Probability and Measure

Stefan Grosskinsky

Cambridge, Michaelmas 2006

These notes and other information about the course are available on
`www.statslab.cam.ac.uk/~stefan/teaching/probmeas.html`

The text is based on – and partially copied from – notes on the same course by James Norris, Alan Stacey and Geoffrey Grimmett.

Contents

Introduction **3**

1 Set systems and measures **5**
1.1 Set systems 5
1.2 Measures 6
1.3 Extension and uniqueness 8
1.4 Lebesgue(-Stieltjes) measure 10
1.5 Independence and Borel-Cantelli lemmas 13

2 Measurable Functions and Random Variables **15**
2.1 Measurable Functions 15
2.2 Random Variables 17
2.3 Convergence of measurable functions 19

3 Integration **22**
3.1 Definition and basic properties 22
3.2 Integrals and limits 25
3.3 Integration in $\mathbb{R}$ and differentiation 28
3.4 Product measure and Fubini's theorem 30

4 L^p-spaces **33**
4.1 Norms and inequalities 33
4.2 L^2 as a Hilbert space 36
4.3 Convergence in L^1 39

5 Characteristic functions and Gaussian random variables **41**
5.1 Definitions 41
5.2 Properties of characteristic functions 42
5.3 Gaussian random variables 43

6 Ergodic theory and sums of random variables **46**
6.1 Motivation 46
6.2 Measure-preserving transformations 46
6.3 Ergodic Theorems 49
6.4 Limit theorems for sums of random variables 51

Appendix A – Example sheets **53**
A.1 Example sheet 1 – Set systems and measures 53
A.2 Example sheet 2 – Measurable functions and integration 55
A.3 Example sheet 3 – Convergence, Fubini, L^p-spaces 57
A.4 Example sheet 4 – Characteristic functions, Gaussian rv's, ergodic theory . . . 59

Appendix B – Hand-outs **61**
B.1 Hand-out 1 – Proof of Carathéodory's extension theorem 61
B.2 Hand-out 2 – Convergence of random variables 63
B.3 Hand-out 3 – Connection between Lebesgue and Riemann integration 65
B.4 Hand-out 4 – Ergodic theorems 66

Introduction

Motivation from two perspectives:

1. Probability
Let $\big(\Omega, \mathcal{P}(\Omega), \mathbb{P}\big)$ be a probability space, where Ω is a set, $\mathcal{P}(\Omega)$ the set of events (power set in this case) and $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ is the probability measure.
If Ω is countable then we have for every $A \in \mathcal{P}(\Omega)$

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}\big(\{\omega\}\big) \ .$$

So calculating probabilities just involves (possibly infinite) sums.
If $\Omega = [0,1]$ and $\mathbb{P}$ is the uniform probability measure on $[0,1]$ then for every $\omega \in \Omega$ it is $\mathbb{P}(\omega) = 0$. So

$$1 = \mathbb{P}\big([0,1]\big) \neq \text{“} \sum_{w \in [0,1]} \mathbb{P}\big(\{\omega\}\big) \text{“}.$$

2. Integration (Analysis)
When $\mathbb{P}$ can be described by a density $\rho : \Omega \to [0,\infty)$ we can handle the situation via

$$\mathbb{P}(A) = \int_A \rho(x)\,dx = \int_\Omega \mathbb{1}_A(x)\,\rho(x)\,dx \ , \qquad (*)$$

where $\mathbb{1}_A$ is the indicator function of the set A. In the example above $\rho(x) \equiv 1$ and this leads to $\mathbb{P}\big([a,b]\big) = \int_0^1 \mathbb{1}_{[a,b]}(x)\,dx = \int_a^b dx = b - a$.
In general this approach only makes sense if the integral $(*)$ exists. Using the theory of Riemann-integration we are fine as long as A is a finite union or intersection of intervals and $\rho(x)$ is e.g. continuous. But e.g. for $A = [0,1] \cap \mathbb{Q}$, the Riemann-integral $\int_0^1 \mathbb{1}_A(x)\,dx$ is not defined, although the probability for this event is intuitively 0.
Moreover, since A is countable, it can be written as $A = \big\{a_n \ : \ n \in \mathbb{N}\big\}$. Define $f_n := \mathbb{1}_{\{a_1,\ldots,a_n\}}$ with $f_n \to \mathbb{1}_A$ for $n \to \infty$. For every n, f_n is Riemann-integrable and $\int_0^1 f_n(x)\,dx = 0$. So it should be the case that

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 0 \stackrel{?}{=} \int_0^1 \mathbb{1}_A(x)\,dx \ ,$$

but the latter integral is not defined. Thus the concept of Riemann-integrals is not satisfactory for two reasons: The set of Riemann-integrable functions is not closed, and there are “too many” functions which are not Riemann-integrable.

Goals of this course

- Generalisation of Riemann-integration to Lebesgue-integration using measure theory, involving a precise treatment of sets A and functions ρ for which $(*)$ is defined
- Using measure theory as the basis of advanced probability and discussing applications in that area

Official schedule

Measure spaces, σ-algebras, π-systems and uniqueness of extension, statement * and proof * of Carathéodory's extension theorem. Construction of Lebesgue measure on $\mathbb{R}$. The Borel σ-algebra of $\mathbb{R}$. Existence of non-measurable subsets of $\mathbb{R}$. Lebesgue-Stieltjes measures and probability distribution functions. Independence of events, independence of σ-algebras. The Borel-Cantelli lemmas. Kolmogorov's zero-one law. [6]

Measurable functions, random variables, independence of random variables. Construction of the integral, expectation. Convergence in measure and convergence almost everywhere. Fatou's lemma, monotone and dominated convergence, differentiation under the integral sign. Discussion of product measure and statement of Fubini s theorem. [6]

Chebyshev s inequality, tail estimates. Jensen's inequality. Completeness of L^p for $1 \leq p \leq \infty$. The Hölder and Minkowski inequalities, uniform integrability. [4]

L^2 as a Hilbert space. Orthogonal projection, relation with elementary conditional probability. Variance and covariance. Gaussian random variables, the multivariate normal distribution. [2]

The strong law of large numbers, proof for independent random variables with bounded fourth moments. Measure preserving transformations, Bernoulli shifts. Statements * and proofs * of maximal ergodic theorem and Birkhoff s almost everywhere ergodic theorem, proof of the strong law. [4]

The Fourier transform of a finite measure, characteristic functions, uniqueness and inversion. Weak convergence, statement of Lévy's convergence theorem for characteristic functions. The central limit theorem. [2]

Appropriate books

P. Billingsley, *Probability and Measure*. Wiley 1995 (hardback).
R.M. Dudley, *Real Analysis and Probability*. CUP 2002 (paperback).
R.L.Schilling, *Measures, Integrals and Martingales*. CUP 2005 (paperback).
R.T. Durrett, *Probability: Theory and Examples*. Wadsworth a. Brooks/Cole 1991 (hardback).
D. Williams, *Probability with Martingales*. CUP (paperback).

From the point of view of analysis, the first chapters of this book might be interesting:
S. Kantorovitz, *Introduction to Modern Analysis*. Oxford 2003 (hardback).

Non-examinable material

- proof of Carathéodory's extension theorem on hand-out 1
- part (a) of the proof of Skorohod's representation theorem on hand-out 2
- connection between Lebesgue and Riemann integration on hand-out 3
- Proof of the maximal ergodic lemma and Birkhoff's almost everywhere ergodic theorem on hand-out 4 and in Section 6

1 Set systems and measures

Let E be an arbitrary set and $\mathcal{E} \subseteq \mathcal{P}(E)$ a set of subsets. To define a measure $\mu : \mathcal{E} \to [0, \infty)$ (see section 1.2) we first need to identify a proper domain of definition.

1.1 Set systems

Definition 1.1. Say that $\mathcal{E}$ is a *ring*, if for all $A, B \in \mathcal{E}$

(i) $\emptyset \in \mathcal{E}$ (ii) $B \setminus A \in \mathcal{E}$ (iii) $A \cup B \in \mathcal{E}$.

Say that $\mathcal{E}$ is an *algebra* (or *field*), if for all $A, B \in \mathcal{E}$

(i) $\emptyset \in \mathcal{E}$ (ii) $A^c = E \setminus A \in \mathcal{E}$ (iii) $A \cup B \in \mathcal{E}$.

Say that $\mathcal{E}$ is a *σ-algebra* (or *σ-field*), if for all A and $A_1, A_2, \ldots \in \mathcal{E}$

(i) $\emptyset \in \mathcal{E}$ (ii) $A^c \in \mathcal{E}$ (iii) $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}$.

Properties. (i) A (σ-)algebra is closed under (countably) finitely many set operations, since

$$A \cap B = \big(A^c \cup B^c\big)^c \in \mathcal{E} \,, \qquad \bigcap_{n\in\mathbb{N}} A_n = \Big(\bigcup_{n\in\mathbb{N}} A_n^c\Big)^c ,^{1}$$

$$A \setminus B = A \cap B^c \in \mathcal{E} \,, \qquad A \triangle B = (A \setminus B) \cup (B \setminus A) \in \mathcal{E} \,.$$

(ii) Thus: $\mathcal{E}$ is a σ-algebra $\Rightarrow$ $\mathcal{E}$ is an algebra $\Rightarrow$ $\mathcal{E}$ is a ring
In general the inverse statementss are false, but in special cases they hold:
$\Leftarrow$ (if E is finite) $\Leftarrow$ (if $E \in \mathcal{E}$)

Examples. (i) $\mathcal{P}(E)$ and $\{\emptyset, E\}$ are the largest and smallest σ-algebras on E, respectively

(ii) $E = \mathbb{R}$, $\mathcal{R} = \Big\{ \bigcup_{i=1}^n (a_i, b_i] \,:\, a_i < b_i,\, i = 1, \ldots, n,\, n \in \mathbb{N}_0 \Big\}$ is the *ring of half-open intervals* ($\emptyset$ is given by the empty intersection $n = 0$).
$\mathcal{R}$ is an algebra if we allow for infinite intervals and identify $\mathbb{R} = (-\infty, \infty]$.

(iii) **Beware:** $\Big\{ \bigcup_{i=1}^\infty (a_i, b_i] \,:\, a_i, b_i \in [-\infty, \infty],\, a_i < b_i,\, i \in \mathbb{N}, \Big\}$ is a **not** a σ-algebra.
(see problem 1.9(a))

Lemma 1.1. *Let $\big\{\mathcal{E}_i \,:\, i \in I\big\}$ be a (possibly uncountable) collection of σ-algebras. Then $\bigcap_{i\in I} \mathcal{E}_i$ is a σ-algebra, whereas $\bigcup_{i\in I} \mathcal{E}_i$ in general is not.*

Proof. Let $\mathcal{E} = \bigcap_{i\in I} \mathcal{E}_i$. We check (i) to (iii) in the above definition:
(i) Since $\emptyset \in \mathcal{E}_i$ for all $i \in I$, $\emptyset \in \mathcal{E}$. (ii) Since for all $A \in \mathcal{E}$, $A^c \in \mathcal{E}_i$ for all $i \in I$, $A^c \in \mathcal{E}$.
(iii) Let $A_1, A_2, \ldots \in \mathcal{E}$. Then $A_k \in \mathcal{E}_i$ for all $k \in \mathbb{N}$ and $i \in I$, hence $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}_i$ for each $i \in I$ and so $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}$. For the second part see problem 1.1 (c). □

Definition 1.2. Let $\mathcal{A} \subseteq \mathcal{P}(E)$. Then the *$\sigma$-algebra generated by* $\mathcal{A}$ is

$$\sigma(\mathcal{A}) := \bigcap_{\substack{\mathcal{E} \supset \mathcal{A} \\ \mathcal{E}\ \sigma-\mathrm{alg.}}} \mathcal{E} \,, \qquad \text{the smallest } \sigma\text{-algebra containing } \mathcal{A}.$$

[1]as long as (ii) is fulfilled "∪" and "∩" are equivalent

Remarks. (i) If $\mathcal{E}$ is a σ-algebra, then $\sigma(\mathcal{E}) = \mathcal{E}$.

(ii) Let $\mathcal{A}_1,\, \mathcal{A}_2 \subseteq \mathcal{P}(E)$ with $\mathcal{A}_1 \subseteq \mathcal{A}_2$. Then $\sigma(\mathcal{A}_1) \subseteq \sigma(\mathcal{A}_2)$.

Examples. (i) Let $\emptyset \neq A \subsetneq E$. Then $\sigma\big(\{A\}\big) = \{\emptyset, E, A, A^c\}$.

(ii) If $\mathcal{E}$ is a σ-algebra, so is $A \cap \mathcal{E} = \{A \cap B \,:\, B \in \mathcal{E}\}$ for each $A \in E$, called the *trace σ-algebra* of A.

The next example is so important, that we spend an extra definition.

Definition 1.3. Let (E, τ) be a *topological space* with *topology* $\tau \subseteq \mathcal{P}(E)$ (set of open sets)[2]. Then $\sigma(\tau)$ is called the *Borel σ-algebra* of E, denoted by $\mathcal{B}(E)$. $A \in \mathcal{B}(E)$ is called a *Borel set*. One usually denotes $\mathcal{B}(\mathbb{R}) = \mathcal{B}$.

Lemma 1.2. *Let $E = \mathbb{R}$, $\mathcal{R}$ the ring of half-open intervals and $\mathcal{I} = \big\{(a, b] \,:\, a < b\big\}$ the set of all half-open intervals. Then $\sigma(\mathcal{R}) = \sigma(\mathcal{I}) = \mathcal{B}$.*

Proof. (i) $\mathcal{I} \subseteq \mathcal{R} \;\Rightarrow\; \sigma(\mathcal{I}) \subseteq \sigma(\mathcal{R})$. On the other hand, each $A \in \mathcal{R}$ can be written as $A = \bigcup_{i=1}^n (a_i, b_i] \in \sigma(\mathcal{I})$. Thus $\mathcal{R} \subseteq \sigma(\mathcal{I}) \;\Rightarrow\; \sigma(\mathcal{R}) \subseteq \sigma(\mathcal{I})$.
(ii) Each $A \in \mathcal{I}$ can be written as $A = (a, b] = \bigcap_{n=1}^\infty (a, b + \frac{1}{n}) \in \mathcal{B} \;\Rightarrow\; \sigma(\mathcal{I}) \subseteq \mathcal{B}$.
Let $A \subseteq \mathbb{R}$ be open, i.e. $\forall\, x{\in}A\; \exists\, \epsilon_x{>}0 \,:\, (x - \epsilon_x, x + \epsilon_x) \subseteq A$. Thus

$$\forall\, x \in A\;\; \exists\, a_x, b_x \in \mathbb{Q} \,:\, \{x\} \subseteq (a_x, b_x] \subseteq A\;.$$

Then $A = \bigcup_{x \in A} (a_x, b_x]$ which is a countable union, since $a_x, b_x \in \mathbb{Q}$.
Thus $A \in \sigma(\mathcal{I}) \;\Rightarrow\; \mathcal{B} \subseteq \sigma(\mathcal{I})$. □

Remarks. (i) The Borel σ-algebra on $A \subseteq \mathbb{R}$ is $\mathcal{B}(A) = A \cap \mathcal{B}$ (trace σ-algebra of A).

(ii) Analogously, $\mathcal{B}(\mathbb{R}^d)$ is generated by $\mathcal{I}^d = \Big\{\prod_{i=1}^d (a_i, b_i] \,:\, a_i < b_i,\, 1 = 1, \ldots, d\Big\}$ and this is consistent with $d = 1$ in the sense that $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}^d$.

Definition 1.4. Let $\mathcal{E} \subseteq \mathcal{P}(E)$ be a σ-algebra. The pair $(E, \mathcal{E})$ is a *measurable space* and elements of $\mathcal{E}$ are *measurable sets*.

If E is finite or countably infinite, one usually takes $\mathcal{E} = \mathcal{P}(E)$ as relevant σ-algebra.

1.2 Measures

Definition 1.5. Let $\mathcal{E}$ be a ring on E. A *set function* is any $\mu : \mathcal{E} \to [0, \infty]$ with $\mu(\emptyset) = 0$.

- μ is called *additive* if for all $A, B \in \mathcal{E}$ with $A \cap B = \emptyset$: $\quad \mu(A \cup B) = \mu(A) + \mu(B)$.
- μ is called *countably additive* (or *σ-additive)* if for all sequences $(A_n)_{n \in \mathbb{N}}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\displaystyle\dot\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$: $\quad \mu\Big(\dot\bigcup_{n \in \mathbb{N}} A_n\Big) = \sum_{n \in \mathbb{N}} \mu(A_n)$.

[2]Having a direct definition of open sets for $E = \mathbb{R}^d$, there is also an axiomatic definition of a topology, namely (i) $\emptyset \in \tau$ and $E \in \tau$ (ii) $\forall\, A, B \in \tau \,:\, A \cap B \in \tau$ (iii) $\bigcup_{i \in I} A_i \in \tau$, given $A_i \in \tau$ for all $i \in I$

Note. μ countably additive $\Rightarrow$ μ additive $\quad \big[A_1 = A, A_2 = B, A_3 = A_4 = \ldots = \emptyset\big]$

Definition 1.6. Let $(E, \mathcal{E})$ be a measurable space. A countably additive set function $\mu : \mathcal{E} \to [0, \infty]$ is called a *measure*, the triple $(E, \mathcal{E}, \mu)$ is called *measure space*.
If $\mu(E) < \infty$, μ is called *finite*. If $\mu(E) = 1$, μ is a *probability measure* and $(E, \mathcal{E}, \mu)$ is a *probability space*. If E is a topological space and $\mathcal{E} = \mathcal{B}(E)$, then μ is called *Borel measure*.

Basic properties. Let $(E, \mathcal{E}, \mu)$ be a measure space.

(i) μ is *non-decreasing*: For all $A, B \in \mathcal{E}$, $A \subseteq B$ it is $\mu(B) = \mu(B \setminus A) + \mu(A) \geq \mu(A)$.
(Note: The version $\mu(B \setminus A) = \mu(B) - \mu(A)$ only makes sense if $\mu(A) < \infty$).

(ii) μ is *subadditive*: For all $A, B \in \mathcal{E}$, $\quad \mu(A \cup B) \leq \mu(A) + \mu(B) \quad$ since

$$\mu(A) + \mu(B) = \underbrace{\mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A)}_{=\mu(A \cup B)} + \mu(A \cap B) = \\ = \mu(A \cup B) + \mu(A \cap B) \geq \mu(A \cup B)\,.$$

(Again: $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ only if $\mu(A \cap B) < \infty$.)

(iii) μ is also *countably subadditive* (see problem 1.6 (b)).

(iv) Let $\mathcal{E}_1 \subseteq \mathcal{E}_2$ be σ-algebras. If μ is a measure on $\mathcal{E}_2$, then it is also on $\mathcal{E}_1$.

(v) For $A \in \mathcal{E}$ the *restriction* $\mu|_A = \mu(. \cap A)$ is a measure on $(E, \mathcal{E})$.

Remark. These properties also hold for countably additive set functions (called *pre-measures*) on a ring , (i) and (ii) also for additive set functions on a ring.

Examples. (i) For every $x \in E$, the *Dirac measure* is given by $\quad \delta_x(A) = \begin{cases} 1 \ , \ x \in A \\ 0 \ , \ x \notin A \end{cases}$.

(ii) Discrete measure theory:
Let E be countable. Every measure μ on $(E, \mathcal{P}(E))$ can be characterized by a *mass function* $m : E \to [0, \infty]$,

$$\mu = \sum_{x \in E} m(x)\, \delta_x \quad \text{or equivalently} \quad \forall\, A \subseteq E \ : \ \mu(A) = \sum_{x \in A} \mu\big(\{x\}\big) = \sum_{x \in A} m(x)\,.$$

If $m(x) \equiv 1$ for all $x \in E$, μ is called *counting measure*.

(iii) Let $E = \mathbb{R}$ and $\mathcal{R}$ be the ring of half-open intervals. For $A \in \mathcal{R}$ write $A = \dot{\bigcup}_{i=1}^n (a_i, b_i]$ with disjoint intervals, $n \in \mathbb{N}_0$. We call this a *standard representation* of A. Although it is not unique, the set function $\mu : \mathcal{R} \to [0, \infty]$ with $\mu(A) := \sum_{i=1}^n (b_i - a_i)$ is independent of the particular representation and thus well defined.
Further, μ is additive and *translation invariant*, i.e. $\forall\, x \in \mathbb{R} \ : \ \mu(A + x) = \mu(A)$, where $A + x := \{x + y \ : \ y \in A\}$. The key question is:

Can μ be extended to a measure on $\mathcal{B} = \sigma(\mathcal{R})$?

In order to attack this question in the next subsection, it is useful to introduce the following property of set functions.

