neq Y\}) = 0$. The corresponding spaces are denoted by $L^p(\Omega, \mathcal{F}, \mathbb{P}) \subseteq \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ and are Banach spaces. ◀

Next, we prove the well known Chebyshev–Markov inequality.

Theorem 5.19 (Chebyshev–Markov inequality)**.** Let $X : \Omega \to \mathbb{R}$ be a random variable. Then, there holds

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a^p} \mathbb{E}[|X|^p]$$

for all $a, p > 0$. ◁

Proof. Let

$$N_a := \{|X| \geq a\}$$

Then,

$$\mathbb{E}[|X|^p] = \int_\Omega |X|^p \, \mathrm{d}\mathbb{P} \geq \int_{N_a} |X|^p \, \mathrm{d}\mathbb{P} \overset{\text{(i)}}{\geq} \int_{N_a} a^p \, \mathrm{d}\mathbb{P} = a^p \mathbb{P}(N_a)$$

where we used Lemma 5.6. □

An immediate consequence is Chebyshev's inequality.

Corollary 5.20 (Chebyshev's inequality)**.** Let $X : \Omega \to \mathbb{R}$ be a random variable. Then, there holds

$$\boxed{\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2} \mathbb{V}\mathrm{ar}[X]} \tag{5.27}$$

for all $a > 0$. ◁

Proof. Apply Theorem 5.19 to $X - \mathbb{E}[X]$ for $p = 2$. □

5.4 Convergence theorems

We give a brief overview on the most important convergence theorems for sequences of random variables. As before, let $(\Omega, \mathcal{F}, \mathbb{P})$ always denote a probability space.

Definition 5.12. Let $X, (X_n)_{n \in \mathbb{N}}$ be random variables.

(i) We say $(X_n)_{n \in \mathbb{N}}$ converges to X pointwise, denoted by $X_n \xrightarrow{\text{p.w.}} X$, if

$$\lim_{n \to \infty} X_n(\omega) = X(\omega)$$

for all $\omega \in \Omega$.

(ii) We say $(X_n)_{n \in \mathbb{N}}$ converges to X $\mathbb{P}$-almost surely, denoted by $X_n \xrightarrow{\text{a.s.}} X$, if

$$\mathbb{P}(\lim_{n \to \infty} X_n = X) = 1$$

→ shorthand for $\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}$

(iii) We say $(X_n)_{n \in \mathbb{N}}$ converges to X in probability, denoted by $X_n \xrightarrow{\text{i.p.}} X$, if

$$\mathbb{P}(|X_n - X| \geq a) \xrightarrow{n \to \infty} 0$$

for all $a > 0$.

(iv) We say $(X_n)_{n \in \mathbb{N}}$ converges to X in distribution, denoted by $X_n \xrightarrow{\text{i.d.}} X$, if

$$\mathbb{E}[f(X_n)] \xrightarrow{n \to \infty} \mathbb{E}[f(X)]$$

for all bounded $f \in C(\mathbb{R})$.

(v) We say $(X_n)_{n \in \mathbb{N}}$ converges to X in L^p, denoted by $X_n \xrightarrow{L^p} X$, if

$$\|X_n - X\|_p \xrightarrow{n \to \infty} 0$$ ◀

Remark 5.11. The following diagram summarizes the implications between the different types of convergence.

$$\begin{array}{ccccccc} \text{pointwise (i)} & \Longrightarrow & \text{almost surely (ii)} & \Longrightarrow & \text{in probability (iii)} & \Longrightarrow & \text{in distribution (iv)} \\ & & & & \Uparrow & & \\ & & & & \text{in } L^p \text{ (v)} & & \end{array}$$

There holds (ii) $\Rightarrow$ (iii). The reverse implication (iii) $\Rightarrow$ (ii) holds along a subsequence.

Moreover, we have (iii) $\Rightarrow$ (iv), while the reverse (iv) $\Rightarrow$ (iii) only holds if X is almost surely constant.

Further, there holds (v) $\Rightarrow$ (iii), where $p = 1$ is sufficient. Vice versa, we have (iii) $\Rightarrow$ (v) if $(X_n)_{n\in\mathbb{N}}$ is uniformly L^p-integrable, i.e. if

$$\sup_{n\in\mathbb{N}} \int_{\{|X_n|>M\}} |X_n|^p \,\mathrm{d}\mathbb{P} \xrightarrow{M\to\infty} 0$$

Finally, there holds (ii) $\Rightarrow$ (v) if there exist an L^p-majorant for $(X_n)_{n\in\mathbb{N}}$, as we will see in Theorem 5.24. ◀

Fact 5.21. The sequence of random variables $(X_n)_{n\in\mathbb{N}}$ converges in distribution to X if and only if the corresponding distribution functions F_n satisfy

$$F_n(x) \xrightarrow{n\to\infty} F(x)$$

for all $x \in \mathbb{R}$ at which the distribution function F of X is continuous. ◁

We have the following result known as monotone convergence theorem.