Definition 1.7. Let $\mathcal{E}$ be a ring on E and $\mu : \mathcal{E} \to [0,\infty]$ an additive set function. μ is *continuous at* $A \in \mathcal{E}$, if

(i) μ is *continuous from below*:
given any $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ in $\mathcal{E}$ with $\bigcup_{n\in\mathbb{N}} A_n = A \in \mathcal{E} \quad (A_n \nearrow A)$,
then $\lim_{n\to\infty} \mu(A_n) = \mu(A)$

(ii) μ is *continuous from above*:
given any $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$ in $\mathcal{E}$ with $\bigcap_{n\in\mathbb{N}} A_n = A \in \mathcal{E} \quad (A_n \searrow A)$ and $\mu(A_n) < \infty$ for some $n \in \mathbb{N}$, then $\lim_{n\to\infty} \mu(A_n) = \mu(A)$

Lemma 1.3. *Let $\mathcal{E}$ be a ring on E and $\mu : \mathcal{E} \to [0,\infty]$ an additive set function. Then:*

(i) μ is countably additive $\quad\Rightarrow\quad$ μ is continuous at all $A \in \mathcal{E}$

(ii) μ is continuous from below at all $A \in \mathcal{E}$ $\quad\Rightarrow\quad$ μ is countably additive

(iii) μ is cont. from above at $\emptyset$ and $\mu(A) < \infty$ for all $A \in \mathcal{E}$ $\quad\Rightarrow\quad$ μ is countably additive

Remark. The condition $\mu(A_n) < \infty$ for some $n \in \mathbb{N}$ in (ii) of the definition is necessary for measures to be continuous. Consider e.g. $E = \mathbb{N}$, $A_k = \{k, k+1, \ldots\}$ and μ the counting measure. Then $\mu(A_k) = \infty$ for all $k \in \mathbb{N}$, but $\mu(A) = \mu(\emptyset) = 0$.

Proof. (i) Given $A_n \nearrow A$ in $\mathcal{E}$, then $A = (A_1 \setminus A_0) \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \ldots \quad (A_0 = \emptyset)$

$$\Rightarrow \quad \mu(A) = \sum_{n=0}^{\infty} \mu(A_{n+1} \setminus A_n) = \lim_{m\to\infty} \mu\Big(\dot{\bigcup_{n=0}^{m-1}} (A_{n+1} \setminus A_n)\Big) = \lim_{m\to\infty} \mu(A_m) \ .$$

Given $A_n \searrow A$ in $\mathcal{E}$ and $\mu(A_m) < \infty$ for some $m \in \mathbb{N}$. Let $B_n := A_m \setminus A_n$ for $n \geq m$. Then $B_n \nearrow (A_m \setminus A)$ for $n \to \infty$ and thus, following the above,

$$\mu(A_m) - \mu(A_n) = \mu(B_n) \stackrel{n\to\infty}{\longrightarrow} \mu(A_m \setminus A) = \mu(A_m) - \mu(A) \ .$$

Since $\mu(A_m) < \infty$ this implies $\lim_{n\to\infty} \mu(A_n) = \mu(A)$.
(ii) see problem 1.6 (a)
(iii) analogous to (i) and (ii) □

1.3 Extension and uniqueness

Theorem 1.4. Carathéodory's extension theorem
Let $\mathcal{E}$ be a ring on E and $\mu : \mathcal{E} \to [0,\infty]$ be a countably additive set function. Then there exists a measure μ' on $\big(E, \sigma(\mathcal{E})\big)$ such that $\quad \mu'(A) = \mu(A) \quad$ for all $A \in \mathcal{E}$.

Proof. The proof is not examinable and is given on Hand-out 1 in the appendix.

To formulate a result on uniqueness two further notions are useful.

Definition 1.8. Let E be a set. $\mathcal{E} \subseteq \mathcal{P}(E)$ is called a π-*system* if $\quad \forall\, A, B \in \mathcal{A} \ : \ A \cap B \in \mathcal{A}$.
$\mathcal{E}$ is called a d-*system* if

(i) $E \in \mathcal{E}$, (ii) $\forall\, A, B \in \mathcal{E},\ A \subseteq B \ : \ B \setminus A \in \mathcal{E}$,
(iii) $\forall\, A_1, A_2, \ldots \in \mathcal{E} \ : \ A_1 \subseteq A_2 \subseteq \ldots \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$.

Remarks.

(i) The set $\mathcal{I} \cup \{\emptyset\} = \big\{(a, b] \ : \ a < b\big\} \cup \{\emptyset\}$ is a π-system on $\mathbb{R}$ and we have shown in Lemma 1.2 that $\sigma(\mathcal{I}) = \mathcal{B}$.

(ii) $\mathcal{E}$ is a σ-algebra $\Leftrightarrow$ $\mathcal{E}$ is a π- and a d-system (see problem 1.8 (a)).

Lemma 1.5. Dynkin's π-system lemma
Let $\mathcal{E}$ be a π-system. Then for any d-system $\mathcal{D} \supseteq \mathcal{E}$ it is $\mathcal{D} \supseteq \sigma(\mathcal{E})$.

Proof. The intersection $\Delta(\mathcal{E}) = \bigcap_{\mathcal{D} \supset \mathcal{E}} \mathcal{D}$ of all d-systems containing $\mathcal{E}$ is itself a d-system. We shall show that $\Delta(\mathcal{E})$ is also a π-system. Then it is also a σ-algebra and for any d-system $\mathcal{D} \supseteq \mathcal{E}$ we have $\mathcal{D} \supseteq \Delta(\mathcal{E}) \supseteq \sigma(\mathcal{E})$, thus proving the lemma. Consider

$$\mathcal{D}' = \big\{B \in \Delta(\mathcal{E}) \ : \ B \cap A \in \Delta(\mathcal{E}) \text{ for all } A \in \mathcal{E}\big\} \subseteq \Delta(\mathcal{E}) \ .$$

Then $\mathcal{E} \subseteq \mathcal{D}'$ because $\mathcal{E}$ is a π-system. We check that $\mathcal{D}'$ is a d-system, and hence $\mathcal{D}' = \Delta(\mathcal{E})$.
(i) clearly $E \in \mathcal{D}'$;
(ii) suppose $B_1, B_2 \in \mathcal{D}'$ with $B_1 \subseteq B_2$, then for $A \in \mathcal{E}$ we have

$$(B_2 \setminus B_1) \cap A = (B_2 \cap A) \setminus (B_1 \cap A) \in \Delta(\mathcal{E}) \ ,$$

because $\Delta(\mathcal{E})$ is a d-system, so $B_2 \setminus B_1 \in \mathcal{D}'$;
(iii) finally, if $B_n \in \mathcal{D}'$ and $B_n \nearrow B$, then for $A \in \mathcal{E}$

$$B_n \cap A \nearrow B \cap A \in \Delta(\mathcal{E}) \quad \Rightarrow \quad B \in \mathcal{D}' \ .$$

Now consider

$$\mathcal{D}'' = \big\{B \in \Delta(\mathcal{E}) \ : \ B \cap A \in \Delta(\mathcal{E}) \text{ for all } A \in \Delta(\mathcal{E})\big\} \subseteq \mathcal{D}' \ .$$

Then $\mathcal{E} \subseteq \mathcal{D}''$ because $\mathcal{D}' = \Delta(\mathcal{E})$. We can check that $\mathcal{D}''$ is a d-system, just as we did for $\mathcal{D}'$. Hence $\mathcal{D}'' = \Delta(\mathcal{E})$ which shows that $\Delta(\mathcal{E})$ is a π-system. $\square$

Theorem 1.6. Uniqueness of extension
Let $\mathcal{E} \subseteq \mathcal{P}(E)$ be a π-system. Suppose that μ_1, μ_2 are measures on $\sigma(\mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$. If $\mu_1 = \mu_2$ on $\mathcal{E}$ then $\mu_1 = \mu_2$ on $\sigma(\mathcal{E})$.
Equivalently, if $\mu(E) < \infty$ the measure μ on $\sigma(\mathcal{E})$ is uniquely determined by its values on the π-system $\mathcal{E}$.

Proof. Consider $\mathcal{D} = \big\{A \in \sigma(\mathcal{E}) \,:\, \mu_1(A) = \mu_2(A)\big\} \subseteq \sigma(\mathcal{E})$. By hypothesis, $E \in \mathcal{D}$. For $A, B \in \mathcal{D}$ with $A \subseteq B$ we have

$$\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A) < \infty \;,$$

thus also $B \setminus A \in \mathcal{D}$. If $A_n \in \mathcal{D}$, $n \in \mathbb{N}$, with $A_n \nearrow A$, then

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A_n) = \lim_{n\to\infty} \mu_2(A_n) = \mu_2(A) \quad \Rightarrow \quad A \in \mathcal{D} \;.$$

Thus $\mathcal{D} \subseteq \sigma(\mathcal{E})$ is a d-system containing the π-system $\mathcal{E}$, so $\mathcal{D} = \sigma(\mathcal{E})$ by Dynkin's lemma. □

These theorems provide general tools for the construction and characterisation of measures and will be applied in a specific context in the next subsection.

1.4 Lebesgue(-Stieltjes) measure

Theorem 1.7. *There exists a unique Borel measure μ on $(\mathbb{R}, \mathcal{B})$ such that*

$$\mu\big((a,b]\big) = b - a \;, \quad \textit{for all } a, b \in \mathbb{R} \textit{ with } a < b.$$

The measure μ is called *Lebesgue measure* on $\mathbb{R}$.

Proof. (Existence) Let $\mathcal{R}$ be the ring of half-open intervals. Consider the set function $\mu(A) = \sum_{i=1}^n (b_i - a_i)$, where $A = \biguplus_{i=1}^n (a_i, b_i]$, $n \in \mathbb{N}_0$. We aim to show that μ is countably additive on $\mathcal{R}$, which then proves existence by Carathéodory's extension theorem.
Since $\mu(A) < \infty$ for all $A \in \mathcal{R}$, by Lemma 1.3 (iii) it suffices to show that μ is continuous from above at $\emptyset$. Suppose not. Then there exists $\epsilon > 0$ and $A_n \searrow \emptyset$ with $\mu(A_n) \geq 2\epsilon$ for all n. For each n we can find $C_n \in \mathcal{R}$ with $\overline{C}_n \subseteq A_n$ and $\mu(A_n \setminus C_n) \leq \epsilon 2^{-n}$ (see problem 1.9 (b)). Then

$$\mu\big(A_n \setminus (C_1 \cap \ldots \cap C_n)\big) \leq \sum_{k=1}^{n} \mu(A_n \setminus C_k) \leq \sum_{k=1}^{n} \mu(A_k \setminus C_k) \leq \sum_{k=1}^{\infty} \epsilon 2^{-k} = \epsilon \;,$$

and since $\mu(A_n) \geq 2\epsilon$ we have $\mu(C_1 \cap \ldots \cap C_n) \geq \epsilon$ and in particular $C_1 \cap \ldots \cap C_n \neq \emptyset$. Thus $K_n = \overline{C}_1 \cap \ldots \cap \overline{C}_n$, $n \in \mathbb{N}$ is a monotone sequence of compact non-empty sets. Thus there exists a sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in K_n$ which has at least one accumulation point x^*, since all $x_n \in K_1$ which is compact. Since $K_n \searrow$, $x^* \in \bigcap_{n\in\mathbb{N}} K_n$. Thus $\emptyset \neq \bigcap_{n\in\mathbb{N}} K_n \subseteq \bigcap_{n\in\mathbb{N}} A_n$ which is a contradiction to $A_n \searrow \emptyset$.
(Uniqueness) For each $n \in \mathbb{Z}$ define

$$\mu_n(A) := \mu\big((n, n+1] \cap A\big) \quad \text{for all } A \in \mathcal{B} \;.$$

Then μ_n is a probability measure on $(\mathbb{R}, \mathcal{B})$, so, by Theorem 1.6, μ_n is uniquely determinded by its values on the π-system $\mathcal{I}$ generating $\mathcal{B}$. Since $\mu(A) = \sum_{n\in\mathbb{Z}} \mu_n(A)$ it follows that μ is also uniquely determined. □

Definition 1.9. $A \subseteq \mathbb{R}$ is called *null* if $A \subseteq B \in \mathcal{B}$ with $\mu(B) = 0$. Denote by $\mathcal{N}$ the set of all null sets. Then $\mathcal{L} = \big\{B \cup N \,:\, b \in \mathcal{B},\, N \in \mathcal{N}\big\}$ is the *completion* of $\mathcal{B}$ and is called *Lebesgue σ-algebra*. The sets in $\mathcal{L}$ are called *Lebesgue-measurable sets* or *Lebesgue sets*.

Remark. The Lebesgue measure μ can be extended to $\mathcal{L}$ via (see problem 1.10)

$$\lambda(B \cup N) := \mu(B) \quad \text{for all} \quad B \in \mathcal{B},\ N \in \mathcal{N}\ .$$

In some books only λ is called *Lebesgue measure*. This makes sense e.g. in analysis, where one usually works with a fixed measure space $(\mathbb{R}, \mathcal{L}, \lambda)$.

Theorem 1.8. $\mathcal{B} \subsetneq \mathcal{L} \subsetneq \mathcal{P}(\mathbb{R})$. *Moreover*[3]

$$\mathrm{card}(\mathcal{B}) = \mathrm{card}(\mathbb{R}) = c \quad \textit{whereas} \quad \mathrm{card}(\mathcal{L}) = \mathrm{card}\big(\mathcal{P}(\mathbb{R})\big) = 2^c\ .$$

Proof. (i) According to problem 1.4, $\mathcal{B}$ is separable, i.e. generated by a countable set system $\mathcal{E}$. Thus $\mathrm{card}(\mathcal{B}) \leq \mathrm{card}\big(\mathcal{P}(\mathcal{E})\big) = c$. On the other hand, $\{x\} \in \mathcal{B}$ for all $x \in \mathbb{R}$ and thus $\mathrm{card}(\mathcal{B}) = c$. With problem 1.10 the Cantor set $C \subseteq \mathbb{R}$ is uncountable and thus $\mathrm{card}(C) = c$. Since also $\mu(C) = 0$, with Definition 1.9 we have $\mathcal{P}(C) \subseteq \mathcal{L}$ and thus $\mathrm{card}(\mathcal{L}) = \mathrm{card}\big(\mathcal{P}(\mathbb{R})\big) = 2^c > c$.
(ii) Using the axiom of choice we construct a subset of $U = [0, 1]$ which is not in $\mathcal{L}$:
Define the equivalence relation $\sim$ on U by $x \sim y$ if $x - y \in \mathbb{Q}$.
Write $\{E_i : i \in I\}$ for the equivalence classes of $\sim$ and let $R = \{e_i : i \in I\}$ be a collection of representatives $e_i \in E_i$, chosen by the axiom of choice. Then U can be partitioned

$$U = \dot{\bigcup_{i\in I}} E_i = \dot{\bigcup_{i\in I}} \dot{\bigcup_{q\in\mathbb{Q}\cap[0,1)}} (e_i + q) = \dot{\bigcup_{q\in\mathbb{Q}\cap[0,1)}} \underbrace{\dot{\bigcup_{i\in I}}(e_i + q)}_{=R+q} = \dot{\bigcup_{q\in\mathbb{Q}\cap[0,1)}} (R + q)\ ,$$

where $+$ is to be understood modulo 1. Suppose $R \in \mathcal{L}$, then $R + q \in \mathcal{L}$ and $\lambda(R) = \lambda(R + q)$ for all $q \in \mathbb{Q}$ by translation invariance of λ. But by countable additivity of λ this means

$$\sum_{q\in\mathbb{Q}\cap[0,1)} \lambda(R) = \lambda(U) = 1\ ,$$

which leads to a contradiction for either $\lambda(R) = 0$ or $\lambda(R) > 0$. Thus $R \notin \mathcal{L}$. □

Remarks. (i) Every set of positive measure has non-measurable subsets and, moreover:

$$\mathcal{P}(A) \subseteq \mathcal{L} \quad \Leftrightarrow \quad \lambda(A) = 0\ .$$

(ii) There exists no translation invariant, countably additive set function on $\mathcal{P}(\mathbb{R})$ with $\mu\big([0, 1]\big) \in (0, \infty)$.

Definition 1.10. $F : \mathbb{R} \to \mathbb{R}$ is called *distribution function* if

(i) F is non-decreasing (ii) F is right-continuous, i.e. $\lim_{x \searrow x_0} F(x) = F(x_0)$.

F is a *probability distribution function* if in addition

(iii) $\lim_{x\to\infty} F(x) = 1\ ,\quad \lim_{x\to-\infty} F(x) = 0\ .$

[3]Notation: The *cardinality* of a countable set (such as $\mathbb{N}$ or $\mathbb{Q}$) is $\aleph_0$, for a continuous set (such as $\mathcal{P}(\mathbb{N})$, $\mathbb{R}$ or the Cantor set C) it is $2^{\aleph_0} = c$. For power sets of the latter we just write $\mathrm{card}\big(\mathcal{P}(\mathbb{R})\big) = 2^c$.

Proposition 1.9. *Let μ be a Radon measure*[4] *on $(\mathbb{R}, \mathcal{B})$. Then for every $r \in \mathbb{R}$*

$$F_r(x) := \begin{cases} \mu\big((r,x]\big) & , \ x > r \\ -\mu\big((x,r]\big) & , \ x \leq r \end{cases} \qquad \big(\textit{in particular } F_r(r) = -\mu(\emptyset) = 0\big)$$

is a distriburion function with $\mu\big((a,b]\big) = F_r(b) - F_r(a)$, *$a < b$* .
Also $F_r + C$ is a distribution function with that property for all $C \in \mathbb{R}$.
The F_r differ only in an additive constant, namely $F_r(x) = F_0(x) - F_0(r)$ *for all $r \in \mathbb{R}$.*

Proof. $F_r \nearrow$ by monotonicity of μ and $\mu\big((a,b]\big) = F_r(b) - F_r(a)$ for $b > a$ by definition. For $x_n \searrow x > r$ it is $F_r(x_n) = \mu\big((r,x_n]\big) \searrow \mu\big((r,x]\big) = F_r(x)$ by continuity of measures. For $x_n \searrow r$ we have $(r,x_n] \searrow \emptyset$ such that $F_r(x_n) \searrow \mu(\emptyset) = 0 = F_r(r)$.
The F_r differ only in a constant, since for all $x > r$

$$\mu(\emptyset) = \mu\big((r,x]\big) - \mu\big((r,x]\big) = F_0(x) - F_0(r) - F_r(x) + F_r(r) = 0 \ ,$$

and $F_r(r) = 0 = F_0(r) - F_0(r)$. Both statements follow analogously for $x < 0$. □

Remarks. (i) The distribution functions for the Lebesgue measure are $F_r(x) = x + r$, $r \in \mathbb{R}$.

(ii) Note that for $x_n \nearrow x$ we have $(x_n, x] \nearrow \{x\} \neq \emptyset$, so that F_r as defined in Proposition 1.9 is in general not left-continuous.

(iii) If μ is a probability measure one usually uses the *cumulative distribution function*

$$CDF_\mu(x) := F_{-\infty}(x) = \mu\big((-\infty, x]\big) \ .$$

E.g. for the Dirac measure δ_a concentrated in $a \in \mathbb{R}$, $CDF_{\delta_a}(x) = \begin{cases} 0 \ , \ x < a \\ 1 \ , \ x \geq a \end{cases}$.

On the other hand, a distribution function uniquely determines a Radon measure.

Theorem 1.10. *Let $F : \mathbb{R} \to \mathbb{R}$ be a distribution function. Then there exists a unique Radon measure μ_F on $(\mathbb{R}, \mathcal{B})$, such that*

$$\mu_F\big((a,b]\big) = F(b) - F(a) \quad \textit{for all } a < b \ .$$

The measure μ_F is called the *Lebesgue-Stieltjes measure* of F.

Proof. As for Lebesgue measure, the set function μ_F is well defined on the ring $\mathcal{R}$ via

$$\mu(A) := \sum_{i=1}^{n} \big(F(b_i) - F(a_i)\big) \quad \text{where} \quad A = \dot{\bigcup_{i=1}^{n}} (a_i, b_i] \ .$$

The proof is then the same as that of Theorem 1.7 for Lebesgue measure. □

Remark. Analogous to Definition 1.9, $\mathcal{B}$ can also be completed with respect to the Lebesgue-Stieltjes measure μ_F. However, the completion $\mathcal{L}_{\mu_F}$ depends on the measure μ_F. Although $\mathcal{L}$ is much larger than $\mathcal{B}$ (see Theorem 1.8), it is therefore preferable to work with the measure space $(\mathbb{R}, \mathcal{B})$, since it is defined independent of the measure and all μ_F will have the same domain of definition.

[4] i.e. $\mu(K) < \infty$ for $K \in \mathcal{B}$ compact

1.5 Independence and Borel-Cantelli lemmas

Let $(E, \mathcal{E}, \mathbb{P})$ be a probability space. It provides a model for an experiment whose outcome is random. E describes the set of possible outcomes, $\mathcal{E}$ the set of events (observable sets of outcomes) and $\mathbb{P}(A)$ is the probability of an event $A \in \mathcal{E}$.

Definition 1.11. The events $(A_i)_{i\in I}$, $A_i \in \mathcal{E}$, are said to be *independent* if

$$\mathbb{P}\Big(\bigcap_{i\in J} A_i\Big) = \prod_{i\in J} \mathbb{P}(A_i) \quad \text{for all finite, nonempty } J \subseteq I \ .$$

The σ-algebras $(\mathcal{E}_i)_{i\in I}$, $\mathcal{E}_i \subseteq \mathcal{E}$ are said to be *independent* if the events $(A_i)_{i\in I}$ are independent for any choice $A_i \in \mathcal{E}_i$.

A useful way to establish independence of two σ-algebra is given below.

Theorem 1.11. *Let $\mathcal{E}_1,\ \mathcal{E}_2 \subseteq \mathcal{E}$ be π-systems and suppose that*

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2) \quad \textit{whenever } A_1 \in \mathcal{E}_1,\ A_2 \in \mathcal{E}_2 \ .$$

Then $\sigma(\mathcal{E}_1)$ and $\sigma(\mathcal{E}_2)$ are independent.

Proof. Fix $A_1 \in \mathcal{E}_1$ and define the measures μ, ν by

$$\mu(A) = \mathbb{P}(A_1 \cap A)\ , \quad \nu(A) = \mathbb{P}(A_1)\,\mathbb{P}(A) \quad \text{for all } A \in \mathcal{E} \ .$$

μ and ν agree on the π-system $\mathcal{E}_2$ with $\mu(E) = \nu(E) = \mathbb{P}(A_1) < \infty$. So, by uniqueness of extension (Theorem 1.6), for all $A_1 \in \mathcal{E}_1$ and $A_2 \in \sigma(\mathcal{E}_2)$

$$\mathbb{P}(A_1 \cap A_2) = \mu(A_2) = \nu(A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2) \ .$$

Now fix $A_2 \in \sigma(\mathcal{E}_2)$ and repeat the same argument with

$$\mu'(A) := \mathbb{P}(A \cap A_2)\ , \quad \nu'(A) := \mathbb{P}(A)\,\mathbb{P}(A_2)$$

to show that for all $A_1 \in \sigma(\mathcal{E}_1)$ we have $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)$. □

Remark. In particular, the σ-algebras $\sigma\big(\{A_1\}\big)$ and $\sigma\big(\{A_2\}\big)$ generated by single events are independent if and only if A_1 and A_2 are independent.

Background. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. Then $\lim\limits_{n\to\infty} a_n$ does not necessarily exist, e.g. for $a_n = (-1)^n$. To nevertheless study asymptotic properties of $(a_n)_{n\in\mathbb{N}}$ consider

$$\underline{a}_n = \inf_{k\geq n} a_k \quad \text{and} \quad \overline{a}_n = \sup_{k\geq n} a_k \ .$$

Then $\underline{a}_n \nearrow$ and $\overline{a}_n \searrow$ are monotone and both have limits in $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Define

$$\liminf_{n\to\infty} a_n := \lim_{n\to\infty} \inf_{k\geq n} a_k \quad \text{and} \quad \limsup_{n\to\infty} a_n := \lim_{n\to\infty} \sup_{k\geq n} a_k \ ,$$

which are equal to the smallest and largest accumulation point of $(a_n)_{n\in\mathbb{N}}$, respectively. In general $\liminf\limits_{n\to\infty} a_n \leq \limsup\limits_{n\to\infty} a_n$ since $\underline{a}_m \leq a_{m\vee n} \leq \overline{a}_n$ for all $m, n \in \mathbb{N}$. They may be different, as e.g. $\liminf_{n\to\infty}(-1)^n = -1 < 1 = \limsup_{n\to\infty}(-1)^n$. Both are equal if and only if $\lim_{n\to\infty} a_n$ exists. Note that a sequence may have much more than two accumulation points, e.g. for $a_n = \sin n$ the set of accumulation points is $[-1,1]$, $\liminf_{n\to\infty} a_n = -1$ and $\limsup_{n\to\infty} a_n = 1$.

We use the same concept for a sequence $(A_n)_{n\in\mathbb{N}}$ of subsets of a set E. Note that $\bigcap_{k\geq n} A_k \nearrow$ and $\bigcup_{k\geq n} A_k \searrow$ are monotone in n.

Definition 1.12. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of sets in E. Then define the sets

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

Remark. This definition can be interpreted as

$$\liminf_{n\to\infty} A_n = \big\{x\ :\ \exists n\in\mathbb{N}\ \forall k\geq n\ x\in A_k\big\} \ = \big\{x\ :\ x\in A_n \text{ for all but finitely many } n\big\}$$
$$\limsup_{n\to\infty} A_n = \big\{x\ :\ \forall n\in\mathbb{N}\ \exists k\geq n\ x\in A_k\big\} \ = \big\{x\ :\ x\in A_n \text{ for infinitely many } n\big\} \ .$$

One also writes $\liminf_{n\to\infty} A_n =$'$A_n$ ev.' (eventually) and $\limsup_{n\to\infty} A_n =$'$A_n$ i.o.' (infinitely often).

Properties. (i) Let $\mathcal{E}$ be a σ-algebra. If $A_n \in \mathcal{E}$ for all $n\in\mathbb{N}$ then $\liminf_{n\to\infty} A_n,\ \limsup_{n\to\infty} A_n \in \mathcal{E}$.

(ii) $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$ since $x\in A_n$ eventually $\Rightarrow x\in A_n$ infinitely often

(iii) $\big(\limsup_{n\to\infty} A_n\big)^c = \liminf_{n\to\infty} A_n^c$ since $x\in A_n$ finitely often $\Leftrightarrow x\in A_n^c$ eventually.

Lemma 1.12. First Borel-Cantelli lemma
Let $(E,\mathcal{E},\mu)$ be a measure space and $(A_n)_{n\in\mathbb{N}}$ a sequence of sets in $\mathcal{E}$.

If $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$ *then* $\mu(A_n \text{ i.o.}) = 0$.

Proof. $\mu(A_n \text{ i.o.}) = \mu\Big(\bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_k\Big) \leq \mu\Big(\bigcup_{k\geq n} A_k\Big) \leq \sum_{k\geq n} \mu(A_k) \to 0$ for $n\to\infty$. □

Lemma 1.13. Second Borel-Cantelli lemma
Let $(E,\mathcal{E},\mathbb{P})$ be a probability space and suppose that $(A_n)_{n\in\mathbb{N}}$ are independent.

If $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$, *then* $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof. We use the inequality $1-a \leq e^{-a}$. With $(A_n)_{n\in\mathbb{N}}$ also $(A_n^c)_{n\in\mathbb{N}}$ are independent (see problem 1.11). Then we have for all $n\in\mathbb{N}$

$$\mathbb{P}\Big(\bigcap_{k\geq n} A_k^c\Big) = \prod_{k\geq n} \big(1-\mathbb{P}(A_k)\big) \leq \exp\Big[-\sum_{k\geq n} \mathbb{P}(A_k)\Big] = 0\ .$$

Hence $\mathbb{P}(A_n \text{ i.o.}) = 1 - \mathbb{P}\Big(\liminf_{n\to\infty} A_n^c\Big) = 1 - \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} A_k^c\Big) = 1$. □

2 Measurable Functions and Random Variables

2.1 Measurable Functions

Definition 2.1. Let $(E,\mathcal{E})$ and $(F,\mathcal{F})$ be measurable spaces. A function $f : E \to F$ is called *measurable (with respect to $\mathcal{E}$ and $\mathcal{F}$)* or *$\mathcal{E}/\mathcal{F}$-measurable* if

$$\forall\, A \in \mathcal{F} \ : \quad f^{-1}(A) = \big\{x \in E \,:\, f(x) \in A\big\} \in \mathcal{E} \qquad \big(\text{short: } f^{-1}(\mathcal{F}) \subseteq \mathcal{E}\big)\ .$$

Often $(F,\mathcal{F}) = (\mathbb{R},\mathcal{B})$ or $(\overline{\mathbb{R}},\overline{\mathcal{B}})$ with the *extended real line* $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty,\infty\}$ and $\overline{\mathcal{B}} = \big\{B \cup C \,:\, b \in \mathcal{B},\, C \subseteq \{-\infty,\infty\}\big\}$. If in addition E is a topological space with $\mathcal{E} = \mathcal{B}(E)$, f is called *Borel function*.

Remarks. (i) Every function $f : E \to F$ is measurable w.r.t. $\mathcal{P}(E)$ and $\mathcal{F}$.

(ii) Preimages of functions preserve the set operations

$$f^{-1}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \bigcup_{n\in\mathbb{N}} f^{-1}(A_n)\ , \quad f^{-1}(A^c) = \big(f^{-1}(A)\big)^c\ ,$$

since e.g. $\{x \in E \,:\, f(x) \notin A\} = \{x \in E \,:\, f(x) \in A\}^c$. Note that this second property does not hold for images since in general $f(A^c) \subseteq \big(f(A)\big)^c$ for $A \in \mathcal{E}$.

(iii) With (ii) the following holds for any function $f : E \to F$:
If $\mathcal{F}$ is a σ-algebra on F then $\sigma(f) := f^{-1}(\mathcal{F})$ is a σ-algebra on E, called *σ-algebra generated by f*. This is the smallest σ-algebra on E w.r.t. which f is measurable.
If $\mathcal{E}$ is a σ-algebra on E then $\mathcal{C} = \big\{A \subseteq F \,:\, f^{-1}(A) \in \mathcal{E}\big\}$ is the largest σ-algebra on F w.r.t. which f is measurable. Note that $\mathcal{C} \neq f(\mathcal{E})$, which is in general not a σ-algebra.

Lemma 2.1. *Let $f : E \to F$ and $\mathcal{F} = \sigma(\mathcal{A})$ for some $\mathcal{A} \subseteq \mathcal{P}(F)$.*
Then f is $\mathcal{E}/\mathcal{F}$-measurable if and only if $\quad f^{-1}(A) \in \mathcal{E} \quad$ *for all $A \in \mathcal{A}$.*

Proof. According to Remark (iii), $\mathcal{C} := \big\{A \subseteq F \,:\, f^{-1}(A) \in \mathcal{E}\big\}$ is a σ-algebra on F. Now if $\mathcal{A} \subseteq \mathcal{C}$ then $\sigma(\mathcal{A}) = \mathcal{F} \subseteq \mathcal{C}$ and f is $\mathcal{E}/\mathcal{F}$-measurable. On the other hand, if f is $\mathcal{E}/\mathcal{F}$-measurable then certainly $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{A} \subseteq \mathcal{F}$. □

Lemma 2.2. *$f : E \to \mathbb{R}$ is $\mathcal{E}/\mathcal{B}$-measurable if and only if one of the following holds:*

(i) $f^{-1}\big((-\infty,c]\big) = \big\{x \in E \,:\, f(x) \leq c\big\} \in \mathcal{E}$ *for all* $c \in \mathbb{R}$,
(ii) $f^{-1}\big((-\infty,c)\big) = \big\{x \in E \,:\, f(x) < c\big\} \in \mathcal{E}$ *for all* $c \in \mathbb{R}$,
(iii) $f^{-1}\big([c,\infty)\big) = \big\{x \in E \,:\, f(x) \geq c\big\} \in \mathcal{E}$ *for all* $c \in \mathbb{R}$,
(iv) $f^{-1}\big((c,\infty)\big) = \big\{x \in E \,:\, f(x) > c\big\} \in \mathcal{E}$ *for all* $c \in \mathbb{R}$.

Proof. (i) In problem 1.3 it was shown that $\mathcal{B} = \sigma\big(\{(-\infty,c] \,:\, c \in \mathbb{R}\}\big)$. The statement then follows with Lemma 3.1.
(ii) – (iv) Show analogously that $\mathcal{B}$ is generated by the respective sets. □

Lemma 2.3. *Let E and F be topological spaces and $f : E \to F$ be continuous (i.e. $f^{-1}(U) \subseteq E$ open whenever $U \subseteq F$ open). Then f is measurable w.r.t. $\mathcal{B}(E)$ and $\mathcal{B}(F)$.*

Proof. Let $\tau_F = \big\{U \subseteq F : U \text{ open}\big\}$ be the topology on F. Then for all $U \in \tau_F$, $f^{-1}(U)$ is open and thus $f^{-1}(U) \in \mathcal{B}(E)$. Since $\mathcal{B}(F) = \sigma(\tau_F)$, f is measurable with Lemma 2.1. □

However, not every measurable function is continuous. Next we introduce another important class of functions which turn out to be measurable.

Definition 2.2. Let $\mathbb{1}_A : E \to \mathbb{R}$ be the indicator function of $A \subseteq E$.
$f : E \to \mathbb{R}$ is called *simple* if

$$f = \sum_{i=1}^{n} c_i \, \mathbb{1}_{A_i} \qquad \text{for some } n \in \mathbb{N},\ c_i \in \mathbb{R} \text{ and } A_1, \ldots, A_n \subseteq E\ .$$

We call this a *standard representation* of f if the $A_i \neq \emptyset$, $c_i \neq c_j$ and $A_i \cap A_j = \emptyset$ for $i \neq j$. f is called $\mathcal{E}$*-simple* if there exists a standard representation such that $A_i \in \mathcal{E}$ for all $i = 1, \ldots, n$. Let $\mathcal{S}(\mathcal{E})$ denote the set of all $\mathcal{E}$-simple functions.

Remark. (i) Simple fct's are more general than *step functions*, where the A_i are intervals.

(ii) Standard representations are not unique, the order of indices may change and a c_i may or may not take the value 0.

Lemma 2.4. *Let $(E, \mathcal{E})$ be a measurable space.*

(i) *A simple function $f : E \to \mathbb{R}$ is $\mathcal{E}/\mathcal{B}$-measurable if and only if $f \in \mathcal{S}(\mathcal{E})$.*

(ii) *$\mathcal{S}(\mathcal{E})$ is a vector space, i.e. if $f_1, f_2 \in \mathcal{S}(\mathcal{E})$ then $\lambda_1 f_1 + \lambda_2 f_2 \in \mathcal{S}(\mathcal{E})$ for all $\lambda_1,\ \lambda_2 \in \mathbb{R}$. In addition $f_1\, f_2 \in \mathcal{S}(\mathcal{E})$.*

Proof. (i) Let $f = \sum\limits_{i=1}^{n} c_i \, \mathbb{1}_{A_i}$ be a simple function in standard representation.
If $A_i \in \mathcal{E}$ for $i = 1, \ldots, n$ then for all $B \in \mathcal{B}$, $\quad f^{-1}(B) = \bigcup\limits_{i: c_i \in B} A_i \in \mathcal{E}$.
On the other hand, if $A_i \notin \mathcal{E}$ for some i then $f^{-1}\big(\{c_i\}\big) = A_i \notin \mathcal{E}$ and f is not measurable.
(ii) Let $f_1, f_2 \in \mathcal{S}(\mathcal{E})$ with standard representations $f_1 = \sum_{i=1}^{n} c_i \, \mathbb{1}_{A_i}$ and $f_2 = \sum_{i=1}^{m} d_i \, \mathbb{1}_{B_i}$.
Define $C_{ij} = A_i \cap B_j$. Then the $\{C_{ij}\}$ are disjoint and each $C_{ij} \in \mathcal{E}$ as well as all possible unions of C_{ij}s. $f_1\, f_2$ and $\lambda_1 f_1 + \lambda_2 f_2$ are constant on each C_{ij} and thus $\mathcal{E}$-simple. □

Remark. In particular, $\mathbb{1}_A : E \to \mathbb{R}$ is $\mathcal{E}/\mathcal{B}$-measurable if and only if $A \in \mathcal{E}$.

Lemma 2.5. *Let $f_1 : E_1 \to E_2$ and $f_2 : E_2 \to E_3$. If f_1 is $\mathcal{E}_1/\mathcal{E}_2$-measurable and f_2 is $\mathcal{E}_2/\mathcal{E}_3$-measurable, then $f_2 \circ f_1 : E_1 \to E_3$ is $\mathcal{E}_1/\mathcal{E}_3$-measurable.*

Proof. For every $A \in \mathcal{E}_3$, $(f_2 \circ f_1)^{-1}(A) = f_1^{-1}\big(f_2^{-1}(A)\big) \in \mathcal{E}_1$ since $f_2^{-1}(A) \in \mathcal{E}_2$. □

Proposition 2.6. *Let $f_n : E \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, be $\mathcal{E}/\overline{\mathcal{B}}$-measurable. Then so are*

$$\text{(i)}\quad c\,f_1 \quad \textit{for all } c \in \mathbb{R} \qquad \text{(ii)}\ \ f_1 + f_2 \qquad \text{(iii)}\ \ f_1\,f_2$$
$$\text{(iv)}\ \ \inf_{n\in\mathbb{N}} f_n \qquad \text{(v)}\ \ \sup_{n\in\mathbb{N}} f_n \qquad \text{(vi)}\ \ \liminf_{n\to\infty} f_n \qquad \text{(vii)}\ \ \limsup_{n\to\infty} f_n\ ,$$

whenever they are defined ($\infty - \infty$ and $\frac{\pm\infty}{\pm\infty}$ are not well defined).

Remarks. (i) Notation: $\inf_{n\in\mathbb{N}} f_n : x \mapsto \inf\big\{f_n(x) : n \in \mathbb{N}\big\} \in \overline{\mathbb{R}}$ and
$$\liminf_{n\to\infty} f_n : x \mapsto \lim_{n\to\infty}\Big(\inf\big\{f_k(x) : k \geq n\big\}\Big) \in \overline{\mathbb{R}}\ .$$

(ii) In particular $f_1 \vee f_2 = \max\{f_1, f_2\}$ and $f_1 \wedge f_2 = \min\{f_1, f_2\}$ are measurable. Whenever it exists, also $\lim_{n\to\infty} f_n$ is measurable.

(iii) $f = f^+ - f^-$ meas. $\Leftrightarrow$ $f^+ = f \vee 0$ and $f^- = (-f) \vee 0$ meas. $\Rightarrow$ $|f| = f^+ + f^-$ meas. The inverse of the last implication is in general false, e.g. $f^+ = \mathbb{1}_A$ and $f^- = \mathbb{1}_{A^c}$ for $A \notin \mathcal{E}$.

Proof. (i) If $c \neq 0$ we have for all $y \in \mathbb{R}$

$$\big\{x \in E : c\,f_1(x) \leq y\big\} = \big\{x \in E : f_1(x) \leq y/c\big\} \in \mathcal{E} \quad \text{since} \quad f_1 \text{ measurable .}$$

If $c = 0$ it is $\big\{x \in E : 0 \leq y\big\} = \begin{cases} E\ , & y \geq 0 \\ \emptyset\ , & y < 0 \end{cases} \in \mathcal{E}$, so $c\,f_1$ is measurable for all $c \in \mathbb{R}$.
(ii) see example sheet
(iii) $f_1\,f_2 = \frac{1}{4}\big((f_1 + f_2)^2 - (f_1 - f_2)^2\big)$ is measurable with (i), (ii), Lemma 2.3 and 2.5, since $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$ is continuous and thus measurable.
(iv) – (vii) see example sheet □

Definition 2.3. Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces and let μ be a measure on $(E, \mathcal{E})$. Then any $\mathcal{E}/\mathcal{F}$-measurable function $f : E \to F$ induces the *image measure* $\nu = \mu \circ f^{-1}$ on $\mathcal{F}$, given by $\nu(A) = \mu\big(f^{-1}(A)\big)$ for all $A \in \mathcal{F}$.

Remark. ν is a measure since $f^{-1}(A) \in \mathcal{E}$ for all $A \in \mathcal{F}$ and f^{-1} preserves set operations as has been shown above.

2.2 Random Variables

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(E, \mathcal{E})$ a measurable space.

Definition 2.4. An $\mathcal{A}/\mathcal{E}$-measurable function $X : \Omega \to E$ is called *random variable* in E or simply *random variable* (if $E = \mathbb{R}$). The image measure $\mu_X = \mathbb{P} \circ X^{-1}$ on $(E, \mathcal{E})$ is called *law* or *(probability) distribution* of X.
For $E = \mathbb{R}$ the cumulative probability distribution function $F_X = CDF_{\mu_X} : \mathbb{R} \to [0, 1]$ with

$$F_X(x) = \mu_X\big((-\infty, x]\big) = \mathbb{P}\big(\{\omega \in \Omega : X(\omega) \leq x\}\big) = \mathbb{P}(X \leq x)$$

is called *distribution function of* X.

Remark. By Proposition 1.9 and Theorem 1.10, μ_X is characterised by F_X. Usually random variables are given by their distribution function without specifying $(\Omega, \mathcal{A}, \mathbb{P})$ and the function $X : \Omega \to \mathbb{R}$.

Definition 2.5. The random variables $(X_n)_{n\in\mathbb{N}}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in $(E, \mathcal{E})$, are called *independent* if the σ-algebras $\sigma(X_n) = X_n^{-1}(\mathcal{E}) \subseteq \mathcal{A}$ are independent.

Lemma 2.7. *Real random variables $(X_n)_{n\in\mathbb{N}}$ are independent if and only if*

$$\mathbb{P}\big(X_1 \le x_1, \ldots, X_k \le x_k\big) = \mathbb{P}\big(X_1 \le x_1\big) \cdots \mathbb{P}\big(X_k \le x_k\big)$$

for all $x_1, \ldots, x_k \in \mathbb{R}$, $k \in \mathbb{N}$.

Proof. see problem 2.5 for two random variables X_1, X_2.
This extends to $X_1, \ldots, X_k$, noting that by continuity of measures e.g. for $k = 3$

$$\mathbb{P}\big(X_1 \le x_1, X_3 \le x_3\big) = \lim_{x_2\to\infty} \mathbb{P}\big(X_1 \le x_1, X_2 \le x_2, X_3 \le x_3\big) =$$
$$= \lim_{x_2\to\infty} \mathbb{P}(X_1 \le x_1)\mathbb{P}(X_2 \le x_2)\mathbb{P}(X_3 \le x_3) = \mathbb{P}(X_1 \le x_1)\mathbb{P}(X_3 \le x_3)\,. \qquad \square$$

Remark. This Lemma provides a characterisation of independence using only distribution functions. But to relate this to the definition, the functions X_n have to be defined on the same probability space. Although one usually does not bother to define them, at least it has to be possible to do so. This is guaranteed by the next theorem which is, although rarely used in practice, of great conceptual importance. It is split in two parts, the secon is Theorem 2.?.