Theorem 5.22 (Monotone Convergence Theorem)**.** Let $(X_n)_{n\in\mathbb{N}}$, $X_n \geq 0$, be a monotonically increasing sequence of random variables. Then, there holds

$$\mathbb{E}\left[\sup_{n\in\mathbb{N}} X_n\right] = \sup_{n\in\mathbb{N}} \mathbb{E}[X_n] \tag{5.28}$$

◁

Proof. For every $n \in \mathbb{N}$, we have $X_n = \sup_{m\in\mathbb{N}} X_{n,m}$ for some monotonically increasing sequence $(X_{n,m})_{m\in\mathbb{N}}$ of non-negative simple functions by Lemma 5.3. Moreover,

$$\sup_{n\in\mathbb{N}} X_n = \sup_{n\in\mathbb{N}} \left(\sup_{m\in\mathbb{N}} X_{n,m} \right) = \sup_{n,m\in\mathbb{N}} X_{n,m}$$

If we define $Y_k := \max\{X_{n,m} : 1 \leq n, m \leq k\}$ then $(Y_k)_{k\in\mathbb{N}}$ is a monotonically increasing sequence of non-negative simple functions and

$$\sup_{k\in\mathbb{N}} Y_k = \sup_{k\in\mathbb{N}} \max_{1\leq n,m\leq k} X_{n,m} = \sup_{n,m\in\mathbb{N}} X_{n,m} = \sup_{n\in\mathbb{N}} X_n$$

We have

$$\mathbb{E}[X_n] \leq \mathbb{E}\left[\sup_{n\in\mathbb{N}} X_n\right] = \mathbb{E}\left[\sup_{k\in\mathbb{N}} Y_k\right] \overset{(5.5)}{=} \sup_{k\in\mathbb{N}} \mathbb{E}[Y_k] \leq \sup_{k\in\mathbb{N}} \mathbb{E}[X_k]$$

for every $n \in \mathbb{N}$. The left inequality follows by monotonicity of the integral (Lemma 5.5 (iii)). For the right inequality, observe that $Y_k \leq X_k$ for every $k \in \mathbb{N}$, since $X_{n,m} \leq X_n \leq X_k$ whenever $1 \leq n, m \leq k$. Thus, again by Lemma 5.5 (iii), $\mathbb{E}[Y_k] \leq \mathbb{E}[X_k]$ for every $k \in \mathbb{N}$.

Taking the supremum over $n \in \mathbb{N}$ on the left-hand side and renaming the (dummy) index on the right-hand side yields

$$\sup_{n\in\mathbb{N}} \mathbb{E}[X_n] \leq \mathbb{E}\left[\sup_{n\in\mathbb{N}} X_n\right] \leq \sup_{n\in\mathbb{N}} \mathbb{E}[X_n]$$

and hence the claimed identity. □

The next result is Fatou's lemma, which is a consequence of the Monotone Convergence Theorem 5.22.

Theorem 5.23 (Fatou's lemma)**.** Let $(X_n)_{n\in\mathbb{N}}$, $X_n \geq 0$, be a sequence of random variables. Then, there holds

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n] \tag{5.29}$$

◁

Proof. Let $Y_n := \inf_{k \geq n} X_k$. Then $(Y_n)_{n \in \mathbb{N}}$ is a monotonically increasing sequence of random variables and we have

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \overset{(3.2)}{=} \mathbb{E}\left[\sup_{n\in\mathbb{N}} Y_n\right] \overset{(5.28)}{=} \sup_{n\in\mathbb{N}} \mathbb{E}[Y_n] \leq \sup_{n\in\mathbb{N}} \inf_{k\geq n} \mathbb{E}[X_k] \overset{(3.2)}{=} \liminf_{n\to\infty} \mathbb{E}[X_n]$$

since, by Lemma 5.5 (iii), $\mathbb{E}[Y_n] \leq \mathbb{E}[X_k]$ for all $k \geq n$. □

Theorem 5.24 (Dominated Convergence Theorem)**.** Let $p \in [1, \infty)$ and let $X_n \xrightarrow{\text{a.s.}} X$ as well as $|X_n| \leq Y$ $\mathbb{P}$-almost surely for all $n \in \mathbb{N}$ and some $Y \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$. Then, $X_n \xrightarrow{L^p} X$. ◁

Proof. Since $|X_n| \leq Y$ for all $n \in \mathbb{N}$, we have $|X| \leq Y$ $\mathbb{P}$-almost surely. Consequently, there also holds $\mathbb{E}[|X|^p] \leq \mathbb{E}[Y^p] < \infty$. Set $Z_n := |X_n - X|^p$. We need to show $\mathbb{E}[Z_n] \to 0$. There holds

$$0 \leq Z_n \leq (|X| + Y)^p =: Z$$

with $\mathbb{E}[Z] < \infty$. Invoking Theorem 5.23, we have

$$\mathbb{E}\left[\liminf_{n\to\infty}(Z - Z_n)\right] \overset{(5.29)}{\leq} \liminf_{n\to\infty} \mathbb{E}[Z - Z_n] = \mathbb{E}[Z] - \limsup_{n\to\infty} \mathbb{E}[Z_n]$$