Theorem 2.8. Skorohod representation theorem – part 1
For all probability distribution functions $F_1, F_2, \ldots : \mathbb{R} \to [0,1]$ there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and random variables $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ such that F_n is the distribution function of X_n for each n. The X_n can be chosen to be independent.

Proof. see problem 2.4, for independence see hand-out 2

Remark. $(X_n)_{n\ge0}$ is often regarded as a *stochastic process* with state space E and discrete time n. The *σ-algebra generated by* $X_0, \ldots, X_n$,

$$\mathcal{F}_n = \sigma(X_0, \ldots, X_n) = \sigma\Big(\bigcup_{i=0}^{n} X_i^{-1}(\mathcal{E})\Big) \subseteq \mathcal{A}\,,$$

contains events depending measurably on $X_0, \ldots, X_n$ and represents what is known about the process by time n. $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for each n and the family $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is called the *filtration* generated by the process $(X_n)_{n\ge0}$.

Definition 2.6. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables. Define

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots) \quad (\searrow \text{ in } n) \quad \text{and} \quad \mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n \subseteq \mathcal{A}\,.$$

Then $\mathcal{T}$ is called the *tail σ-algebra* of $(X_n)_{n\in\mathbb{N}}$ and elements in $\mathcal{T}$ are called *tail events*.

Example. $A = \{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\} \in \mathcal{T}$ since $A = \{\omega : \lim_{n\to\infty} X_{N+n}(\omega) \text{ exists}\}$ for every fixed $N \in \mathbb{N}$. Similarly, $\{\omega : \limsup_{n\to\infty} X_n(\omega) = 137\} \in \mathcal{T}$.

Theorem 2.9. Kolmogorov's 0-1-law
Suppose $(X_n)_{n\in\mathbb{N}}$ is a sequence of independent random variables. Then every tail event has probability 0 or 1. Moreover, any $\mathcal{T}$-measurable random variable Y is almost surely constant, i.e. $\mathbb{P}(Y = c) = 1$ for some $c \in \mathbb{R}$.

Proof. The σ-algebra $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is generated by the π-system of events

$$A = \{X_1 \le x_1, \ldots, X_n \le x_n\} ,$$

whereas $\mathcal{T}_n$ is generated by the π-system of events

$$B = \{X_{n+1} \le x_{n+1}, \ldots, X_{n+k} \le x_{n+k}\} , \quad k \in \mathbb{N} .$$

Since $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ for all such A and B by independence, $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent by Theorem 1.11 for all $n \in \mathbb{N}$. Hence $\mathcal{F}_n$ and $\mathcal{T}$ are independent, since $\mathcal{T} \subseteq \mathcal{T}_{n+1}$.
Since $\bigcup_n \mathcal{F}_n$ is a π-system generating the σ-algebra $\mathcal{F}_\infty = \sigma(X_n : n \in \mathbb{N})$, $\mathcal{F}_\infty$ and $\mathcal{T}$ are independent, again by Theorem 1.11. But $\mathcal{T} \subseteq \mathcal{F}_\infty$ and thus every $A \in \mathcal{T}$ is independent of itself, i.e.

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\,\mathbb{P}(A) \quad \Rightarrow \quad \mathbb{P}(A) \in \{0, 1\} .$$

Let Y be a $\mathcal{T}$-measurable random variable. Then $F_Y(y) = \mathbb{P}(Y \le y)$ takes values in $\{0, 1\}$, so $\mathbb{P}(Y = c) = 1$ for $c = \inf\{y \in \mathbb{R} : F_Y(y) = 1\}$. □

Remark. Kolmogorov's 0-1-law involves the σ-algebras generated by random variables, rather than the random variables themselves. Thus it can be formulated without using r.v.'s:
Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of independent σ-algebras in $\mathcal{A}$. Let A be a tail event, i.e.

$$A \in \mathcal{T} , \quad \textit{where} \quad \mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n \quad \textit{with} \quad \mathcal{T}_n = \sigma\Big(\bigcup_{n\ge m} \mathcal{F}_m\Big) .$$

Then $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

2.3 Convergence of measurable functions

Let $(E, \mathcal{E}, \mu)$ be a measure space.

Remark. *'Convergence' to infinity*
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\overline{\mathbb{R}}$. Here and in the following we say that

$$x_n \to \infty \quad \text{if} \quad \forall\, y \in \mathbb{R}\ \exists\, N \in \mathbb{N}\ \forall\, n \ge N \ :\ x_n \ge y ,$$

and $x_n \nearrow \infty$ if in addition $x_{n+1} \ge x_n$ for all $n \in \mathbb{N}$ (analogously $x_n \to -\infty$ and $x_n \searrow -\infty$). Remember that x_n is *unbounded form above* if $\forall\, y \in \mathbb{R}\ \forall\, N \in \mathbb{N}\ \exists\, n \ge N \ :\ x_n \ge y$, i.e. $\forall\, y \in \mathbb{R} \ :\ x_n \ge y$ for infinitely many n, whereas $x_n \to \infty$ means that $\forall\, y \in \mathbb{R} \ :\ x_n \ge y$ for all but finitely many n. This is not convergence in the usual sense, since either $|\infty - x_n|$ is not well defined or is equal to ∞.

Definition 2.7. We say that $A \in \mathcal{E}$ holds *almost everywhere* (short *a.e.*), if $\mu(A^c) = 0$. If μ is a probability measure ($\mu(E) = 1$) one uses *almost surely* (short *a.s.*) instead.
Let $f, f_1, f_2, \ldots : E \to \overline{\mathbb{R}}$ be measurable functions. Say that

(i) $f_n \to f$ *everywhere* or *pointwise* if $\quad f_n(x) \to f(x) \quad$ for all $x \in E$,

(ii) $f_n \to f$ *almost everywhere* ($f_n \xrightarrow{a.e.} f$), *almost surely* ($f_n \xrightarrow{a.s.} f$) for $\mu(E) = 1$, if

$$\mu\big(\{x \in E : f_n(x) \not\to f(x)\}\big) = \mu(f_n \not\to f) = 0 \,,$$

(iii) $f_n \to f$ *in measure, in probability* ($f_n \xrightarrow{P} f$) for $\mu(E) = 1$, if

$$\forall\, \epsilon > 0 \,:\, \mu(|f_n - f| > \epsilon) \to 0 \quad \text{for} \quad n \to \infty \,.$$

Theorem 2.10. *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions.*

(i) Assume that $\mu(E) < \infty$. If $f_n \to f$ a.e. then $f_n \to f$ in measure.

(ii) If $f_n \to f$ in measure then there exists a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $f_{n_k} \to f$ a.e..

Proof. (i) Set $g_n = f_n - f$ and suppose $g_n \to 0$ a.e.. Then for every $\epsilon > 0$

$$\mu\big(|g_n| \le \epsilon\big) \ge \mu\Big(\bigcap_{m\ge n}\{|g_m| \le \epsilon\}\Big) \nearrow \mu\big(|g_n| \le \epsilon \; ev.\big) \ge \mu(g_n \to 0) = \mu(E) \,.$$

Hence $\mu(|g_n| > \epsilon) \to 0$ as $n \to \infty$ and $f_n \to f$ in measure.
(ii) Suppose $g_n = f_n - f \to 0$ in measure. Thus $\mu(|g_n| > \frac{1}{k}) \to 0$ for every $k \in \mathbb{N}$ and we can find a subsequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\mu\big(|g_{n_k}| > \tfrac{1}{k}\big) < 2^{-k} \quad \text{and thus} \quad \sum_{k\in\mathbb{N}} \mu\big(|g_{n_k}| > \tfrac{1}{k}\big) < \infty \,.$$

So, by the first Borel-Cantelli lemma (Lemma 1.12) $\quad \mu\big(|g_{n_k}| > \frac{1}{k}\; i.o.\big) = 0$.
$\big\{|g_{n_k}| > \frac{1}{k}\; i.o.\big\}^c \subseteq \{g_{n_k} \to 0\}$ and thus

$$\mu\big(g_{n_k} \not\to 0\big) \le \mu\big(|g_{n_k}| > \tfrac{1}{k}\; i.o.\big) = 0 \,,$$

so $g_{n_k} \to 0$ a.e.. □

Definition 2.8. Let $X, X_1, X_2, \ldots$ be random variables with distribution functions $F, F_1, F_2, \ldots$.
Say that X_1 and X_2 are *identically distributed*, written as $X_1 \sim X_2$, if $F_1(x) = F_2(x)$, $x \in \mathbb{R}$.
Say that $X_n \to X$ *in distribution* (short $X_n \xrightarrow{D} X$) if for all continuity points x of F,

$$F_n(x) = \mathbb{P}(X_n \le x) \to \mathbb{P}(X \le x) = F(x) \quad \text{as } n \to \infty \,.$$

Proposition 2.11. *Let $X, X_1, X_2, \ldots$ be random variables on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then* $\quad X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{D} X \quad$ *and* $\quad X_n \xrightarrow{D} c \in \mathbb{R} \quad \Rightarrow \quad X_n \xrightarrow{P} c$.

Proof. The first statement is proved on hand-out 2, the second follows directly from

$$\mathbb{P}\big(|X_n - c| > \epsilon\big) = \mathbb{P}\big(X_n > c + \epsilon\big) + \mathbb{P}\big(X_n < c - \epsilon\big) \stackrel{n\to\infty}{\longrightarrow} 0 \quad \text{for all } \epsilon > 0\ . \qquad \square$$

Theorem 2.12. Skorohod representation theorem – part 2
Let $X, X_1, X_2, \ldots$ be random variables such that $X_n \to X$ in distribution. Then there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and random variables $Y \sim X, Y_1 \sim X_1, Y_2 \sim X_2, \ldots$ defined on Ω such that $Y_n \to Y\ a.s.$.

Proof. see problem 2.4

Example. Let $X_1, X_2, \ldots \in \{0, 1\}$ be independent random variables with

$$\mathbb{P}(X_n = 0) = 1 - 1/n \quad \text{and} \quad \mathbb{P}(X_n = 1) = 1/n\ .$$

Then $\forall\, \epsilon > 0 : \mathbb{P}\big(|X_n| > \epsilon\big) = 1/n \to 0$ as $n \to \infty$, i.e. $X_n \to 0$ in measure.
On the other hand, $\sum_n \mathbb{P}(X_n = 1) = \infty$ and $\{X_n = 1\}$ are independent events.
Thus with the second Borel-Cantelli lemma

$$\mathbb{P}(X_n \not\to 0) \geq \mathbb{P}(X_n = 1 \text{ i.o.}) = 1\ , \quad \text{and thus} \quad X_n \not\to 0 \text{ a.s.}\ .$$

3 Integration

3.1 Definition and basic properties

Let $(E, \mathcal{E}, \mu)$ be a measure space.

Theorem 3.1. *Let $f : E \to [0, \infty]$ be $\mathcal{E}/\overline{\mathcal{B}}$-measurable. Then $f_n(x) = \dfrac{\lfloor f(x)\, 2^n \rfloor}{2^n} \wedge n$ defines a sequence of $\mathcal{E}$-simple, non-negative functions, such that $f_n \nearrow f$ pointwise as $n \to \infty$.*

Proof. We can write $\quad f_n(x) = \sum_{k=0}^{2^n n} \mathbb{1}_{A_{k,n}}(x)\, 2^{-n}\, k \quad$ where

$$A_{k,n} = f^{-1}\Big(\big[2^{-n}\, k\ ,\ 2^{-n}(k+1)\big)\Big) \text{ for } k < 2^n n \quad \text{and} \quad A_{2^n n,n} = f^{-1}\Big(\big[n\ ,\ \infty\big]\Big)\ .$$

Since f is measurable, so are the sets $A_{k,n}$, and thus $f_n \in \mathcal{S}(\mathcal{E})$ for all $n \in \mathbb{N}$.
From the first representation it follows immediately that $f_{n+1}(x) \geq f_n(x)$ for all $x \in E$ and that $|f_n(x) - f(x)| \leq 2^{-n}$ for $n \geq f(x)$, or $f_n(x) \uparrow \infty$ for $f(x) = \infty$. Thus $f_n \nearrow f$. □

This motivates the following definition.

Definition 3.1. Let $f : E \to [0, \infty]$ be an $\mathcal{E}/\overline{\mathcal{B}}$-measurable function. We define the *integral* of f, written as $\quad \mu(f) = \int_E f\, d\mu = \int f\, d\mu = \int_E f(x)\, \mu(dx)\ ,\quad$ by

$$\int_E f\, d\mu := \sup\Big\{ \int_E g\, d\mu\ :\ g \in \mathcal{S}(\mathcal{E}),\ 0 \leq g \leq f \Big\}\ ,$$

where $\int_E g\, d\mu\ :=\ \sum_{k=1}^{n} c_k\, \mu(A_k)$ is the integral of an $\mathcal{E}$-simple function $g : E \to \mathbb{R}$ with standard representation $\quad g(x) = \sum_{k=1}^{n} c_k\, \mathbb{1}_{A_k}$. We adopt $\infty \cdot 0 = 0 \cdot \infty = 0$.

Remarks. (i) $\int g\, d\mu$ is independent of the representation of the $\mathcal{E}$-simple function g.

(ii) If $f, g : E \to \mathbb{R}$ are $\mathcal{E}$-simple then: $\qquad f \leq g \quad \Rightarrow \quad \int_E f\, d\mu \leq \int_E g\, d\mu \quad$ and

$$\int_E \big(c_1\, f + c_2\, g\big)\, d\mu = c_1 \int_E f\, d\mu + c_2 \int_E g\, d\mu \qquad \text{for all } c_1, c_2 \in \mathbb{R}\ .$$

Lemma 3.2. *Let $f : E \to [0, \infty]$ be measurable and $(f_n)_{n\in\mathbb{N}}$ a sequence in $\mathcal{S}(\mathcal{E})$ with $0 \leq f_n \nearrow f$. Then $\quad \int_E f_n\, d\mu \nearrow \int_E f\, d\mu$.*

Proof. $\int_E f_n\, d\mu \leq \int_E f\, d\mu$ for all $n \in \mathbb{N}$ by definition of $\int_E f\, d\mu$. It remains to show that for any $\mathcal{E}$-simple $g = \sum_{k=1}^{n} a_k \mathbb{1}_{A_k} \leq f$ (with standard representation and $a_k \neq 0$)

$$\lim_{n\to\infty} \int_E f_n\, d\mu \geq \int_E g\, d\mu\ .$$

Choose $\epsilon > 0$ and set $B_n := \{x \in E : f_n(x) \geq g(x) - \epsilon\}$. Thus $B_n \nearrow E$ and for any $A \in \mathcal{E}$: $\mu(B_n \cap A) \nearrow \mu(A)$.
Case (i): $\int_E g\, d\mu = \infty \quad \Rightarrow \quad \mu(A_r) = \infty$ for some $r \in \{1, \ldots, n\}$ and $a_r > 0$. Then

$$\int_E f_n\, d\mu \geq \int_E f_n\, \mathbb{1}_{B_n \cap A_r}\, d\mu \geq \int_E (g - \epsilon)\, \mathbb{1}_{B_n \cap A_r}\, d\mu = (a_r - \epsilon)\, \mu(B_n \cap A_r) \to \infty$$

as $n \to \infty$, provided $\epsilon < a_r$.
Case (ii): $\int_E g\, d\mu < \infty \quad \Rightarrow$ for $A = \bigcup_{k=1}^n A_k$ it is $\mu(A) < \infty$. Then

$$\int_E f_n\, d\mu \geq \int_E f_n\, \mathbb{1}_{B_n \cap A}\, d\mu \geq \int_E (g - \epsilon)\, \mathbb{1}_{B_n \cap A}\, d\mu =$$
$$= \int_E g\, \mathbb{1}_{B_n \cap A}\, d\mu - \epsilon\, \mu(B_n \cap A) \longrightarrow \int_E g\, d\mu - \epsilon\, \mu(A) \quad \text{as } n \to \infty\,.$$

This is true for ϵ arbitrarily small and thus $\lim_{n\to\infty} \int_E f_n\, d\mu \geq \int_E g\, d\mu$. □

Definition 3.2. Let $f : E \to \overline{\mathbb{R}}$ be a measurable function. Then f^+ and f^- are measurable with Proposition 2.6. f is called *(μ-)integrable* if $\int_E f^+\, d\mu < \infty$ and $\int_E f^-\, d\mu < \infty$ and the *integral* of f is defined as

$$\int_E f\, d\mu = \int_E f^+\, d\mu - \int_E f^-\, d\mu \in \mathbb{R}\,.$$

For random variables $X : \Omega \to \mathbb{R}$ the integral $\int_\Omega X\, d\mathbb{P} = \mathbb{E}(X)$ is also called *expectation*.
For $A \in \mathcal{E}$ and f integrable, $f\, \mathbb{1}_A$ is integrable and we write $\int_A f\, d\mu := \int_E f\, \mathbb{1}_A\, d\mu$.

Remark. The integral can be well defined even if f is not integrable, namely if either $\int_E f^+\, d\mu = \infty$ or $\int_E f^-\, d\mu = \infty$, it takes a value $\pm\infty$. In particular a measurable function $f : E \to [0, \infty]$ is integrable if and only if $\int_E f\, d\mu < \infty$.

Theorem 3.3. Basic properties of integration
Let $f, g : E \to \overline{\mathbb{R}}$ be integrable functions on $(E, \mathcal{E}, \mu)$.

(i) Linearity: $f + g$ *and, for any* $c \in \mathbb{R}$, $c\, f$ *are integrable with*

$$\int_E (f + g)\, d\mu = \int_E f\, d\mu + \int_E g\, d\mu\,, \quad \int_E (c\, f)\, d\mu = c \int_E f\, d\mu\,.$$

(ii) Monotonicity: $f \geq g \quad \Rightarrow \quad \int_E f\, d\mu \geq \int_E g\, d\mu\,.$

(iii) $f \geq 0$ *and* $\int_E f\, d\mu = 0 \quad \Rightarrow \quad f = 0$ *a.e.* $\quad \Rightarrow \quad \int_E f\, d\mu = 0\,.$

Let $f : E \to \overline{\mathbb{R}}$ be measurable. Then

(iv) f *integrable* $\quad \Leftrightarrow \quad |f|$ *integrable, and in this case* $\int_E |f|\, d\mu \geq \left| \int_E f\, d\mu \right|.$

Proof. (i) If $f, g \geq 0$ choose sequences of $\mathcal{E}$-simple functions with $0 \leq f_n \nearrow f$ and $0 \leq g_n \nearrow g$. Then $f_n + g_n$ is $\mathcal{E}$-simple for all $n \in \mathbb{N}$ and

$$\int_E (f_n + g_n)\, d\mu = \int_E f_n\, d\mu + \int_E g_n\, d\mu\,.$$

Since $0 \leq f_n + g_n \nearrow f + g$ it follows by Lemma 3.2: $\displaystyle\int_E (f+g)\, d\mu = \int_E f\, d\mu + \int_E g\, d\mu\,.$
For $f, g : E \to \overline{\mathbb{R}}$ we have $\quad (f+g)^+ - (f+g)^- = f^+ - f^- + g^+ - g^- \quad$ and thus

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+\,.$$

Since each of the terms is non-negative we also have

$$\int_E (f+g)^+\, d\mu + \int_E f^-\, d\mu + \int_E g^-\, d\mu = \int_E (f+g)^-\, d\mu + \int_E f^+\, d\mu + \int_E g^+\, d\mu$$

and the statement follows by reordering the terms.
For $c\, f$, $c \geq 0$, analogously, and for $c < 0$ use $\quad 0 = \displaystyle\int_E (f - f)\, d\mu = \int_E f\, d\mu + \int_E (-f)\, d\mu\,.$
(ii) With $f \geq g$ using (i): $\quad \displaystyle\int_E f\, d\mu = \underbrace{\int_E (f-g)\, d\mu}_{\geq 0} + \int_E g\, d\mu \geq \int_E g\, d\mu\,.$
(iii) Let $f \geq 0$ and suppose that $\mu(f > 0) > 0$. Then, since $\{f > 0\} = \bigcup_{m\in\mathbb{N}} \{f \geq 1/m\}$, we have $\mu(f \geq 1/n) = \epsilon > 0$ for some $n \in \mathbb{N}$. Thus $f \geq \frac{1}{n}\, \mathbb{1}_{f \geq 1/n}$ and $\int_E f\, d\mu \geq \epsilon/n > 0$.
On the other hand let $f = 0\ a.e. \quad \Rightarrow \quad f^+,\, f^- = 0\ a.e. \quad \Rightarrow \quad \int_E f^+\, d\mu = \int_E f^-\, d\mu = 0$ by definition, since $\int_E g\, d\mu = 0$ for all $\mathcal{E}$-simple $0 \leq g \leq f$.
(iv) Follows with $\quad |f| = f^+ + f^- \geq \begin{cases} f^+ - f^- = f \\ f^- - f^+ = -f \end{cases}$ and (i), (ii). □

Remark. Let $f, g : E \to \overline{\mathbb{R}}$ be measurable and $f = g\ a.e.$. Then f is integrable if and only if g is integrable, and then $\int_E f\, d\mu = \int_E g\, d\mu$, which is a direct consequence of (iii). In particular, whenever $f \geq 0\ a.e.$ the integral $\int_E f\, d\mu \in [0, \infty]$ is well defined.

Proposition 3.4. *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be measure spaces and suppose that $\nu = \mu \circ f^{-1}$ is the image measure of a measurable $f : E \to F$. Then*

$$\int_F g\, d\nu = \int_E (g \circ f)\, d\mu \quad \textit{for all integrable } g : F \to \overline{\mathbb{R}}\,.$$

Remark. In particular for random variables X with distribution μ_X this leads to the useful formula $\quad \mathbb{E}\big(g(X)\big) = \displaystyle\int_{\mathbb{R}} g(x)\, \mu_X(dx)\,.$

Proof. For $g = \mathbb{1}_A$, $A \in \mathcal{F}$, the identity $\nu(A) = \mu\big(f^{-1}(A)\big)$ is the definition of ν.
The identity extends to all $\mathcal{F}$-simple functions by linearity of integration, then to all measurable $g : F \to [0, \infty]$ with Lemma 3.2, using the approximations $g_n = 2^{-n} \lfloor 2^n g \rfloor \wedge n$, and finally to all integrable $g = g^+ - g^- : F \to \overline{\mathbb{R}}$ again by linearity. □

Examples. This notion of integration includes in particular Riemann integrals as is discussed in section 3.3, but is much more general than the latter.

(i) $\int_E f\,d\delta_y = f(y)$ for all integrable $f : E \to \overline{\mathbb{R}}$ and $y \in E$.
For $g = \sum_{k=1}^{n} c_k \mathbb{1}_{A_k} \in \mathcal{S}(\mathcal{E})$, $\int_E g\,d\delta_y = \sum_{k=1}^{n} c_k \delta_y(A_k) = g(y)$, since in standard representation $\delta_y(A_k) = 1$ for exactly one k. So for $\mathcal{E}$-simple $f_n \nearrow f : E \to [0,\infty]$

$$\int_E f\,d\delta_y = \lim_{n\to\infty} \int_E f_n\,d\delta_y = \lim_{n\to\infty} f_n(y) = f(y) \quad \text{by Lemma 3.2 ,}$$

which extends to $f : E \to \overline{\mathbb{R}}$ by linearity of integration.

(ii) Let $(E,\mathcal{E}) = \big(\mathbb{N},\mathcal{P}(\mathbb{N})\big)$ and μ be the counting measure. Then for all $f : \mathbb{N} \to \mathbb{R}$

$$\int_E f\,d\mu = \lim_{n\to\infty} \int_E f\,\mathbb{1}_{\{1,\ldots,n\}}\,d\mu = \lim_{n\to\infty} \sum_{k=1}^{n} f(k)\,\mu\big(\{k\}\big) = \sum_{k=1}^{\infty} f(k)\ .$$

By our definition: f integrable $\Leftrightarrow$ $|f|$ integrable $\Leftrightarrow$ $\sum_{k=1}^{\infty} |f(k)| < \infty$.
So $f(k) = (-1)^k/k$ is not integrable, although $\sum_{k=1}^{\infty} (-1)^k/k = \ln 2$ ($'='\ \infty - \infty$).

3.2 Integrals and limits

We are interested under which conditions $f_n \to f$ implies $\int f_n\,d\mu \to \int \lim f_n\,d\mu$.
Let $(E,\mathcal{E},\mu)$ be a measure space.

Theorem 3.5. Monotone convergence
*Let $f, f_1, f_2, \ldots : E \to \overline{\mathbb{R}}$ be measurable with $f_n \geq 0$ a.e. for all $n \in \mathbb{N}$ and $f_n \nearrow f$ a.e..
Then* $\int_E f_n\,d\mu \nearrow \int_E f\,d\mu$.

Proof. Suppose first that $f_n \geq 0$ and $f_n \nearrow f$ pointwise.
For each $n \in \mathbb{N}$ let $(f_n^k)_{k\in\mathbb{N}}$ be a sequence of $\mathcal{E}$-simple functions with $0 \leq f_n^k \nearrow f_n$ as $k \to \infty$ and let $g_n := \max\big\{f_1^n, \ldots, f_n^n\big\}$. Then g_n is an increasing sequence of $\mathcal{E}$-simple functions with $f_m^n \leq g_n \leq f_n$ for each $m \leq n$, $n \in \mathbb{N}$. Taking the limit $n \to \infty$ we get

$$f_m \leq g \leq f \quad \text{for each } m \in \mathbb{N} \quad \text{with } g = \lim_{n\to\infty} g_n : E \to [0,\infty]\ .$$

Taking the limit $m \to \infty$ gives $g = f$. Hence $\int_E f\,d\mu = \lim_{n\to\infty} \int_E g_n\,d\mu$ by Lemma 3.2. But

$$\int_E f_m^n\,d\mu \leq \int_E g_n\,d\mu \leq \int_E f_n\,d\mu \quad \text{for each } m \leq n,\ n \in \mathbb{N} \text{ and so with}$$

$$n \to \infty\ :\qquad \int_E f_m\,d\mu \leq \int_E f\,d\mu \leq \lim_{n\to\infty} \int_E f_n\,d\mu$$

$$m \to \infty\ :\qquad \lim_{m\to\infty} \int_E f_m\,d\mu \leq \int_E f\,d\mu \leq \lim_{n\to\infty} \int_E f_n\,d\mu\ .$$

Now, let $f_n \geq 0$ $a.e.$ and $f_n \nearrow f$ $a.e.$. Since $N = \bigcup_n \big(\{f_n < 0\} \cup \{f_n > f_{n+1}\}\big) \cup \{f_n \not\to f\}$ is a countable union of null sets, $\mu(N) = 0$. Then use monotone convergence on N^c to get

$$\int_E f_n \, d\mu = \int_{N^c} f_n \, d\mu \nearrow \int_{N^c} f \, d\mu = \int_E f \, d\mu \ . \qquad \square$$

Lemma 3.6. Fatou's lemma
Let $f_1, f_2, \ldots : E \to \overline{\mathbb{R}}$ be measurable functions with $f_n \geq 0$ $a.e.$ for all $n \in \mathbb{N}$. Then

$$\int_E \liminf_n f_n \, d\mu \leq \liminf_n \int_E f_n \, d\mu \ .$$

Proof. As previously, $N = \bigcup_n \{f_n < 0\}$ is a null set, so suppose $f_n \geq 0$ pointwise w.l.o.g..
Let $g_n := \inf_{k \geq n} f_k$. Then the g_n are measurable by Proposition 2.6 and $g_n \nearrow \liminf_n f_n$.
So since $f_n \geq g_n$ and by monotone convergence: $\displaystyle\int_E f_n \, d\mu \geq \int_E g_n \, d\mu \to \int_E \liminf_n f_n \, d\mu$
which proves the statement, taking $\liminf_n$ on the left-hand side. $\square$

Theorem 3.7. Dominated convergence
Let $f, f_1, f_2, \ldots : E \to \overline{\mathbb{R}}$ be measurable and $g, g_1, g_2, \ldots : E \to [0, \infty]$ be integrable with $f_n \to f$ a.e., $g_n \to g$ a.e., $|f_n| \leq g_n$ $a.e.$ for all $n \in \mathbb{N}$ and $\int_E g_n \, d\mu \to \int_E g \, d\mu < \infty$.
Then f and the f_n are integrable and $\displaystyle\int_E f_n \, d\mu \to \int_E f \, d\mu \ .$

Proof. f, f_n are integrable since with $|f_n| \leq g_n$ and $|f| \leq g$, $\int |f_n| \, d\mu$, $\int |f| \, d\mu < \infty$.
As before, $N = \bigcup_n \{f_n > |g_n|\} \cup \{f_n \not\to f\} \cup \{g_n \not\to g\}$ is a null set which does not affect the integral, so we assume pointwise validity of the assumptions w.l.o.g..
We have $0 \leq g_n \pm f_n \to g \pm f$, so $\liminf_n (g_n \pm f_n) = g \pm f$. By Fatou's lemma,

$$\int g \, d\mu + \int f \, d\mu = \int \liminf_n (g_n + f_n) \, d\mu \leq \liminf_n \int (g_n + f_n) \, d\mu = \int g \, d\mu + \liminf_n \int f_n \, d\mu \ ,$$

$$\int g \, d\mu - \int f \, d\mu = \int \liminf_n (g_n - f_n) \, d\mu \leq \liminf_n \int (g_n - f_n) \, d\mu = \int g \, d\mu - \limsup_n \int f_n \, d\mu \ .$$

Since $\int g \, d\mu < \infty$ it follows that

$$\int f \, d\mu \leq \liminf_n \int f_n \, d\mu \leq \limsup_n \int f_n \, d\mu \leq \int f \, d\mu \ ,$$

proving that $\int f_n \, d\mu \to \int f \, d\mu$ as $n \to \infty$. $\square$

Remark. If $g_n = g$ for all $n \in \mathbb{N}$ this is *Lebesgue's dominated convergence theorem*.

Corollary 3.8. Bounded convergence
Let $\mu(E) < \infty$ and $f, f_1, f_2 : E \to \overline{\mathbb{R}}$ be a measurable with $f_n \to f$ $a.e.$ and $|f_n| \leq C$ $a.e.$ for some $C \in \mathbb{R}$ and all $n \in \mathbb{N}$. Then f and the f_n are integrable and $\displaystyle\int_E f_n \, d\mu \to \int_E f \, d\mu \ .$

Proof. Apply dominated convergence with $g_n \equiv C$, noting that $\int g \, d\mu = C \, \mu(E) < \infty$.

The following example shows that the inequality in Lemma 3.6 can be strict and that domination by an integrable function in Theorem 3.7 is crucial.

Example. On $(\mathbb{R}, \mathcal{B}, \mu)$ with Lebesgue measure μ take $f_n = n^2\, \mathbb{1}_{(0,1/n)}$.
Then $f_n \searrow f \equiv 0$ pointwise, but $\int f\, d\mu = 0 < \int f_n\, d\mu = n \to \infty$.

Remarks. Equivalent series versions of Theorems 3.5 and 3.7:

(i) Let $(f_n)_{n\in\mathbb{N}}$, $f_n : E \to [0, \infty]$ measurable. Then $\displaystyle\sum_{n=1}^{\infty} \int_E f_n\, d\mu = \int_E \Big(\sum_{n=1}^{\infty} f_n\Big)\, d\mu$.

(ii) Let $(f_n)_{n\in\mathbb{N}}$, $f_n : E \to \overline{\mathbb{R}}$ measurable. If $\sum_{n=1}^{\infty} f_n$ converges and $\Big|\sum_{k=1}^{n} f_k\Big| \le g$, where g is integrable, then $\sum_{n=1}^{\infty} f_n,\ f_n$ are integrable and $\displaystyle\sum_{n=1}^{\infty} \int_E f_n\, d\mu = \int_E \Big(\sum_{n=1}^{\infty} f_n\Big)\, d\mu$.

Definition 3.3. Let $\mu, \mu_1, \mu_2, \dots$ be measures on $(\mathbb{R}, \mathcal{B})$. Say that μ_n *converges weakly* to μ, written $\mu_n \Rightarrow \mu$, if

$$\int_{\mathbb{R}} f\, d\mu_n \to \int_{\mathbb{R}} f\, d\mu \quad \text{for all } f \in C_b(\mathbb{R}, \mathbb{R})\ , \quad \text{i.e. } f : \mathbb{R} \to \mathbb{R} \text{ bounded and continuous .}$$

Theorem 3.9. *Let $X, X_1, X_2, \dots$ be random variables with distributions $\mu, \mu_1, \mu_2, \dots$.*
Then $X_n \xrightarrow{D} X \quad \Leftrightarrow \quad \mu_n \Rightarrow \mu \quad \big(\Leftrightarrow \quad \mathbb{E}\big(f(X_n)\big) \to \mathbb{E}\big(f(X)\big)$ *by Prop. 3.4*$\big)$.

Proof. Suppose $X_n \xrightarrow{D} X$. Then by the Skorohod theorem 2.12 there exist $Y \sim X$ and $Y_n \sim X_n$ on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that, $f(Y_n) \to f(Y)$ $a.e.$ since $f \in C_b(\mathbb{R}, \mathbb{R})$ (see also problem 2.6). Thus by bounded convergence

$$\int_{\mathbb{R}} f\, d\mu_n = \int_{\Omega} f(Y_n)\, d\mathbb{P} \to \int_{\Omega} f(Y)\, d\mathbb{P} = \int_{\mathbb{R}} f\, d\mu \qquad \text{so} \quad \mu_n \Rightarrow \mu\ .$$

Suppose $\mu_n \to \mu$ and let y be a continuity point of F_X.
For $\delta > 0$, approximate $\mathbb{1}_{(-\infty,y]}$ by $f_\delta(x) = \begin{cases} \mathbb{1}_{(-\infty,y]}(x) & ,\ x \notin (y, y+\delta) \\ 1 + (y-x)/\delta & ,\ x \in (y, y+\delta) \end{cases}$ such that

$$\Big|\int_{\mathbb{R}} (\mathbb{1}_{(-\infty,y]} - f_\delta)\, d\mu\Big| \le \Big|\int_{\mathbb{R}} g_\delta\, d\mu\Big| \quad \text{where} \quad g_\delta(x) = \begin{cases} 1 + (x-y)/\delta & ,\ x \notin (y-\delta, y) \\ 1 + (y-x)/\delta & ,\ x \in [y, y+\delta) \\ 0 & ,\ \text{otherwise} \end{cases} .$$

The same inequality holds for μ_n for all $n \in \mathbb{N}$. Then as $n \to \infty$

$$\big|F_{X_n}(y) - F_X(y)\big| = \Big|\int_{\mathbb{R}} \mathbb{1}_{(-\infty,y]}\, d\mu_n - \int_{\mathbb{R}} \mathbb{1}_{(-\infty,y]}\, d\mu\Big| \le$$
$$\le \Big|\int_{\mathbb{R}} g_\delta\, d\mu_n\Big| + \Big|\int_{\mathbb{R}} g_\delta\, d\mu\Big| + \Big|\int_{\mathbb{R}} f_\delta\, d\mu_n - \int_{\mathbb{R}} f_\delta\, d\mu\Big| \to 2\Big|\int_{\mathbb{R}} g_\delta\, d\mu\Big|\ ,$$

since $f_\delta, g_\delta \in C_b(\mathbb{R}, \mathbb{R})$. Now, $\big|\int_{\mathbb{R}} g_\delta\, d\mu\big| \le \mu\big((y-\delta, y+\delta)\big) \to 0$ as $\delta \to 0$, since $\mu\big(\{y\}\big) = 0$, so $X_n \xrightarrow{D} X$. □

3.3 Integration in $\mathbb{R}$ and differentiation

Theorem 3.10. Differentiation under the integral sign
Let $(E, \mathcal{E}, \mu)$ be a measure space, $U \subseteq \mathbb{R}$ be open and suppose that $f : U \times E \to \mathbb{R}$ satisfies:

(i) $x \mapsto f(t, x)$ *is integrable for all* $t \in U$,

(ii) $t \mapsto f(t, x)$ *is differentiable for all* $x \in E$,

(iii) $\left|\frac{\partial f}{\partial t}(t, x)\right| \leq g(x)$ *for some integrable* $g : E \to \mathbb{R}$ *and all* $x \in E$, $t \in U$.

Then $\frac{\partial f}{\partial t}(t, .)$ *is integrable for all* t*, the function* $F : U \to \mathbb{R}$ *defined by* $F(t) = \int_E f(t, x)\, \mu(dx)$ *is differentiable and* $\frac{d}{dt}F(t) = \int_E \frac{\partial f}{\partial t}(t, x)\, \mu(dx)$.

Proof. Take a sequence $h_n \to 0$ and set

$$g_n(t, x) := \frac{f(t + h_n, x) - f(t, x)}{h_n} - \frac{\partial f}{\partial t}(t, x)\ .$$

Then for all $x \in E, t \in U$, $g_n(t, x) \to 0$ and $|g_n(t, x)| \leq 2\, g(x)$ for all $n \in \mathbb{N}$ by the MVT. $\frac{\partial f}{\partial t}(t, .)$ is measurable as the limit of measurable functions, and integrable since $\left|\frac{\partial f}{\partial t}\right| \leq g$. Then by dominated convergence, as $n \to \infty$

$$\frac{F(t + h_n) - F(t)}{h_n} - \int_E \frac{\partial f}{\partial t}(t, x)\, \mu(dx) = \int_E g_n(t, x)\, \mu(dx) \to 0\ . \qquad \square$$

Remarks. (i) The integral on $\mathbb{R}$ w.r.t. Lebesgue measure μ is called *Lebesgue integral*.

We write $\int_{\mathbb{R}} f\, d\mu = \int_{-\infty}^{\infty} f(x)\, dx$ and $\int_{\mathbb{R}} f\, \mathbb{1}_{(a,b]}\, d\mu = \int_a^b f(x)\, dx\ .$

(ii) Linearity of the integral then implies: $\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx$
for all $c \in \mathbb{R}$, using the convention $\int_b^a f(x)\, dx = -\int_a^b f(x)\, dx$.

Theorem 3.11. Fundamental theorem of calculus

(i) Let $f : [a, b] \to \mathbb{R}$ *be a continuous function and set* $F_a(t) = \int_a^t f(x)\, dx\ .$
Then F_a *is differentiable on* $[a, b]$ *with* $F_a' = f$.

(ii) Let $F : [a, b] \to \mathbb{R}$ *be differentiable with continuous derivative* f*. Then*

$$\int_a^b f(x)\, dx = F(b) - F(a)\ .$$

Proof. (i) Fix $t \in [a,b)$. $\forall\,\epsilon>0\;\exists\,\delta>0\;:\;|x-y|<\delta\;\Rightarrow\;|f(x)-f(t)|<\epsilon$. So for $0<h\leq\delta$,

$$\Big|\frac{F_a(t+h)-F_a(t)}{h}-f(t)\Big| = \frac{1}{h}\Big|\int_t^{t+h}\big(f(x)-f(t)\big)\,dx\Big| \leq \frac{\epsilon}{h}\int_t^{t+h}dx = \epsilon\;.$$

Analogous for negative h and $t \in (a,b]$, thus $F_a' = f$.
(ii) $(F-F_a)'(t)=0$ for all $t\in(a,b)$ so by the MVT

$$F(b)-F(a) = F_a(b)-F_a(a) = \int_a^b f(x)\,dx\;. \qquad \square$$

So the methods of calculating Riemann integrals also apply to Lebesgue integrals.

Proposition 3.12. Partial integration and change of variable

(i) Let $u,v \in C^1\big([a,b],\mathbb{R}\big)$, i.e. differentiable with continuous derivative, then

$$\int_a^b u(x)\,v'(x)\,dx = \big[u(b)\,v(b)-u(a)\,v(a)\big] - \int_a^b u'(x)\,v(x)\,dx\;.$$

(ii) Let $\phi\in C^1\big([a,b],\mathbb{R}\big)$ be strictly increasing. Then

$$\int_{\phi(a)}^{\phi(b)} f(y)\,dy = \int_a^b f\big(\phi(x)\big)\,\phi'(x)\,dx \quad \textit{for all } f\in C\big([\phi(a),\phi(b)],\mathbb{R}\big)\;.$$

Proof. see problems 2.12 and 2.13

Definition 3.4. Let $(E,\mathcal{E},\mu)$ be a measure space and $f:E\to[0,\infty)$ be integrable. We say a measure ν on $(E,\mathcal{E})$ has *density f with respect to μ*, short $\nu = f\cdot\mu$, if

$$\nu(A) = \int_A f\,d\mu \quad \text{for all } A\in\mathcal{E}\;.$$

Lemma 3.13. *Let $(E,\mathcal{E},\mu)$ be a measure space. For every integrable $f:E\to[0,\infty)$, $\nu : A\mapsto\int_A f\,d\mu$ is a measure on $(E,\mathcal{E})$ with μ-density f and*

$$\int_E g\,d\nu = \int_E f\,g\,d\mu \quad \textit{for all integrable } g:E\to\overline{\mathbb{R}}\;.$$

Let μ be a Radon measure on $(\mathbb{R},\mathcal{B})$ with distribution function $F\in C^1(\mathbb{R},\mathbb{R})$. Then μ has density $f=F'$ with respect to Lebesgue measure.

Proof. For the first part see problem 2.15(a).
With Theorem 1.10 and 3.11, $\mu\big((a,b]\big) = F(b)-F(a) = \int_a^b f(x)\,dx\;.$
So ν coincides with $f\cdot\mu$ on the π-system $\mathcal{I}\cup\{\emptyset\} = \big\{(a,b]\;:\;a<b\big\}\cup\{\emptyset\}$ that generates $\mathcal{B}$. Thus by uniqueness of extension, $\nu = f\cdot\mu$ on $\mathcal{B}$ and ν has μ-density f. $\square$

3.4 Product measure and Fubini's theorem

Let $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ be finite measure spaces and $E = E_1 \times E_2$.

Definition 3.5. The *product σ-algebra* $\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 := \sigma(\mathcal{A})$ is generated by the π-system

$$\mathcal{A} = \big\{ A_1 \times A_2 \,:\, A_1 \in \mathcal{E}_1,\, A_2 \in \mathcal{E}_2 \big\} \,.$$

Example. If $E_1 = E_2 = \mathbb{R}$ and $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{B}$ then $\mathcal{E}_1 \otimes \mathcal{E}_2 = \mathcal{B}(\mathbb{R}^2)$.

Lemma 3.14. *Let $f : E \to \overline{\mathbb{R}}$ be $\mathcal{E}$-measurable. Then the following holds:*

(i) $f(x_1, .) : E_2 \to \overline{\mathbb{R}}$ is $\mathcal{E}_2$-measurable for all $x_1 \in E_1$.