Furthermore, since $X_n \xrightarrow{n\to\infty} X$ $\mathbb{P}$-almost surely, we have that $Z_n \xrightarrow{n\to\infty} 0$ $\mathbb{P}$-almost surely and, consequently,

$$\mathbb{E}\left[\liminf_{n\to\infty}(Z - Z_n)\right] = \mathbb{E}[Z]$$

Inserting this in the previous inequality yields $\limsup_{n\to\infty} \mathbb{E}[Z_n] \leq 0$ and consequently

$$\lim_{n\to\infty} \mathbb{E}[Z_n] = 0$$

since $Z_n \geq 0$. □

6 Laws of Large Numbers

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space (Definition 1.7).

6.1 Weak Law of Large Numbers

Theorem 6.1 (Weak Law of Large Numbers)**.** Let $(X_n)_{n\in\mathbb{N}}$ with $X_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}[X_n] = \mu$ for all $n \in \mathbb{N}$. Assume they are pairwise uncorrelated, i.e. $\mathbb{E}[(X_j - \mu)(X_k - \mu)] = 0$ for all $j \neq k$. Assume that $\mathbb{V}\mathrm{ar}[X_n] \leq C$ for all $n \in \mathbb{N}$ and some $C > 0$. Then, for $S_n := \sum_{k=1}^n X_k$, there holds

$$\frac{S_n}{n} \xrightarrow{n\to\infty} \mu$$

in L^2 and in probability, see Definition 5.12 (v) and (iii). ◁

Proof.

$$\begin{aligned}
\left\|\frac{S_n}{n} - \mu\right\|_2 &\overset{(5.26)}{=} \sqrt{\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right]} \\
&= \sqrt{\mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^n (X_k - \mu)\right)^2\right]} \\
&= \sqrt{\frac{1}{n^2}\,\mathbb{E}\left[\sum_{j=1}^n\sum_{k=1}^n (X_j - \mu)(X_k - \mu)\right]} \\
&= \sqrt{\frac{1}{n^2}\sum_{j=1}^n\sum_{k=1}^n \mathbb{E}[(X_j - \mu)(X_k - \mu)]} \\
&= \sqrt{\frac{1}{n^2}\sum_{k=1}^n \underbrace{\mathbb{E}[(X_k - \mu)^2]}_{=\mathbb{V}\mathrm{ar}[X_k]\leq C}} \\
&\leq \sqrt{\frac{C}{n}} \xrightarrow{n\to\infty} 0
\end{aligned}$$

and hence $S_n/n \xrightarrow{L^2} \mu$. According to Remark 5.11, this implies $S_n/n \xrightarrow{\text{i.p.}} \mu$. □

Remark 6.1. Instead of $\mathbb{V}\mathrm{ar}[X_n] \leq C$, it is sufficient that $\sum_{k=1}^n \mathbb{V}\mathrm{ar}[X_k] = o(n^2)$.

In case that $\mathbb{E}[X_n]$ is not constant, we have $\frac{1}{n}\sum_{k=1}^n (X_k - \mathbb{E}[X_k]) \xrightarrow{L^2} 0$. ◀

A consequence of Theorem 6.1 is the approximation theorem of Weierstraß, Theorem 6.2, which states that polynomials are dense in the space of continuous functions on $[0, 1]$ with respect to the supremum norm.

Theorem 6.2 (Weierstraß approximation)**.** Let $f \in C([0, 1])$. Then for every $\varepsilon > 0$, there exists a polynomial p on $[0, 1]$ such that $|f(x) - p(x)| < \varepsilon$ for all $x \in [0, 1]$. ◁

Proof. Define the Bernstein polynomials according to

$$B_{k,n}(t) := \binom{n}{k} t^k (1 - t)^{n-k}$$

for $n \in \mathbb{N}$, $k = 0, \ldots, n$ and $t \in [0, 1]$.

$$f_n(t) := \sum_{k=0}^n f(k/n) B_{k,n}(t) = \sum_{k=0}^n f(k/n) \binom{n}{k} t^k (1 - t)^{n-k}$$