(ii) If f is bounded, $f_1(x_1) := \displaystyle\int_{E_2} f(x_1, x_2)\, \mu_2(dx_2)$ is bounded and $\mathcal{E}_1$-measurable.

Proof. (i) For fixed $x_1 \in E_1$ define $T_{x_1} : E_2 \to E$ by $T_{x_1} x_2 = (x_1, x_2)$.
For $A = A_1 \times A_2 \in \mathcal{A}$, $T_{x_1}^{-1} A = \begin{cases} A_2 & , x_1 \in A_1 \\ \emptyset & , x_1 \notin A_1 \end{cases} \in \mathcal{E}_2$ and thus with Lemma 2.1 T_{x_1} is $\mathcal{E}_2/\mathcal{E}$-measurable. So $f(x_1, .) = f(T_{x_1}(.))$ is $\mathcal{E}_2/\overline{\mathcal{B}}$-measurable with Lemma 2.5.
(ii) By (i) and since f is bounded, f_1 is well defined and bounded, since $\mu_2(E_2) < \infty$.
For $f = \mathbb{1}_A$, $f_1(x_1) = \mu_2\big(T_{x_1}^{-1}(A)\big)$. Denote $\mathcal{D} = \big\{ A \in \mathcal{E} \,:\, f_1 \text{ is measurable} \big\}$, which can be checked to be a d-system. Since $f_1(x_1) = \mathbb{1}_{A_1}(x_1)\, \mu_2(A_2)$ for $A = A_1 \times A_2$, $\mathcal{A} \subseteq \mathcal{D}$ and thus $\mathcal{E} = \sigma(\mathcal{A}) = \mathcal{D}$ with Dynkin's lemma (1.5).
By linearity of integration the statement also holds for non-negative $\mathcal{E}$-simple functions, and by monotone convergence for all bounded, measurable f using

$$f_1(x_1) := \int_{E_2} f^+(x_1, x_2)\, \mu_2(dx_2) - \int_{E_2} f^-(x_1, x_2)\, \mu_2(dx_2) \,. \qquad \square$$

Theorem 3.15. Product measure
There exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on $\mathcal{E}$, such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\, \mu_2(A_2) \quad \text{for all } A_1 \in \mathcal{E}_1 \text{ and } A_2 \in \mathcal{E}_2 \,,$$

defined as $\displaystyle \mu(A) := \int_{E_1} \int_{E_2} \mathbb{1}_A(x_1, x_2)\, \mu_2(dx_2)\, \mu_1(dx_1) \,.$

Proof. With Lemma 3.14, μ is a well defined function of A. Using monotone convergence μ can be seen to be countably additive and is thus a measure.
Since $\mathbb{1}_{A_1 \times A_2} = \mathbb{1}_{A_1} \mathbb{1}_{A_2}$ the above property is fulfilled for all $A_1 \in \mathcal{E}_1$ and $A_2 \in \mathcal{E}_2$.
Since $\mathcal{A} = \{A_1 \times A_2 : A_i \in \mathcal{E}_i\}$ is a π-system generating $\mathcal{E}$ and $\mu(E) < \infty$, μ is uniquely determined by its values on $\mathcal{A}$ following Theorem 1.6 (Uniqueness of extension). $\square$

Remark. $f : E_1 \times E_2 \to \mathbb{R}$ is measurable if and only if $\hat{f} : E_2 \times E_1 \to \mathbb{R}$ with $\hat{f}(x_2, x_1) = f(x_1, x_2)$ is measurable and for integrable f: $\displaystyle\int_{E_2 \times E_1} \hat{f}\, d\mu_2 \otimes \mu_1 = \int_{E_1 \times E_2} f\, d\mu_1 \otimes \mu_2 \,.$

Theorem 3.16. Fubini's theorem

(i) Let $f : E \to [0,\infty]$ be $\mathcal{E}$-measurable. Then

$$\int_E f\,d\mu = \int_{E_1}\int_{E_2} f(x_1,x_2)\,\mu_2(dx_2)\,\mu_1(dx_1) = \int_{E_2}\int_{E_1} f(x_1,x_2)\,\mu_1(dx_1)\,\mu_2(dx_2)$$

taking values in $[0,\infty]$.

(ii) Let $f : E \to \overline{\mathbb{R}}$ be $\mathcal{E}$-measurable. If at least one of the following integrals is finite

$$\int_E |f|\,d\mu,\ \int_{E_1}\int_{E_2} |f(x_1,x_2)|\,\mu_2(dx_2)\,\mu_1(dx_1),\ \int_{E_2}\int_{E_1} |f(x_1,x_2)|\,\mu_1(dx_1)\,\mu_2(dx_2)$$

then all three are finite and f is integrable. Furthermore,
$f(x_1,.)$ is μ_2-integrable for μ_1-almost all x_1 and $\int_{E_2} f(.,x_2)\,\mu_2(dx_2)$ is μ_1-integrable,
$f(.,x_2)$ is μ_1-integrable for μ_2-almost all x_2 and $\int_{E_1} f(x_1,.)\,\mu_1(dx_1)$ is μ_2-integrable,
and the formula in (i) holds.

Proof. (i) If $f = \mathbb{1}_A$ for some $A \in \mathcal{E}$ the formula holds by definition of μ and can be extended to non-negative measurable f as in the proof of Lemma 3.14 (ii).
(ii) Since $|f|$ is non-negative, the formula in (i) holds and all integrals coincide and are finite. By Lemma 3.14 $f^\pm(x_1,.)$ is measurable and μ_2-integrable since

$$\int_{E_2} f^\pm(x_1,x_2)\,\mu_2(dx_2) \le \int_{E_2} \big|f(x_1,x_2)\big|\,\mu_2(dx_2) < \infty \quad \text{for } \mu_1\!-\!a.e.\ x_1 \in E_1$$

Furthermore $\displaystyle\int_{E_1}\int_{E_2} f^\pm(x_1,x_2)\,\mu_2(dx_2)\,\mu_1(dx_1) \le \int_E |f|\,d\mu < \infty\,.$
The same follows for $f(.,x_2)$ and finally the formula in (i) holds for $f^\pm$ and thus for $f = f^+ - f^-$ by linearity. □

Remarks. (i) Product measures and Fubini can be extended to σ-*finite* measure spaces, i.e. for all $A \in \mathcal{E}_1$ there exist $A_n \in \mathcal{E}_1$, $n \in \mathbb{N}$ with $\mu_1(A_n) < \infty$ for all n and $A = \bigcup_n A_n$.

(ii) However, without σ-finiteness Fubini's theorem does in general not hold. Consider e.g. the measure $\nu(\emptyset) = 0$, $\nu(A) = \infty$ for $a \neq \emptyset$ on $(\mathbb{R},\mathcal{B})$. This is not σ-finite and with Lebesgue measure μ on $(\mathbb{R},\mathcal{B})$ we have

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_Q(x+y)\,\mu(dx)\,\nu(dy) = 0 \quad \text{whereas} \quad \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_Q(x+y)\,\nu(dy)\,\mu(dx) = \infty\,.$$

(iii) The operation of taking products of measure spaces is associative

$$\mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \mathcal{E}_3 := \big(\mathcal{E}_1 \otimes \mathcal{E}_2\big) \otimes \mathcal{E}_3 = \mathcal{E}_1 \otimes \big(\mathcal{E}_2 \otimes \mathcal{E}_3\big) \quad \text{(also for measures)}\,.$$

So products can be taken without specifying the order, e.g. $\big(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu^d\big)$.

Example. $I = \int_{\mathbb{R}} e^{-x^2}\, dx = \sqrt{\pi}$, since by Fubini's theorem and polar-doordinates

$$I^2 = \int_{\mathbb{R}^2} e^{-(x^2+y^2)}\, dx\, dy = \int_{r=0}^{\infty} \int_{\phi=0}^{2\pi} e^{-r^2}\, r\, dr\, d\phi = 2\pi \big[- e^{-r^2}/2 \big]_0^{\infty} = \pi\ .$$

Proposition 3.17. *Let* $(E_1, \mathcal{E}_1, \mu_1)$ *be a* σ*-finite measure space. Then*

$$\int_E f\, d\mu = \int_0^{\infty} \mu(f \geq x)\, dx \quad \textit{for all } \mathcal{E}_1/\mathcal{B}\textit{-measurable } f : \mathcal{E}_1 \to [0, \infty)\ .$$

Proof. see problem 2.15(b)

Remark. Together with Proposition 3.4, this consequence of Fubini's theorem is particularly useful to calculate expectations of random variables.

4 L^p-spaces

4.1 Norms and inequalities

Let $(E, \mathcal{E}, \mu)$ be a measure space.

Theorem 4.1. Chebyshev's inequality
Let $f : E \to [0, \infty]$ be measurable. Then for any $\lambda \geq 0$: $\quad \lambda\, \mu(f \geq \lambda) \leq \int_E f\, d\mu\,.$

Proof. Integrate $f \geq \lambda\, \mathbb{1}_{\{f \geq \lambda\}}$. □

Example. Let X be a random variable with $m = \mathbb{E}(X) < \infty$. Then take $f = |X - m|^2$ to get $\quad \mathbb{P}(f \geq \lambda^2) = \mathbb{P}\big(|X - m| \geq \lambda\big) \leq \mathrm{Var}(X)/\lambda^2 \quad$ for all $\lambda > 0$.

Definition 4.1. For $1 \leq p \leq \infty$ we denote by $L^p = L^p(E, \mathcal{E}, \mu)$ the set of measurable functions $f : E \to \overline{\mathbb{R}}$ with finite L^p*-norm*, $\|f\|_p < \infty$, where

$$\|f\|_p = \Big(\int_E |f|^p\, d\mu \Big)^{1/p} \quad \text{for } p < \infty\,, \qquad \|f\|_\infty = \inf \big\{ \lambda \in \mathbb{R} : |f| \leq \lambda\ a.e. \big\}\,.$$

We say that f_n *converges to f in L^p*, $f_n \xrightarrow{L^p} f$, if $\quad \|f - f_n\|_p \to 0$ as $n \to \infty$.

Remarks. (i) At the end of this section we will see in what sense $\|.\|_p$ is a norm on L^p.

(ii) For $f \in C(\mathbb{R})$, $\|f\|_\infty = \sup_{x \in E} |f(x)|$.

(iii) For $1 \leq p < \infty$: $\quad \|f\|_p \leq \mu(E)^{1/p}\, \|f\|_\infty$.

(iv) Let $f \in L^p$, $1 \leq p < \infty$. Then $\quad \mu\big(|f| \geq \lambda\big) \leq (\|f\|_p/\lambda)^p \quad$ for all $\lambda > 0$ by Chebyshev's inequality.
For $f \in L^p(\mathbb{R})$ this includes that $f(x)$ essentially tends to zero as $|x| \to \infty$ in the sense $\|f\, \mathbb{1}_{|x| \geq y}\|_\infty \to 0$ as $y \to \infty$. For random variables $X \in L^p$ the relation $\mathbb{P}\big(|X| \geq \lambda\big) = \mathcal{O}(\lambda^{-p}) \quad$ as $\lambda \to \infty$ is called a *tail estimate* $\quad$ (see also problem 3.6) .

Definition 4.2. A function $f : \mathbb{R} \to \mathbb{R}$ is *convex* if, for all $x, y \in \mathbb{R}$ and $t \in [0, 1]$

$$f\big(t\, x + (1 - t)\, y\big) \leq t\, f(x) + (1 - t)\, f(y)\,.$$

Remark. Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Then f is continuous (in particular measurable) and

$$\forall\, x_0 \in \mathbb{R}\ \exists\, a \in \mathbb{R}\ :\ f(x) \geq a\, (x - x_0) + f(x_0)\,.$$

Theorem 4.2. Jensen's inequality
Let X be an integrable r.v. and $f : \mathbb{R} \to \mathbb{R}$ convex. Then $\quad \mathbb{E}\big(f(X)\big) \geq f\big(\mathbb{E}(X)\big)$.

Proof. With $m = \mathbb{E}(X) < \infty$ choose $a \in \mathbb{R}$ such that $\quad f(X) \geq a\,(X - m) + f(m)$. In particular $\mathbb{E}\big(f(X)^-\big) \leq |a|\, \mathbb{E}(|X|) + |f(m)| < \infty$ and $\mathbb{E}\big(f(X)\big) \in \overline{\mathbb{R}}$ is well defined. Moreover

$$\mathbb{E}\big(f(X)\big) \geq a\, \big(\mathbb{E}(X) - m\big) + f(m) = f(m) = f\big(\mathbb{E}(X)\big)\,.$$ □

Theorem 4.3. Hölder's inequality
Let $p, q \in [1, \infty]$ *be conjugate indices, i.e.* $\frac{1}{p} + \frac{1}{q} = 1$ *or equivalently* $q = p/(p-1)$*. Then for all measurable* $f, g : E \to \mathbb{R}$ *,* $\displaystyle\int_E |f\,g|\,d\mu = \|f\,g\|_1 \le \|f\|_p\,\|g\|_q\,.$

Equality holds if and only if $\dfrac{|f(x)|^p}{\|f\|_p^p} = \dfrac{|g(x)|^q}{\|g\|_q^q}$ *a.e.* .

Proof. For $p = 1$, $q = \infty$ the result follows with $|f\,g| \le |f|\,\|g\|_\infty$ *a.e.*. If $\|f\|_p$, $\|g\|_q = 0$ or ∞ the result is trivial, so in the following $p, q \in (1, \infty)$ and $\|f\|_p$, $\|f\ g_q \in (0, \infty)$.
For given $0 < a, b < \infty$ let $a = e^{s/p}$, $b = e^{t/q}$ and by convexity of e^x we get

$$e^{s/p+t/q} \le \frac{e^s}{p} + \frac{e^t}{q} \quad \text{and thus} \quad a\,b \le \frac{a^p}{p} + \frac{b^q}{q}\,. \qquad \text{(Young's inequality)}$$

By strict convexity of e^x equality holds if and only if $s = t \quad \Leftrightarrow \quad b = a^{p-1}$.
Now insert $a = |f|/\|f\|_p$ and $b = |g|/\|g\|_q$ and integrate

$$\|f\,g\|_1 \le \|f\|_p\,\|g\|_q \Big(\frac{\int |f|^p\,d\mu}{p\,\|f\|_p^p} + \frac{\int |g|^q\,d\mu}{q\,\|g\|_q^q} \Big) = \|f\|_p\,\|g\|_q \big(\tfrac{1}{p} + \tfrac{1}{q}\big) = \|f\|_p\,\|g\|_q\,.$$

After integration equality holds if and only if $b = a^{p-1}$ *a.e.*, finishing the proof. □

Corollary 4.4. Minkowski's inequality
For $p \in [1, \infty]$ *and measurable* $f, g : E \to \mathbb{R}$ *we have* $\|f + g\|_p \le \|f\|_p + \|g\|_p$.

Remark. This is the triangle inequality for p-norms. For every norm this implies the *negative triangle inequality* $\|f_n - f\| \ge \big|\|f_n\| - \|f\|\big|$, and thus $\|f_n\| \to \|f\|$ whenever $\|f_n - f\| \to 0$.

Proof. The cases $p = 1, \infty$ follow directly from the triangle inequality on $\mathbb{R}$, so assume $p \in (1, \infty)$ and $\|f\|_p, \|g\|_p < \infty$, $\|f + g\|_p > 0$.
Then $|f + g|^p \le \big(2\,(|f| \vee |g|)\big)^p \le 2^p\big(|f|^p + |g|^p\big)$ so $\|f + g\|_p < \infty$.
The result then follows from

$$\begin{aligned}\|f + g\|_p^p &= \int_E |f + g|^p\,d\mu \le \int_E |f|\,|f + g|^{p-1}\,d\mu + \int_E |g|\,|f + g|^{p-1}\,d\mu \le \\ &\le \Big(\|f\|_p + \|g\|_p\Big)\Big(\int_E |f + g|^{(p-1)q}\,d\mu\Big)^{1/q} \qquad \text{using Hölder with } q = \tfrac{p}{p-1} \\ &= \Big(\|f\|_p + \|g\|_p\Big)\,\|f + g\|_p^{p-1}\,.\end{aligned}$$

□

Corollary 4.5. Monotonicity of L^p-norms
Let $1 \le p < q \le \infty$*. Then for all* $f \in L^p(\mu)$ *we have* $\|f\|_p \le \mu(E)^{1/p-1/q}\|f\|_q$ *,*
which includes $L^q(\mu) \subseteq L^p(\mu)$ *in the case* $\mu(E) < \infty$ *.*

Proof. For $q = \infty$ the result follows from Remark (iii) on the previous page. For $q < \infty$ apply Hölder with indices $\tilde{p} = q/p$ and $\tilde{q} = \tilde{p}/(\tilde{p} - 1) = q/(q - p)$ to get

$$\|f\|_p^p = \int_E |f|^p \cdot 1\,d\mu \le \Big(\int_E |f|^{p\tilde{p}}\,d\mu\Big)^{1/\tilde{p}} \mu(E)^{1/\tilde{q}} = \|f\|_q^p\,\mu(E)^{1-p/q}\,.$$

Since $x \mapsto x^p$ is monotone increasing for $x \ge 0$ this implies the result. □

Examples. (i) There is no monotonicity of L^p-norms if $\mu(E)=\infty$. Take e.g. $f(x)=1/x$ on $(0,\infty)$ with Lebesgue measure. Then $\|f\mathbb{1}_{[1,\infty)}\|_1 = \infty > 1 = \|f\mathbb{1}_{[1,\infty)}\|_2$ and $\|\sqrt{f}\mathbb{1}_{(0,1)}\|_1 = 2 < \infty = \|\sqrt{f}\mathbb{1}_{(0,1)}\|_2$.

(ii) Consider the counting measure $\mu=\sum_n \delta_n$ on the measurable space $\big(\mathbb{N},\mathcal{P}(\mathbb{N})\big)$. Then

$$\|f\|_p = \Big(\sum_{n=1}^\infty \big|f(n)\big|^p\Big)^{1/p} \quad \text{for } p<\infty \quad \text{and} \quad \|f\|_\infty = \sup_{n\in\mathbb{N}} \big|f(n)\big| \ .$$

So $L^p(\mathbb{N},\mu)=\ell^p$ is the space of sequences with finite p-norm.

Definition 4.3. For $f,g\in L^p$ we say that g is a *version* of f if $g=f\ \ a.e.$. This defines an equivalence relation on L^p and we denote by $\mathcal{L}^p = L^p/[0]$ the quotient space of all equivalence classes $[f]=\{g\in L^p\ :\ g-f=0\ a.e.\}$.

Proposition 4.6. $(\mathcal{L}^p,\|.\|_p)$ *is a normed vector space.*

Proof. If $f,g\in L^p$ with $f=g\ a.e.$ then $\|f\|_p=\|g\|_p<\infty$ by Theorem 3.3, so $\|.\|_p$ is well defined on $\mathcal{L}^p$. In particular $f=0\ a.e.$ implies $\|f\|_p=0$. Furthermore $\|\lambda f\|_p = |\lambda|\,\|f\|_p$ for all $\lambda\in\mathbb{R}$ by linearity of integration and $\|f+g\|_p\le\|f\|_p+\|g\|_p$ by Minkowski's inequality. These properties extend to equivalence classes. In particular $[f],[g]\in\mathcal{L}^p$ implies that $\lambda[f]=[\lambda f]$ and $[f]+[g]=[f+g]$ are in $\mathcal{L}^p$, so that $\mathcal{L}^p$ is a vector space. □

Remark. In the following we follow the usual abuse of notation and identify $\mathcal{L}^p$ with L^p.

Theorem 4.7. Completeness of L^p
$(L^p,\|.\|_p)$ *is a Banach space, i.e. a complete normed vector space, for every* $p\in[1,\infty]$.

Proof. The case $p=\infty$ is left as problem 3.3, in the following $p<\infty$.
Let $(f_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in L^p such that $\|f_n-f_m\|_p\to 0$ as $n,m\to\infty$.

Choose a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $S:=\sum_{k=1}^\infty \|f_{n_{k+1}}-f_{n_k}\|_p<\infty$.

By Minkowski's inequality, for any $K\in\mathbb{N}$, $\Big\|\sum_{k=1}^K |f_{n_{k+1}}-f_{n_k}|\Big\|_p \le S$.

By monotone convergence this bound holds also for $K\to\infty$, so $\sum_{k=1}^\infty |f_{n_{k+1}}-f_{n_k}|<\infty\ a.e.$

So for a.e. $x\in\mathbb{R}$, $f_{n_k}(x)$ is Cauchy and thus converges by completeness of $\mathbb{R}$. We define

$$f(x) := \begin{cases} \lim_{k\to\infty} f_{n_k}(x) & \text{, if the limit exists} \\ 0 & \text{, otherwise} \end{cases} .$$

Given $\epsilon>0$, we can find $N\in\mathbb{N}$ such that $\int |f_n-f_m|^p\,d\mu<\epsilon$ for all $m\ge n\ge N$, and in particular $\int |f_n-f_{n_k}|^p\,d\mu<\epsilon$ for sufficiently large k. Hence by Fatou's Lemma

$$\int |f_n-f|^p\,d\mu = \int \liminf_k |f_n-f_{n_k}|^p\,d\mu \le \liminf_k \int |f_n-f_{n_k}|^p\,d\mu < \epsilon \ \text{ for all } n\ge N \ .$$

Hence $\|f_n - f\|_p \to 0$ since $\epsilon > 0$ was arbitrary and $f \in L^p$ since for n large enough

$$\|f\|_p \le \|f - f_n\|_p + \|f_n\|_p \le 1 + \|f_n\|_p < \infty . \qquad \square$$

4.2 L^2 as a Hilbert space

Let $(E, \mathcal{E}, \mu)$ be a measure space and $L^2 = L^2(E, \mathcal{E}, \mu)$.