For fixed $t \in [0, 1]$, let $\mathbb{P}_{t,j}$ be the probability measure on $\{0, 1\}$ given by

$$\mathbb{P}_{t,j}(\{x_j\}) := t^{x_j}(1 - t)^{1-x_j}$$

for $x_j \in \{0, 1\}$ and $j = 1, \ldots, n$. We equip $\{0, 1\}^n$ with the product measure $\mathbb{P}_t = \bigotimes_{j=1}^n \mathbb{P}_{t,j}$ (Example 2.5), given by

$$\mathbb{P}_t\left(\bigtimes_{j=1}^n \{x_j\}\right) \overset{(2.2)}{=} \prod_{j=1}^n \mathbb{P}_{t,j}(\{x_j\}) = \prod_{j=1}^n t^{x_j}(1 - t)^{1-x_j}$$

for $(x_1, \dots, x_n) \in \{0,1\}^n$. Let $X_1, \dots, X_n$ be the coordinate maps on $\{0,1\}^n$ and set $S_n := \sum_{j=1}^n X_j$. $\mathbb{E}_t$ then denotes expectation with respect to $\mathbb{P}_t$, i.e.

$$\mathbb{E}_t[g] \overset{(5.12)}{=} \int_{\{0,1\}^n} g \, \mathrm{d}\mathbb{P}_t = \sum_{\omega \in \{0,1\}^n} g(\omega)\mathbb{P}_t(\{\omega\}) = \sum_{\omega \in \{0,1\}^n} g(\omega) \prod_{j=1}^n t^{X_j(\omega)}(1-t)^{1-X_j(\omega)}$$

for every $g : \{0,1\}^n \to \mathbb{R}$. Setting $g = f(S_n/n)$, we have

$$\begin{aligned}
\mathbb{E}_t[f(S_n/n)] &= \sum_{\omega \in \{0,1\}^n} f(S_n(\omega)/n) \prod_{j=1}^n t^{X_j(\omega)}(1-t)^{1-X_j(\omega)} \\
&= \sum_{k=0}^n \sum_{\substack{\omega \in \{0,1\}^n \\ S_n(\omega)=k}} f(S_n(\omega)/n) \prod_{j=1}^n t^{X_j(\omega)}(1-t)^{1-X_j(\omega)} \\
&= \sum_{k=0}^n \sum_{\substack{\omega \in \{0,1\}^n \\ S_n(\omega)=k}} f(k/n)\ t^k(1-t)^{n-k} \\
&= \sum_{k=0}^n \binom{n}{k} f(k/n)\ t^k(1-t)^{n-k} \\
&= f_n(t)
\end{aligned}$$

We have

$$|f_n(t) - f(t)| = \left|\mathbb{E}_t\left[f\left(\frac{1}{n}S_n\right) - f(t)\right]\right| \overset{(5.24)}{\leq} \mathbb{E}_t\left[\left|f\left(\frac{1}{n}S_n\right) - f(t)\right|\right]$$

Now let $\varepsilon > 0$. Since f is continuous on the compact interval $[0,1]$, it is uniformly continuous by the Heine–Cantor theorem. Hence, there exists $\delta > 0$ such that $|f(t) - f(s)| < \varepsilon/2$ whenever $|t - s| < \delta$. Splitting the expectation according to whether $|S_n/n - t| < \delta$ ("good" event) or $|S_n/n - t| \geq \delta$ ("bad" event) and employing the $\triangle$-inequality, we have

$$\begin{aligned}
\mathbb{E}_t\left[\left|f\left(\frac{1}{n}S_n\right) - f(t)\right|\right] &\leq \frac{\varepsilon}{2}\mathbb{P}_t\left(\left|\frac{1}{n}S_n - t\right| < \delta\right) + \max_{s\in[0,1]}(|f(s) - f(t)|)\mathbb{P}_t\left(\left|\frac{1}{n}S_n - t\right| \geq \delta\right) \\
&\leq \frac{\varepsilon}{2} + \max_{s\in[0,1]}(|f(s)| + |f(t)|)\mathbb{P}_t\left(\left|\frac{1}{n}S_n - t\right| \geq \delta\right) \\
&\leq \frac{\varepsilon}{2} + 2\|f\|_\infty \mathbb{P}_t\left(\left|\frac{1}{n}S_n - t\right| \geq \delta\right)
\end{aligned}$$

Since $X_1, \dots, X_n$ are independent under $\mathbb{P}_t$ and $\mathbb{E}_t[X_j] = t$, $\mathbb{V}\mathrm{ar}_t[X_j] = t(1-t)$ we have $\mathbb{E}_t[\frac{1}{n}S_n] = t$ and

$$\mathbb{V}\mathrm{ar}_t\left[\frac{1}{n}S_n\right] = \frac{1}{n^2}\sum_{j=1}^n \mathbb{V}\mathrm{ar}_t[X_j] = \frac{t(1-t)}{n}$$

and applying Chebyshev's inequality (Corollary 5.20), we obtain

$$\mathbb{P}_t\left(\left|\frac{1}{n}S_n - t\right| \geq \delta\right) \leq \frac{1}{\delta^2}\,\mathbb{V}\mathrm{ar}_t\left[\frac{1}{n}S_n\right] = \frac{1}{\delta^2}\frac{t(1-t)}{n} \leq \frac{1}{4\delta^2 n} \xrightarrow{n\to\infty} 0$$

Choose $N \in \mathbb{N}$ such that $\frac{2\|f\|_\infty}{4\delta^2 N} \leq \frac{\varepsilon}{2}$. Then, for all $n \geq N$ and all $t \in [0,1]$,

$$|f_n(t) - f(t)| \leq \frac{\varepsilon}{2} + \frac{2\|f\|_\infty}{4\delta^2 n} \leq \varepsilon$$

Thus $f_n \to f$ uniformly on $[0,1]$. □

A practical application of the Weak Law of Large Numbers (Theorem 6.1) is the Monte Carlo method.

Theorem 6.3 (Monte Carlo method). Let $f \in \mathcal{L}^2([0,1], \mathcal{B}([0,1]), \lambda)$ and let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and uniformly $[0,1]$-distributed random variables. Then

$$\frac{1}{n}\sum_{k=1}^n f(X_k) \xrightarrow{L^2} \int_0^1 f(x)\,\mathrm{d}x \tag{6.1}$$

◁

Proof. Since f is measurable, the random variables $(f(X_n))_{n\in\mathbb{N}}$ are pairwise independent by Theorem 4.4. As mentioned in Example 4.2, this implies pairwise uncorrelatedness. They are also identically distributed. Moreover, by Corollary 5.11 and since $\mathbb{P}_{X_n} = \lambda|_{[0,1]}$,

$$\mathbb{E}[f(X_n)] \overset{(5.12)}{=} \int_\Omega f(X_n)\,\mathrm{d}\mathbb{P} \overset{(5.18)}{=} \int_{[0,1]} f\,\mathrm{d}\lambda \overset{(5.9)}{=} \int_0^1 f(x)\,\mathrm{d}x =: \mu$$

as well as

$$\mathbb{E}[f(X_n)^2] \overset{(5.12)}{=} \int_\Omega f(X_n)^2\,\mathrm{d}\mathbb{P} \overset{(5.18)}{=} \int_{[0,1]} f^2\,\mathrm{d}\lambda \overset{(5.9)}{=} \int_0^1 f(x)^2\,\mathrm{d}x < \infty$$

because $f \in \mathcal{L}^2([0,1],\mathcal{B}([0,1]),\lambda)$. Hence

$$\mathbb{V}\mathrm{ar}[f(X_n)] = \mathbb{E}[f(X_n)^2] - (\mathbb{E}[f(X_n)])^2 \le \mathbb{E}[f(X_n)^2] =: C < \infty$$

is uniformly bounded (i.e. bounded by the constant C independent of n). Thus the assumptions of Theorem 6.1 are satisfied and we obtain (6.1). □

Remark 6.2. The proof works in arbitrary dimensions and with general distributions, as long as f is bounded in L^2. The Monte Carlo method is therefore applied as a dimnsion-robust quadratur method, with dimension independent rate of convergence $\mathcal{O}(1/\sqrt{n})$. ◀

6.2 Strong Law of Large Numbers

Theorem 6.4 (Strong Law of Large Numbers)**.** Let $(X_n)_{n\in\mathbb{N}}$ be pairwise independent and identically distributed with $\mu = \mathbb{E}[X_n]$ and $\mathbb{E}[|X_n|] < \infty$. Then, for $S_n := \sum_{k=1}^n X_k$, there holds

$$\frac{S_n}{n} \xrightarrow{n\to\infty} \mu$$

$\mathbb{P}$-almost surely, see Definition 5.12 (ii). ◁

An important consequence of the Strong Law of Large Numbers is the convergence of empirical distributions, known as the Glivenko–Cantelli theorem (Theorem 6.6).

Definition 6.1. Let $(\mu_n)_{n\in\mathbb{N}}$ and μ be probability measures on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. We say that μ_n converges weakly to μ, denoted by

$$\mu_n \xrightarrow{\text{weakly}} \mu$$

if

$$\int_\mathbb{R} f\,\mathrm{d}\mu_n \xrightarrow{n\to\infty} \int_\mathbb{R} f\,\mathrm{d}\mu$$

for all bounded $f \in C(\mathbb{R})$. ◀

Remark 6.3. Weak convergence of probability measures (Definition 6.1) is the measure-theoretic formulation of convergence in distribution of random variables (Definition 5.12). Indeed, by (iv) and Corollary 5.11, there holds

$$Y_n \xrightarrow{\text{i.d.}} Y$$

if and only if we have

$$\mathbb{P}_{Y_n} \xrightarrow{\text{weakly}} \mathbb{P}_Y$$

for the image measures. ◀

The corresponding formulation of Fact 5.21 is

Fact 6.5. Let $(\mu_n)_{n\in\mathbb{N}}$ and μ be probability measures on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Let F_{μ_n} and F_μ be the distribution functions (Definition 2.10) of μ_n and μ, respectively. Then

$$\mu_n \xrightarrow{\text{weakly}} \mu$$

if and only if

$$F_{\mu_n}(x) \xrightarrow{n\to\infty} F_\mu(x)$$

for every $x \in \mathbb{R}$ at which F_μ is continuous. ◁

Theorem 6.6 (Glivenko–Cantelli)**.** Let $(X_n)_{n\in\mathbb{N}}$ be independent and identically distributed random variables. Define the empirical distribution via

$$\rho_n(\omega, A) := \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k(\omega)}(A) = \frac{1}{n}|\{k \in \{1,\dots,n\} : X_k(\omega) \in A\}|$$

which is the relative frequency with which the first n observations fall into A, where $A \in \mathcal{B}(\mathbb{R})$.