Proposition 4.8. *The form $\langle .,. \rangle : L^2 \times L^2 \to \mathbb{R}$ with $\langle f, g \rangle = \int_E f\, g\, d\mu$ is an inner product on L^2, and the inner product space $(L^2, \langle .,. \rangle)$ is a Hilbert space, i.e. complete with respect to the norm $\|f\|_2 = \sqrt{\langle f, f \rangle}$.*

Proof. By Hölder's inequality we have for all $f, g \in L^2$

$$\big|\langle f, g \rangle\big| \le \int_E |f\, g|\, d\mu \le \|f\|_2 \|g\|_2 \qquad \text{(Cauchy-Schwarz inequality)} .$$

Thus $\langle .,. \rangle$ is finite and well defined on L^2, symmetric by definition and bilinear by linearity of integration. Further $\langle f, f \rangle = \|f\|_2^2 \ge 0$ with equality if and only if $f = 0$ $a.e.$ and $(L^2, \|.\|_2)$ is complete by Theorem 4.7. $\square$

Proposition 4.9. *For $f, g \in L^2$ we have* ***Pythagoras' rule***

$$\|f + g\|_2^2 = \|f\|_2^2 + 2\,\langle f, g \rangle + \|g\|_2^2 ,$$

and the ***parallelogram law***

$$\|f + g\|_2^2 + \|f - g\|_2^2 = 2\big(\|f\|_2^2 + \|g\|_2^2\big) .$$

Proof. Follows directly from $\|f \pm g\|_2^2 = \langle f \pm g, f \pm g \rangle$. $\square$

Definition 4.4. We say $f, g \in L^2$ are *orthogonal* if $\langle f, g \rangle = 0$. For $V \subseteq L^2$, we define

$$V^\perp = \big\{ f \in L^2 \,:\, \langle f, v \rangle = 0 \text{ for all } v \in V \big\} .$$

$V \subseteq L^2$ is called *closed* if, for every sequence $(f_n)_{n\in\mathbb{N}}$ in V, with $f_n \to f$ in L^2, we have $f = v$ $a.e.$, for some $v \in V$.

Remark. For all $V \subseteq L^2$, $V^\perp$ is a closed subspace, since $f, g \in V^\perp$ includes $\lambda_1 f + \lambda_2 g \in V^\perp$ for all $\lambda_1, \lambda_2 \in \mathbb{R}$ and for $(f_n)_{n\in\mathbb{N}}$ in $V^\perp$ with $f_n \to f$ in L^2 we have for all $v \in V$

$$\big|\langle f, v \rangle\big| = \big\langle f - f_n, v \rangle\big| \le \|f_n - f\|_2 \|v\|_2 \to 0 \quad \text{as } n \to \infty .$$

Theorem 4.10. Orthogonal projection
Let V be a closed subspace of L^2. Then each $f \in L^2$ has a decomposition $f = v + u$, with $v \in V$ and $u \in V^\perp$. The decomposition is unique up to a version and v is called the ***orthogonal projection*** *of f on V. Moreover, $\|f - v\|_2 \le \|f - g\|_2$ for all $g \in V$, with equality iff $g = v$ $a.e.$.*

Proof. Uniqueness: Suppose $f = v + u = \tilde{v} + \tilde{u}$ $a.e.$ with $v, \tilde{v} \in V$ and $u, \tilde{u} \in V^{\perp}$. Then with Pythagoras' rule

$$0 = \|v - \tilde{v} + u - \tilde{u}\|_2^2 = \|v - \tilde{v}\|_2^2 + \|u - \tilde{u}\|_2^2 \quad \Rightarrow \quad \tilde{u} = u, \tilde{v} = v \; a.e.\,.$$

Existence: Choose a sequence $v_n \in V$ such that, as $n \to \infty$,

$$\|f - v_n\|_2 \to d(f, V) := \inf\big\{\|f - g\|_2 \, : \, g \in V\big\}\,.$$

By the parallelogram law,

$$\big\|2\big(f - (v_n + v_m)/2\big)\big\|_2^2 + \|v_n - v_m\|_2^2 = 2\big(\|f - v_n\|_2^2 + \|f - v_m\|_2^2\big)\,.$$

But $\big\|2\big(f - (v_n + v_m)/2\big)\big\|_2^2 \geq 4\, d(f, V)^2$, so we must have $\|v_n - v_m\|_2 \to 0$ as $n, m \to \infty$. By completeness, $\|v_n - g\|_2 \to 0$, for some $g \in L^2$, and by closure $g = v$ a.e., for some $v \in V$. Hence

$$\|f - v\|_2 = \lim_{n\to\infty} \|f - v_n\|_2 = d(f, V) \leq \|f - h\|_2 \quad \text{for all } h \in V\,.$$

In particular, for all $t \in \mathbb{R}$, $h \in V$, we have

$$d(f, V)^2 \leq \big\|f - (v + t\, h)\big\|_2^2 = d(f, V)^2 - 2t\,\langle f - v, h\rangle + t^2\,\|h\|_2^2\,.$$

So we must have $\langle f - v, h\rangle = 0$, and $u = f - v \in V^{\perp}$, as required. □

Definition 4.5. For $\mathbb{R}$-valued random variables $X, Y \in L^2(\mathbb{P})$ with means $m_X = \mathbb{E}(X)$ and $m_Y = \mathbb{E}(Y)$ we define *variance*, *covariance* and *correlation* by

$$\mathrm{var}(X) = \mathbb{E}\big((X - m_X)^2\big)\,, \quad \mathrm{cov}(X, Y) = \mathbb{E}\big((X - m_X)(Y - m_Y)\big)\,,$$
$$\mathrm{corr}(X, Y) = \mathrm{cov}(X, Y)\big/\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}\,.$$

For an $\mathbb{R}^n$-valued random variable $X = (X_1, \ldots, X_n) \in L^2(\mathbb{P})$ (this means that each coordinate $X_i \in L^2(\mathbb{P})$) the variance is given by the *covariance matrix*

$$\mathrm{var}(X) = \big(\mathrm{cov}(X_i, X_j)\big)_{i,j=1,..,n}\,.$$

Remarks. (i) $\mathrm{var}(X) = 0$ if and only if $X = m_X$ $a.s.$.

(ii) $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$, $\mathrm{cov}(X, X) = \mathrm{var}(X)$ and if X and Y are independent, then $\mathrm{cov}(X, Y) = 0$.

(iii) By Hölder $\big|\mathrm{cov}(X, Y)\big| \leq \big|\mathrm{var}(X)\,\mathrm{var}(Y)\big|$ and thus $\mathrm{corr}(X, Y) \in [-1, 1]$.

Proposition 4.11. *Every covariance matrix is symmetric and non-negative definite.*

Proof. Symmetry by definition. For $X = (X_1, \ldots, X_n) \in L^2(\mathbb{P})$ and $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$

$$a^T \mathrm{var}(X) a = \sum_{i,j=1}^{n} a_i a_j \mathrm{cov}(X_i, X_j) = \mathrm{var}(a^t X) \geq 0\,,$$

since $a^T X = \sum_i a_i X_i \in L^2(\mathcal{P})$. □

Revision. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $G \in \mathcal{A}$ be some event. For $\mathbb{P}(G) > 0$ the *conditional probability* $\mathbb{P}(.\,|\,G)$, given by

$$\mathbb{P}(A\,|\,G) = \frac{\mathbb{P}(A \cap G)}{\mathbb{P}(G)} \quad \text{for all } A \in \mathcal{A} ,$$

is a probability measure on $(\Omega, \mathcal{A})$.

Definition 4.6. For a random variable $X : \Omega \to \mathbb{R}$ we denote

$$\mathbb{E}(X\,|\,G) = \int_\Omega X \, d\mathbb{P}(.|G) = \Big(\int_\Omega X \, \mathbb{1}_G \, d\mathbb{P} \Big) \Big/ \mathbb{P}(G) = \mathbb{E}(X \, \mathbb{1}_G)/\mathbb{P}(G) ,$$

whenever $\mathbb{P}(G) > 0$, and we set $\mathbb{E}(X\,|\,G) = 0$ when $\mathbb{P}(G) = 0$.
Let $(G_i)_{i \in I}$ be a countable family of disjoint events with $\bigcup_i G_i = \Omega$ and $\mathcal{G} = \sigma(G_i \,:\, i \in I)$. Then the *conditional expectation* of a r.v. X given $\mathcal{G}$ is given by

$$\mathbb{E}(X\,|\,\mathcal{G}) = \sum_{i \in I} \mathbb{E}(X\,|\,G_i) \, \mathbb{1}_{G_i} .$$

Remarks. (i) $\mathbb{E}(X\,|\,\mathcal{G})$ is a $\mathcal{G}/\mathcal{B}$-measurable r.v., taking constant values on each G_i.
In particular, for $\mathcal{G} = \sigma(\Omega) = \{\emptyset, \Omega\}$, $\quad \mathbb{E}\big(X\,|\,\{\emptyset, \Omega\}\big) = \mathbb{E}(X\,|\,\Omega)\,\mathbb{1}_\Omega = \mathbb{E}(X)$.

(ii) For every $A \in \mathcal{G}$ it is $A = \bigcup_{i \in J} G_i$ for some $J \subseteq I$. Thus

$$\int_A \mathbb{E}(X\,|\,\mathcal{G}) \, d\mathbb{P} = \sum_{i \in I} \mathbb{E}(X \, \mathbb{1}_{G_i}) \int_A \mathbb{1}_{G_i} \, d\mathbb{P} \big/ \mathbb{P}(G_i) = \sum_{i \in J} \mathbb{E}(X \, \mathbb{1}_{G_i}) = \int_A X \, d\mathbb{P} .$$

In particular, if $\mathbb{E}(X) < \infty$, $\mathbb{E}(X\,|\,\mathcal{G})$ is integrable and $\mathbb{E}\big(\mathbb{E}(X\,|\,\mathcal{G})\big) = \mathbb{E}(X)$.

(iii) For a σ-algebra $\mathcal{G} \subseteq \mathcal{A}$, $L^2(\mathcal{G}, \mathbb{P})$ is complete and therefore a closed subspace of $L^2(\mathcal{A}, \mathbb{P})$. If $X \in L^2(\mathcal{A}, \mathbb{P})$ then $\mathbb{E}(X\,|\,\mathcal{G}) \in L^2(\mathcal{G}, \mathbb{P})$.

Proposition 4.12. *If $X \in L^2(\mathcal{A}, \mathbb{P})$ then $\mathbb{E}(X\,|\,\mathcal{G})$ is a version of the orthogonal projection of X on $L^2(\mathcal{G}, \mathbb{P})$.*

Proof. see problem 3.10

Remarks on the general case

(i) For a general σ-algebra $\mathcal{F} \subseteq \mathcal{A}$ one can show, that for every integrable r.v. X there exists an $\mathcal{F}$-measurable, integrable r.v. Y with $\int_F Y \, d\mathbb{P} = \int_F X \, d\mathbb{P}$ for every $F \in \mathcal{F}$. It is unique up to a version, defining the *conditional expectation* $Y = \mathbb{E}(X\,|\,\mathcal{F})$.
For $X \in L^2(\mathcal{A}, \mathbb{P})$, $\mathbb{E}(X\,|\,\mathcal{F})$ is the orthogonal projection of X on $L^2(\mathcal{F}, \mathbb{P})$.

(ii) If X is $\mathcal{F}$-measurable, $\mathbb{E}(X\,|\,\mathcal{F}) = X$. In particular $\mathbb{E}(X\,|\,\mathcal{A}) = X$.

(iii) For σ-algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{A}$ we have

$$\mathbb{E}\big(\mathbb{E}(X\,|\,\mathcal{F}_2)\,\big|\,\mathcal{F}_1\big) = \mathbb{E}(X\,|\,\mathcal{F}_1) = \mathbb{E}\big(\mathbb{E}(X\,|\,\mathcal{F}_1)\,\big|\,\mathcal{F}_2\big) .$$

4.3 Convergence in L^1

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and consider $L^1 = L^1(\Omega, \mathcal{A}, \mathbb{P})$.
By monotonicity of L^p-norms $X_n \xrightarrow{L^p} X$ implies $X_n \xrightarrow{L^q} X$ for all $1 \leq q \leq p$, so convergence in L^1 is the weakest. From problem 3.4 we know that $X_n \xrightarrow{L^1} X$ implies convergence in probability. The converse holds only under additional assumptions.

Theorem. 4.13. Bounded convergence
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables with $X_n \to X$ in probability. If in addition $|X_n| \leq C\ \ a.s.$ for all $n \in \mathbb{N}$ and some $C < \infty$, then $X_n \to X$ in L^1.

Proof. By Theorem 2.10(ii) X is the almost sure limit of a subsequence, so $|X| \leq C\ \ a.s.$.
For $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$: $\mathbb{P}\big(|X_n - X| > \epsilon/2\big) \leq \epsilon/(4C)$. Then

$$\mathbb{E}\big(|X_n - X|\big) = \mathbb{E}\big(|X_n - X|\,\mathbb{1}_{|X_n - X| > \epsilon/2}\big) + \mathbb{E}\big(|X_n - X|\,\mathbb{1}_{|X_n - X| \leq \epsilon/2}\big) \leq \\ \leq 2\,C\big(\epsilon/(4C)\big) + \epsilon/2 = \epsilon\ .$$

□

Remark. Corollary 3.8 on bounded convergence gives a similar statement under the stronger assumption $X_n \to X\ a.s.$. Although the assumptions in 4.13 are weaker, they are still not necessary for the conclusion to hold. The main motivation of this section is to provide a necessary and sufficient extra condition, such that convergence in probability implies convergence in L^1.

Lemma 4.14. *For $X \in L^1(\mathcal{A}, \mathbb{P})$ set* $\quad I_X(\delta) = \sup\big\{\mathbb{E}(|X|\,\mathbb{1}_A)\ :\ A \in \mathcal{A},\, \mathbb{P}(A) \leq \delta\big\}$.
Then $I_X(\delta) \searrow 0$ as $\delta \searrow 0$.

Proof. Suppose not. Then, for some $\epsilon > 0$, there exist $A_n \in \mathcal{A}$, with $\mathbb{P}(A_n) \leq 2^{-n}$ and $\mathbb{E}\big(|X|\,\mathbb{1}_{A_n}\big) \geq \epsilon$ for all $n \in \mathbb{N}$. By the first Borel-Cantelli lemma, $\mathbb{P}(A_n\ i.o.) = 0$. But then by dominated convergence

$$\epsilon \leq \mathbb{E}\big(|X|\,\mathbb{1}_{\cup_{m\geq n} A_m}\big) \to \mathbb{E}\big(|X|\,\mathbb{1}_{\{A_n\ i.o.\}}\big) = 0 \quad \text{as } n \to \infty\ ,$$

which is a contradiction. □

Definition 4.7. Let $\mathcal{X}$ be a family of random variables on $(\Omega, \mathcal{A}, \mathbb{P})$. For $1 \leq p \leq \infty$ we say that $\mathcal{X}$ is *uniformly bounded in L^p* if $\ \sup\big\{\|X\|_p\ :\ X \in \mathcal{X}\big\} < \infty$. Define

$$I_{\mathcal{X}}(\delta) = \sup\big\{\mathbb{E}(|X|\,\mathbb{1}_A)\ :\ X \in \mathcal{X},\, A \in \mathcal{A},\, \mathbb{P}(A) \leq \delta\big\}\ .$$

We say that $\mathcal{X}$ is *unif. integrable* (*UI*) if $\mathcal{X}$ is unif. bounded in L^1 and $I_{\mathcal{X}}(\delta) \searrow 0$, as $\delta \searrow 0$.

Remarks. (i) $\mathcal{X}$ is unif. bounded in L^1 if and only if $I_{\mathcal{X}}(1) = \sup\big\{\|X\|_1 : X \in \mathcal{X}\big\} < \infty$.

(ii) With Lemma 4.14, any single, integrable random variable is UI, which can easily be extended to finitely many.

(iii) If $(\Omega, \mathcal{A}, \mathbb{P}) = \big((0,1], \mathcal{B}((0,1]), \mu\big)$ then if $I_{\mathcal{X}}(\delta) \searrow 0$ there exists $\delta > 0$ such that

$$\mathbb{E}\big(|X|\big) = \sum_{k=0}^{n-1} \mathbb{E}\big(|X|\,\mathbb{1}_{(k/n,(k+1)/n]}\big) \leq n \quad \text{for } n = \lceil 1/\delta \rceil \quad \text{and all } X \in \mathcal{X}\ ,$$

which includes that $\mathcal{X}$ is uniformly bounded. In general this does not hold.

(iv) Some sufficient conditions: $\mathcal{X}$ is UI, if

- there exists $Y \in L^1(\mathcal{A}, \mathbb{P})$ such that $|X| \le Y$ for all $X \in \mathcal{X}$
  [$\mathbb{E}\big(|X|\,\mathbb{1}_A\big) \le \mathbb{E}\big(Y\,\mathbb{1}_A\big)$ for all $A \in \mathcal{A}$, then use (ii)]
- there exists $p > 1$ such that $\mathcal{X}$ is uniformly bounded in L^p
  [by Hölder, for conjugate indices p and $q < \infty$, $\mathbb{E}\big(|X|\,\mathbb{1}_A\big) \le \|X\|_p\,\mathbb{P}(A)^{1/q}$]

Example. $X_n = n\,\mathbb{1}_{(0,1/n)}$ is uniformly bounded in $L^1\big((0,1], \mathcal{B}((0,1]), \mu\big)$ but not UI.

Proposition 4.15. *A family $\mathcal{X}$ of random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ is UI if and only if*

$$\sup\big\{\mathbb{E}\big(|X|\,\mathbb{1}_{|X|\ge K}\big) \,:\, X \in \mathcal{X}\big\} \to 0\,, \quad \textit{as } K \to \infty\,.$$

Proof. Suppose $\mathcal{X}$ is UI. Given $\epsilon > 0$, choose $\delta > 0$ so that $I_{\mathcal{X}}(\delta) < \epsilon$, then choose $K < \infty$ so that $I_{\mathcal{X}}(1) \le K\delta$. Then with $A = \big\{|X| \ge K\big\}$ we have $\|X\|_1 \ge \mathbb{P}(A)\,K$ so that $\mathbb{P}(A) \le \delta$ and $\mathbb{E}\big(|X|\,\mathbb{1}_A\big) < \epsilon$ for all $X \in \mathcal{X}$. Hence, as $K \to \infty$,

$$\sup\big\{\mathbb{E}\big(|X|\,\mathbb{1}_{|X|\ge K}\big) \,:\, X \in \mathcal{X}\big\} \to 0\,.$$

On the other hand, if this condition holds, $I_{\mathcal{X}}(1) < \infty$, since $\mathbb{E}\big(|X|\big) \le K + \mathbb{E}\big(|X|\,\mathbb{1}_{|X|\ge K}\big)$.
Given $\epsilon > 0$, choose $K < \infty$ so that $\mathbb{E}\big(|X|\,\mathbb{1}_{|X|\ge K}\big) < \epsilon/2$ for all $X \in \mathcal{X}$.
Then choose $\delta > 0$ so that $K\delta < \epsilon/2$. For all $X \in \mathcal{X}$ and $A \in \mathcal{A}$ with $\mathbb{P}(A) < \delta$, we have

$$\mathbb{E}\big(|X|\,\mathbb{1}_A\big) \le \mathbb{E}\big(|X|\,\mathbb{1}_{|X|\ge K}\big) + K\,\mathbb{P}(A) < \epsilon\,.$$

Hence $\mathcal{X}$ is UI. □

Theorem 4.16. *Let X_n, $n \in \mathbb{N}$ and X be random variables. The following are equivalent:*

(i) $X_n \in L^1$ for all $n \in \mathbb{N}$, $X \in L^1$ and $X_n \to X$ in L^1,

(ii) $\{X_n : n \in \mathbb{N}\}$ is UI and $X_n \to X$ in probability.