Then, we have

$$\rho_n(\omega,\cdot) \xrightarrow{n\to\infty} \mathbb{P}_{X_1}(\cdot)$$

weakly for $\mathbb{P}$-almost every $\omega \in \Omega$. This means that the empirical distribution $\rho_n(\omega,\cdot)$ converges weakly to the common distribution of the random variables $(X_n)_{n\in\mathbb{N}}$. ◁

Proof. We prove the pointwise convergence of the sequence of *empirical distribution functions*

$$F_n(\omega, x) = \rho_n(\omega, (-\infty, x]) = \frac{1}{n}\sum_{k=1}^{n} \underbrace{\mathbf{1}_{(-\infty,x]}(X_k(\omega))}_{=:Y_k(\omega)}$$

→ for fixed $x \in \mathbb{R}$, this is a random variable

towards

$$F(x) = \mathbb{P}_{X_1}((-\infty, x]) = \mathbb{P}(\{X_1 \le x\})$$

for every $x \in \mathbb{R}$ where F is continuous and then apply Fact 6.5.

For fixed $x \in \mathbb{R}$, the random variables Y_n are independent and identically distributed. Since $0 \le Y_n \le 1$, we have $\mathbb{E}[|Y_n|] = \mathbb{E}[Y_n] \le \mathbb{E}[1] = 1 < \infty$. Thus, we may apply the Strong Law of Large Numbers and obtain

$$F_n(\omega, x) \xrightarrow{n\to\infty} \mathbb{E}[Y_1] = F(x) \tag{6.2}$$

$\mathbb{P}$-almost surely, see Definition 5.12 (ii). The equality on the RHS of (6.2) holds since

$$\mathbb{E}[Y_1] = \mathbb{E}[\mathbf{1}_{(-\infty,x]}(X_1)] \overset{(5.12)}{=} \int_\Omega \mathbf{1}_{(-\infty,x]}(X_1)\,\mathrm{d}\mathbb{P} = \int_\Omega \mathbf{1}_{\{X_1\le x\}}\,\mathrm{d}\mathbb{P} \overset{(5.1)}{=} \mathbb{P}(\{X_1 \le x\}) = F(x)$$

for each fixed $x \in \mathbb{R}$.

In particular, we find for each $x \in \mathbb{R}$ a set $N_x \in \mathcal{F}$ with $\mathbb{P}(N_x) = 0$ such that (6.2) holds for all $\omega \in \Omega \setminus N_x$. The problem is that the exceptional set N_x may depend on x.

Now letting $N := \bigcup_{x\in\mathbb{Q}} N_x$, there holds $\mathbb{P}(N) \le \sum_{x\in\mathbb{Q}} \mathbb{P}(N_x) = 0$, since $\mathbb{Q}$ is countable. Therefore, for every $\omega \in \Omega \setminus N$, the convergence in (6.2) holds simultaneously for all $x \in \mathbb{Q}$ (not just at continuity points of F). We find for all $\omega \in \Omega \setminus N$ by the monotonicity of probability distribution functions that

$$F(s) = \lim_{n\to\infty} F_n(\omega, s) \le \liminf_{n\to\infty} F_n(\omega, x) \le \limsup_{n\to\infty} F_n(\omega, x) \le \lim_{n\to\infty} F_n(\omega, t) = F(t)$$

for all $x \in \mathbb{R}$ and all $s, t \in \mathbb{Q}$ with $s \le x \le t$. Therefore, if F is continuous at x, we have

$$\lim_{s\nearrow x} F(s) = F(x) = \lim_{t\searrow x} F(t)$$

and, consequently, by the previous inequality, $\lim_{n\to\infty} F_n(\omega, x) = F(x)$ for every $\omega \in \Omega \setminus N$.

Thus, for every $\omega \in \Omega \setminus N$, the distribution functions of $\rho_n(\omega,\cdot)$ converge to the distribution function of $\mathbb{P}_{X_1}$ at every continuity point of F. By Fact 6.5, this is equivalent to

$$\rho_n(\omega,\cdot) \xrightarrow{\text{weakly}} \mathbb{P}_{X_1}$$

for every $\omega \in \Omega \setminus N$. Since $\mathbb{P}(N) = 0$, the convergence holds for $\mathbb{P}$-almost every $\omega \in \Omega$ and hence the claim follows. □

7 Central Limit Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space (Definition 1.7).

Definition 7.1. Let X be a real-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. The moment generating function of X is defined by

$$M_X(t) := \mathbb{E}\big[e^{tX}\big] = \int_\Omega e^{tX}\,\mathrm{d}\mathbb{P} \tag{7.1}$$

for all $t \in \mathbb{R}$ such that $\mathbb{E}\big[e^{tX}\big] < \infty$. Its domain is $D_X := \{t \in \mathbb{R} : \mathbb{E}\big[e^{tX}\big] < \infty\}$. ◀

Example 7.1. If $|X| \le M$ $\mathbb{P}$-almost surely, then the moment generating function exists on $\mathbb{R}$, since

$$\mathbb{E}\big[e^{tX}\big] \le \mathbb{E}\big[e^{|tX|}\big] = \mathbb{E}\big[e^{|t||X|}\big] \le \mathbb{E}\big[e^{|t|M}\big] = e^{|t|M} < \infty$$

for all $t \in \mathbb{R}$. ◀

If X admits a Lebesgue density f_X in the sense of Definition 5.7 and $t \in D_X$, then $e^{t\,\mathrm{id}_\mathbb{R}}$ is non-negative and $\mathbb{P}_X$-integrable. Therefore, by Corollary 5.11, and Theorem 5.15, applied with $\mu = \lambda^1$, $\nu = \mathbb{P}_X$ and $g = e^{t\,\mathrm{id}_\mathbb{R}}$, there holds

$$M_X(t) \overset{(7.1)}{=} \int_\Omega e^{tX}\,\mathrm{d}\mathbb{P} \overset{(5.19)}{=} \int_\mathbb{R} e^{t\,\mathrm{id}_\mathbb{R}}\,\mathrm{d}\mathbb{P}_X \overset{(5.23)}{=} \int_\mathbb{R} e^{t\,\mathrm{id}_\mathbb{R}} f_X\,\mathrm{d}\lambda^1 \overset{(5.8)}{=} \int_\mathbb{R} e^{tx} f_X(x)\,\mathrm{d}x \tag{7.2}$$

Lemma 7.1 (Properties of moment generating functions)**.** Let $X : \Omega \to \mathbb{R}$ be a random variable with moment generating function M_X defined in an open interval $D_X \subset \mathbb{R}$ containing 0.

(i) The k-th derivative $\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t)$ exists on D_X for all $k \in \mathbb{N}$ and

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(0) = \mathbb{E}[X^k] \tag{7.3}$$

(ii) If $X, Y : \Omega \to \mathbb{R}$ are independent with moment generating functions M_X, M_Y on D and $Z := X + Y$, then there holds $M_Z(t) = M_X(t)M_Y(t)$ for all $t \in D$.

(iii) Let $a, b \in \mathbb{R}$ and set $Y := a + bX$. Then, $M_Y(t) = e^{at} M_X(bt)$ for all t such that $bt \in D_X$. ◁

Proof. (i) Since M_X is finite on an open interval around t, the terms $|X|^k e^{sX}$ are dominated for s near t by an integrable exponential bound. Hence, the mean value theorem applied to the difference quotients, together with repeated applications of Theorem 5.24, yields

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} M_X(t) = \mathbb{E}\left[\frac{\partial^k}{\partial t^k} e^{tX}\right] = \mathbb{E}\big[X^k e^{tX}\big]$$

for all $t \in D_X$. Evaluation at $t = 0$ yields (7.3).

(ii) The independence of X and Y yields

$$M_Z(t) = \mathbb{E}\big[e^{tZ}\big] = \mathbb{E}\big[e^{tX+tY}\big] = \mathbb{E}\big[e^{tX}e^{tY}\big] \overset{(5.21)}{=} \mathbb{E}\big[e^{tX}\big]\,\mathbb{E}\big[e^{tY}\big] = M_X(t)M_Y(t)$$

where the fourth equality is the factorization of expectations of non-negative measurable functions of independent random variables, which follows from Definition 4.2 and Fact 5.12.

(iii) By direct computation, we infer

$$M_Y(t) = \mathbb{E}\big[e^{tY}\big] = \mathbb{E}\big[e^{t(a+bX)}\big] = \mathbb{E}\big[e^{at}e^{btX}\big] = e^{at}\,\mathbb{E}\big[e^{btX}\big] = e^{at} M_X(bt)$$ □

Remark 7.2. If $M_X(t)$ exists and is finite on some open interval containing 0, then the distribution $\mathbb{P}_X$ is uniquely determined by the moments $\mathbb{E}[X^k]$, $k \in \mathbb{N}$. ◀

Example 7.3 (Normal distribution)**.** Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma > 0$, i.e. let X admit the Lebesgue density $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$. Then there holds

$$M_X(t) \overset{(7.2)}{=} \frac{1}{\sqrt{2\pi}\sigma} \int_\mathbb{R} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = e^{\mu t + \frac{\sigma^2 t^2}{2}} \frac{1}{\sqrt{2\pi}\sigma} \int_\mathbb{R} e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}}\,\mathrm{d}x = e^{\mu t + \frac{\sigma^2 t^2}{2}} \tag{7.4}$$

for all $t \in \mathbb{R}$. Hence X has moments of all orders by Lemma 7.1 (i), and its distribution is uniquely determined by those moments by Remark 7.2. ◀

Fact 7.2. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables with moment generating functions $(M_n)_{n\in\mathbb{N}}$, which all are finite on some open interval $D_X \subset \mathbb{R}$ containing 0. Suppose there exists a random variable X with moment generating function M_X, also defined on D_X, such that

$$\lim_{n\to\infty} M_n(t) = M_X(t)$$

for all $t \in D_X$ (see Definition 0.2). Then $X_n \xrightarrow{\text{i.d.}} X$ in the sense of Definition 5.12 (iv). ◁

We now prove a version of the central limit theorem for i.i.d. random variables whose moment generating function exists in an open interval around zero.

Theorem 7.3 (Central Limit Theorem)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$. Assume that the moment generating function $M(t) := \mathbb{E}\big[e^{tX_1}\big]$ exists and is finite on an open interval containing 0. Define

$$S_n := \sum_{k=1}^{n} X_k \qquad \text{and} \qquad Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Then $(Z_n)_{n\in\mathbb{N}}$ converges in distribution to the standard normal distribution. ◁

Proof. Without loss of generality, we assume $\mu = 0$ and set $Z_n := S_n/(\sigma\sqrt{n})$.

By Lemma 7.1 (ii), the independence of $(X_n)_{n\in\mathbb{N}}$, and the fact that the random variables are identically distributed, there holds

$$M_{S_n}(t) = \prod_{k=1}^{n} M_{X_k}(t) = (M(t))^n$$

Using Lemma 7.1 (iii), we further infer

$$M_{Z_n}(t) = M_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(M\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n \tag{7.5}$$

Choose $a > 0$ such that $(-a, a)$ is contained in the domain of M and set $J := (-\sigma a, \sigma a)$. Then the moment generating functions M_{Z_n} are finite on the common open interval J for all $n \in \mathbb{N}$, since $t/(\sigma\sqrt{n}) \in (-a, a)$ for all $t \in J$ and all $n \in \mathbb{N}$.

Since M is finite in a neighbourhood of 0, its derivatives at 0 exist and by Lemma 7.1 (i), we have

$$M(0) = \mathbb{E}\big[e^{0\cdot X_1}\big] = 1 \qquad M'(0) = \mathbb{E}[X_1] = 0 \qquad M''(0) = \mathbb{E}[X_1^2] \overset{(5.25)}{=} \sigma^2$$

and the second-order Taylor expansion of M around 0 therefore takes the form

$$M(s) = 1 + \frac{\sigma^2}{2}s^2 + r(s) \tag{7.6}$$

where $r(s) = o(s^2)$ as $s \to 0$. In particular, for fixed $t \in J$,

$$M\left(\frac{t}{\sigma\sqrt{n}}\right) \overset{(7.6)}{=} 1 + \frac{\sigma^2}{2}\frac{t^2}{\sigma^2 n} + r\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + r_n(t)$$

where $r_n(t) := r\big(t/(\sigma\sqrt{n})\big)$ satisfies $nr_n(t) \to 0$ as $n \to \infty$, i.e. $r_n(t) = o(1/n)$. Indeed, this is immediate for $t = 0$, while for $t \neq 0$, setting $s_n := t/(\sigma\sqrt{n})$ gives

$$nr_n(t) = \frac{t^2}{\sigma^2}\frac{r(s_n)}{s_n^2} \xrightarrow{n\to\infty} 0$$

Inserting this into (7.5) and taking the logarithm (which is well-defined since the MGF is strictly positive) gives

$$\log M_{Z_n}(t) = n\log\left(1 + \frac{t^2}{2n} + r_n(t)\right)$$

Since $\log(1+u) = u + o(u)$ as $u \to 0$, letting $u_n(t) := t^2/(2n) + r_n(t)$ and observing that $u_n(t) = O(1/n)$, we obtain

$$\log M_{Z_n}(t) = n\left(\frac{t^2}{2n} + r_n(t) + o\left(\frac{1}{n}\right)\right) = \frac{t^2}{2} + nr_n(t) + o(1) \xrightarrow{n\to\infty} \frac{t^2}{2}$$

Therefore, by the continuity of the exponential function, we arrive at

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty} e^{\log M_{Z_n}(t)} = e^{\lim_{n\to\infty}\log M_{Z_n}(t)} = e^{t^2/2}$$

for every $t \in J$. By Example 7.3 (7.4), this is the MGF of a standard normal random variable $Z \sim \mathcal{N}(0, 1)$. By Fact 7.2, we conclude that $(Z_n)_{n\in\mathbb{N}}$ converges to Z in distribution. □