Proof. Suppose (i) holds. Then $X_n \to X$ in probability, following problem 3.4.
Moreover, given $\epsilon > 0$, there exists N such that $\mathbb{E}\big(|X_n - X|\big) < \epsilon/2$ whenever $n \ge N$. Then we can find $\delta > 0$ so that $\mathbb{P}(A) \le \delta$ implies, using Lemma 4.14,

$$\mathbb{E}\big(|X|\,\mathbb{1}_A\big) \le \epsilon/2\,, \quad \mathbb{E}\big(|X_n|\,\mathbb{1}_A\big) \le \epsilon, \quad \text{for all } n = 1, \ldots, N\,.$$

Then, also for $n \ge N$ and $\mathbb{P}(A) \le \delta$, $\mathbb{E}\big(|X_n|\,\mathbb{1}_A\big) \le \mathbb{E}\big(|X_n - X|\big) + \mathbb{E}\big(|X|\,\mathbb{1}_A\big) \le \epsilon$.
Hence $\{X_n : n \in \mathbb{N}\}$ is UI and we have shown that (i) implies (ii).
Now suppose that (ii) holds. Then there is a subsequence (n_k) such that $X_{n_k} \to X$ $a.s.$.
So, by Fatou's lemma, $\mathbb{E}\big(|X|\big) \le \liminf_k \mathbb{E}\big(|X_{n_k}|\big) < \infty$.
Now with Proposition 4.15, given $\epsilon > 0$, there exists $K < \infty$ such that, for all n,

$$\mathbb{E}\big(|X_n|\,\mathbb{1}_{|X_n|\ge K}\big) < \epsilon/3\,, \quad \mathbb{E}\big(|X|\,\mathbb{1}_{|X|\ge K}\big) < \epsilon/3\,.$$

Consider the unif. bounded sequence $X_n^K = (-K) \vee X_n \wedge K$ and set $X^K = (-K) \vee X \wedge K$. Then $X_n^K \to X^K$ in probability, so, by bounded convergence, there exists $N \in \mathbb{N}$ such that, for all $n \ge N$, $\mathbb{E}\big(|X_n^K - X^K|\big) < \epsilon/3$. But then, for all $n \ge N$,

$$\mathbb{E}\big(|X_n - X|\big) \le \mathbb{E}\big(|X_n|\,\mathbb{1}_{|X_n|\ge K}\big) + \mathbb{E}\big(|X_n^K - X^K|\big) + \mathbb{E}\big(|X|\,\mathbb{1}_{|X|\ge K}\big) < \epsilon\,.$$

Since $\epsilon > 0$ was arbitrary, we have shown that (ii) implies (i). □

5 Characteristic functions and Gaussian random variables

5.1 Definitions

Definition 5.1. For a finite measure μ on $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\big)$, define the *Fourier transform* $\hat{\mu} : \mathbb{R}^n \to \mathbb{C}$ by

$$\hat{\mu}(u) = \int_{\mathbb{R}^n} e^{i\langle u,x\rangle}\, \mu(dx)\ , \quad \text{for all } u \in \mathbb{R}^n\ .$$

Here $\langle u, x\rangle = \sum_{i=1}^{n} u_i\, x_i$ denotes the usual inner product on $\mathbb{R}^n$.
For a random variable X in $\mathbb{R}^n$, the *characteristic function* $\phi_X : \mathbb{R}^n \to \mathbb{C}$ is given by

$$\phi_X(u) = \mathbb{E}\big(e^{i\langle u,X\rangle}\big)\ , \quad \text{for all } u \in \mathbb{R}^n\ .$$

Thus $\phi_X = \hat{\mu}_X$ where μ_X is the distribution of X in $\mathbb{R}^n$.

Remark. For measurability of $f : \mathbb{R} \to \mathbb{C}$ identify $f = (\mathrm{Re} f, \mathrm{Im} f) \in \mathbb{R}^2$ and use $\mathcal{B}(\mathbb{R}^2)$. The integral of such functions is to be understood as

$$\int_{\mathbb{R}^n} f\, \mu(dx) = \int_{\mathbb{R}^n} \mathrm{Re} f\, \mu(dx) + i \int_{\mathbb{R}^n} \mathrm{Im} f\, \mu(dx)\ .$$

Since $e^{ix} = \cos x + i \sin x$ has bounded real and imaginary part it is integrable with respect to every finite measure. Thus also $\hat{\mu}(u)$ and $\phi_X(u)$ are well defined for all $u \in \mathbb{R}^n$ (in contrast to moment generating functions M_X, see problem 3.13).

Definition 5.2. A random variable X in $\mathbb{R}^n$ is called *standard Gaussian* if

$$\mathbb{P}(X \in A) = \int_A \frac{1}{(2\pi)^{n/2}}\, e^{-|x|^2/2}\, dx\ , \quad \text{for all } A \in \mathcal{B}(\mathbb{R}^n)\ .$$

Example. For a standard Gaussian random variable X in $\mathbb{R}$ it is

$$\phi_X(u) = \int_{\mathbb{R}} e^{iux}\, \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = e^{-u^2/2}\, I\ , \quad \text{where} \quad I = \int_{\mathbb{R}} \frac{e^{-(x-iu)^2/2}}{\sqrt{2\pi}}\, dx\ .$$

I can be evaluated by considering the complex integral $\int_\Gamma e^{-z^2/2}\, dz$ around the rectangular contour Γ with corners $R, R-iu, -R-iu, -R$. Since $e^{-z^2/2}$ is analytic, the integral vanishes by Cauchy's theorem for every $R > 0$. In the limit $R \to \infty$, the contributions from the vertical sides of Γ also vanish and thus

$$I = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = 1 \quad \Rightarrow \quad \phi_X(u) = e^{-u^2/2}\ .$$

In the next subsection we will also make use of the following.

Definition 5.3. For $t > 0$ and $x, y \in \mathbb{R}^n$ we define the *heat kernel*

$$p(t, x, y) = \frac{1}{(2\pi t)^{n/2}}\, e^{-|x-y|^2/(2t)} \in \mathbb{R}\ .$$

Remark. From the previous calculation we have $e^{-w^2/2} = \int_{\mathbb{R}} e^{iwu} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du.$ With $w = (x-y)/\sqrt{t}$ and the change of variable $v = u/\sqrt{t}$, we deduce for $n = 1$

$$p(t,x,y) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{ixv}\, e^{-v^2 t/2}\, e^{-iyv}\, dv \ .$$

For $n \geq 1$ we obtain analogously $p(t,x,y) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\langle x,v\rangle}\, e^{-|v|^2 t/2}\, e^{-i\langle y,v\rangle}\, \mu(dv)$.

5.2 Properties of characteristic functions

The characteristic function of a random variable uniquely determines its distribution.

Theorem 5.1. Uniqueness and inversion
Let X be a random variable in $\mathbb{R}^n$. The law μ_X of X is uniquely determined by its characteristic function ϕ_X. Moreover, if ϕ_X is integrable, then X has density function $f_X(x)$, with

$$f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u)\, e^{-i\langle u,x\rangle}\, du \ .$$

Remark. The above formula is also called the *inverse Fourier transformation.*

Proof. Let Y be a standard Gaussian r.v. in $\mathbb{R}^n$, independent of X, and let $g : \mathbb{R}^n \to \mathbb{R}$ be a bounded Borel function. Then, for $t > 0$, by change of variable $y' = x + \sqrt{t}y$ and Fubini,

$$\begin{aligned}
\mathbb{E}(g(X+\sqrt{t}Y)) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} g(x+\sqrt{t}y)(2\pi)^{-n/2}\, e^{-|y|^2/2}\, dy\, \mu_X(dx) = \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} p(t,x,y')\, g(y')\, dy'\, \mu_X(dx) = \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{i\langle u,x\rangle}\, e^{-|u|^2 t/2}\, e^{-i\langle u,y\rangle}\, du\, \mu_X(dx) \right) g(y)\, dy \\
&= \int_{\mathbb{R}^n} \left(\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u)\, e^{-|u|^2 t/2}\, e^{-i\langle u,y\rangle}\, du \right) g(y)\, dy \ .
\end{aligned}$$

By this formula, ϕ_X determines $\mathbb{E}\big(f(X+\sqrt{t}Y)\big)$. For any bounded continuous g, we have

$$\mathbb{E}\big(g(X+\sqrt{t}Y)\big) \to \mathbb{E}\big(g(X)\big) \quad \text{as } t \searrow 0 \ ,$$

so ϕ_X determines $\mathbb{E}\big(g(X)\big)$. Hence ϕ_X determines μ_X due to problem 4.1.
If ϕ_X is integrable and if g is continuous and bounded, then

$$(u,y) \mapsto \big|\phi_X(u)\big|\,\big|g(y)\big| \in L^1(du \otimes dy) \ .$$

So, by dominated convergence, as $t \searrow 0$, the last integral above converges to

$$\int_{\mathbb{R}^n} \left(\frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \phi_X(u)\, e^{-i\langle u,y\rangle}\, du \right) g(y)\, dy \ .$$

Hence X has the claimed density function. □

Remark. Let X, Y be independent r.v.s in $\mathbb{R}^n$. Then the characteristic fct. of the sum is

$$\phi_{X+Y}(u) = \mathbb{E}\big(e^{i\langle u, X+Y\rangle}\big) = \mathbb{E}\big(e^{i\langle u,X\rangle}\, e^{i\langle u,Y\rangle}\big) = \phi_X(u)\,\phi_Y(u)\ .$$

The next result shows that independence of r.v.s is equivalent to factorisation of the joint characteristic function.

Theorem 5.2. *Let $X = (X_1, \ldots, X_n)$ be a r.v. in $\mathbb{R}^n$. Then the following are equivalent:*

(i) $X_1, \ldots, X_n$ *are independent ,*

(ii) $\mu_X = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$,

(iii) $\mathbb{E}\Big(\prod_{k=1}^n f_k(X_k)\Big) = \prod_{k=1}^n \mathbb{E}\big(f_k(X_k)\big)$, *for all bounded Borel functions* $f_1, \ldots, f_n$,

(iv) $\phi_X(u) = \prod_{k=1}^n \phi_{X_k}(u_k)$, *for all* $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$.

Proof. If (i) holds, $\mu_X(A_1 \times \ldots \times A_n) = \prod_k \mu_{X_k}(A_k)$ for all Borel sets $A_1, \ldots, A_n$.
So (ii) holds, since this formula characterizes the product measure by Theorem 3.15.
If (ii) holds, then, for $f_1, \ldots, f_n$ bounded Borel,

$$\mathbb{E}\Big(\prod_k f_k(X_k)\Big) = \int_{\mathbb{R}^n} \prod_k f_k(x_k)\, \mu_X(dx) = \prod_k \int_{\mathbb{R}} f_k(x_k)\, \mu_{X_k}(dx_k) = \prod_k \mathbb{E}\big(f_k(X_k)\big)\ ,$$

so (iii) holds. Statement (iv) is a special case of (iii), with $f_k(x_k) = e^{i\, u_k x_k}$.
Suppose, finally, that (iv) holds and take independent r.v.s $\tilde{X}_1, \ldots, \tilde{X}_n$ with $\mu_{\tilde{X}_k} = \mu_{X_k}$ for all k. Then $\phi_{\tilde{X}_k} = \phi_{X_k}$, and we know that (i) implies (iv) for $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_n)$, so

$$\phi_{\tilde{X}}(u) = \prod_k \phi_{\tilde{X}_k}(u_k) = \prod_k \phi_{X_k}(u_k) = \phi_X(u)\ ,$$

and $\mu_{\tilde{X}} = \mu_X$ by uniqueness of characteristic functions. Hence (i) holds. □

5.3 Gaussian random variables

Definition 5.4. A random variable X in $\mathbb{R}$ is *Gaussian* if it has density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)}\ ,$$

for some $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$. We write $X \sim \mathcal{N}(\mu, \sigma^2)$.
We also admit as Gaussian the degenerate case $X = \mu\ \ a.s.$, corresponding to taking $\sigma^2 = 0$.

Proposition 5.3. *Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Then*

(i) $\mathbb{E}(X) = \mu$,

(ii) $\mathrm{var}(X) = \sigma^2$,

(iii) $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$,

(iv) $\phi_X(u) = e^{iu\mu - u^2\sigma^2/2}$.

Proof. see problem 4.7

Definition 5.5. A random variable X in $\mathbb{R}^n$ is *Gaussian* if $\langle u, X\rangle$ is Gaussian, for all $u \in \mathbb{R}^n$.

Examples. (i) If $X = (X_1, \dots, X_n)$ is Gaussian, in particular X_i is Gaussian for each i.

(ii) Let $X_1, \dots, X_n$ be independent $\mathcal{N}(0,1)$ random variables. Then $X = (X_1, \dots, X_n)$ is Gaussian, since for all $u \in \mathbb{R}^n$

$$\mathbb{E}\big(e^{iv\langle u,X\rangle}\big) = \mathbb{E}\Big(\prod_{k=1}^n e^{ivu_k X_k}\Big) = e^{-v^2|u|^2/2}\ , \quad \text{for all } v \in \mathbb{R}\ .$$

Thus $\langle u, X\rangle \sim \mathcal{N}(0, |u|^2)$ by uniqueness of characteristic functions.

Remark. Let X be a random variable in $\mathbb{R}^n$. Then the covariance matrix

$$\Sigma = \mathrm{var}(X) = \big(\mathrm{cov}(X_i, X_j)\big)_{i,j=1,..,n} = \mathbb{E}\big((X - \mathbb{E}(X))(X - \mathbb{E}(X))^T\big)$$

is symmetric and non-negative definite by Proposition 4.11. Thus Σ has n real eigenvalues $\lambda_i \geq 0$ and the eigenvectors v_i form an ortho-normal basis of $\mathbb{R}^n$, i.e. $\langle v_i, v_j\rangle = \delta_{i,j}$. So

$$x = \langle v_i, x\rangle v_i = (v_i^T x) v_i \quad \text{and} \quad \Sigma x = \sum_{i=1}^n \lambda_i v_i (v_i^T x) \quad \text{for all } x \in \mathbb{R}^n\ .$$

So we can write $\Sigma = \sum_{i=1}^n \lambda_i\, v_i\, v_i^T$, and we define $\Sigma^{1/2} := \sum_{i=1}^n \sqrt{\lambda_i}\, v_i\, v_i^T$.
$P_i = v_i v_i^T \in \mathbb{R}^{n\times n}$ is the projection on the one-dimensional eigenspace corresponding to the eigenvector λ_i. Since $P_i P_j = \delta_{i,j} P_i$, it follows that $\Sigma = \big(\Sigma^{1/2}\big)^2$.

Theorem 5.4. *Let $X : \Omega \to \mathbb{R}^n$ be a Gaussian random variable. Then*

(i) *$AX + b$ is Gaussian, for all $A \in \mathbb{R}^{n\times n}$ and all $b \in \mathbb{R}^n$,*

(ii) *$X \in L^2(\Omega)$ (coordinatewise) and its distribution is determined by the mean $\mu = \mathbb{E}(X) \in \mathbb{R}^n$ and the covariance matrix $\Sigma = \mathrm{var}(X) \in \mathbb{R}^{n\times n}$, we write $X \sim \mathcal{N}(\mu, \Sigma)$,*

(iii) $\phi_X(u) = e^{i\langle u,\mu\rangle - \langle u,\Sigma u\rangle/2}$,

(iv) *if Σ is invertible, then X has a density function on $\mathbb{R}^n$, given by*

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}}\ \exp\Big[-\big\langle x - \mu, \Sigma^{-1}(x - \mu)\big\rangle/2\Big]\ ,$$

(v) *if $X = (Y, Z)$, with Y in $\mathbb{R}^m$ and Z in $\mathbb{R}^p$ ($m{+}p{=}n$), then the block structure*

$$\mathrm{var}(X) = \begin{pmatrix} \mathrm{var}(Y) & 0 \\ 0 & \mathrm{var}(Z)\end{pmatrix} \quad \textit{implies that} \quad Y \textit{ and } Z \textit{ are independent}\ .$$

Proof. We use $\langle u,v\rangle = u^T v$ and $(Av)^T = v^T A^T$ for all $u,v\in\mathbb{R}^n$ and $A\in\mathbb{R}^{n\times n}$.
(i) For all $u\in\mathbb{R}^n$, $\langle u, AX+b\rangle = \langle A^T u, X\rangle + \langle u,b\rangle$ is Gaussian, by Proposition 5.3.
(ii),(iii) Each X_k is Gaussian, so $X\in L^2$. For all $u\in\mathbb{R}^n$ we have $\mathbb{E}\big(\langle u,X\rangle\big) = \langle u,\mu\rangle$ and

$$\mathrm{var}\big(\langle u,X\rangle\big) = \mathbb{E}\Big(u^T(X-\mu)(X-\mu)^T u\Big) = \Big\langle u, \mathbb{E}\big((X-\mu)(X-\mu)^T\big)u\Big\rangle = \langle u,\Sigma u\rangle\ .$$

Since $\langle u,X\rangle$ is Gaussian, by Proposition 5.3, we must have $\langle u,X\rangle \sim \mathcal{N}\big(\langle u,\mu\rangle, \langle u,\Sigma u\rangle\big)$ and

$$\phi_X(u) = \mathbb{E}\big(e^{i\langle u,X\rangle}\big) = \phi_{\langle u,X\rangle}(1) = e^{i\langle u,\mu\rangle - \langle u,\Sigma u\rangle/2}\ .$$

This is (iii) and (ii) follows by uniqueness of characteristic functions.
(iv) Let $Y_1,\ldots,Y_n$ be independent $\mathcal{N}(0,1)$ r.v.s. Then $Y=(Y_1,\ldots,Y_n)$ has density

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^n}}\, e^{-|y|^2/2}\ .$$

Set $\tilde X = \Sigma^{1/2}Y + \mu$, then $\tilde X$ is Gaussian, with $\mathbb{E}(\tilde X)=\mu$ and $\mathrm{var}(\tilde X)=\Sigma$, since

$$\mathrm{cov}(\tilde X_i,\tilde X_j) = \mathbb{E}\big((\Sigma^{1/2}Y)_i(\Sigma^{1/2}Y)_j\big) = \mathbb{E}\Big(\sum_{k,l=1}^n \Sigma^{1/2}_{ik} Y_k \Sigma^{1/2}_{jl} Y_l\Big) = \Sigma_{ij}$$

due to $\mathbb{E}(Y_kY_l)=\delta_{k,l}$. So $\tilde X\sim X$. If Σ is invertible, then $\tilde X$ and hence X has the density claimed in (iv), by the linear change of variables $Y=\Sigma^{-1/2}(X-\mu)$ leading to

$$|y|^2 = \langle y,y\rangle = \big\langle x-\mu, \Sigma^{-1}(x-\mu)\big\rangle \quad\text{and}\quad d^n y = d^n x \det\Sigma^{-1/2} = \frac{d^n x}{\sqrt{\det\Sigma}}\ .$$

(v) Finally, if $X=(Y,Z)$ and $\Sigma=\mathrm{var}(X)$ has the block structure given in (v) then, for all $v\in\mathbb{R}^m$ and $w\in\mathbb{R}^p$,

$$\big\langle (v,w), \Sigma\,(v,w)\big\rangle = \langle v, \Sigma_Y\, v\rangle + \langle w,\Sigma_Z\, w\rangle\ ,\quad \text{where } \Sigma_Y=\mathrm{var}(Y) \text{ and } \Sigma_Z = \mathrm{var}(Z)\ .$$

With $\mu=(\mu_Y,\mu_Z)$, the joint characteristic function ϕ_X then splits into a product

$$\phi_X(v,w) = e^{i\langle v,\mu_Y\rangle - \langle v,\Sigma_Y v\rangle/2}\, e^{i\langle w,\mu_Z\rangle - \langle w,\Sigma_Z w\rangle/2}\ ,$$

so Y and Z are independent by Theorem 5.2. □

Remarks. Let $X=(X_1,\ldots,X_n)\sim\mathcal{N}(\mu,\Sigma)$ be a Gaussian.

(i) $X_1,\ldots,X_n$ are independent if and only if Σ is a diagonal matrix.

(ii) If Σ is invertible, $Y=\Sigma^{-1/2}(X-\mu)\sim\mathcal{N}(0,I_n)$ are independent $\mathcal{N}(0,1)$.

6 Ergodic theory and sums of random variables

6.1 Motivation

Theorem 6.1. Strong law of large numbers
Let $(X_n)_{n\in\mathbb{N}}$ *be a sequence of independent random variables such that, for some constants* $\mu \in \mathbb{R}$, $M > 0$,

$$\mathbb{E}(X_n) = \mu\ , \quad \mathbb{E}(X_n^4) \leq M\ , \quad \textit{for all } n \in \mathbb{N}\ .$$

Then, with $S_n = \sum_{i=1}^{n} X_i$ *we have* $S_n/n \to \mu$ *a.s. as* $n \to \infty$.

Proof. For $Y_n := X_n - \mu$ we have

$$Y_n^4 \leq \big(|X_n| + |\mu|\big)^4 \leq \big(2\max\{|X_n|, |\mu|\}\big)^4 \leq 16\big(|X_n|^4 + |\mu|^4\big)\ ,$$

and thus $\mathbb{E}(Y_n^4) \leq 16(M + \mu^4) = \tilde{M} < \infty$ for all $n \in \mathbb{N}$. With Y_n^4 also Y_n, Y_n^2 and Y_n^3 are integrable and by independence and $\mathbb{E}(Y) = 0$

$$\mathbb{E}\big(Y_i\, Y_j^3\big) = \mathbb{E}\big(Y_i\, Y_j\, Y_k^2\big) = \mathbb{E}\big(Y_i\, Y_j\, Y_k\, Y_l\big) = 0\ ,$$

for distinct indices i, j, k, l. Hence

$$\mathbb{E}\big((S_n - n\mu)^4\big) = \mathbb{E}\bigg(\sum_{i,j,k,l} Y_i\, Y_j\, Y_k\, Y_l\bigg) = \mathbb{E}\bigg(\sum_i Y_i^4 + \binom{4}{2}\sum_{i<j} Y_i^2\, Y_j^2\bigg)\ ,$$

and by Jensen's inequality $\tilde{M} \geq \mathbb{E}\big((Y_i^2)^2\big) \geq \mathbb{E}\big(Y_i^2\big)^2$, so using independence

$$\mathbb{E}\big((S_n - n\mu)^4\big) \leq n\tilde{M} + 6\,\frac{n(n-1)}{2}\,\tilde{M} \leq 3n^2\tilde{M}\ .$$

Thus $\mathbb{E}\bigg(\sum_n (S_n/n - \mu)^4\bigg) \leq 3\tilde{M}\sum_n 1/n^2 < \infty$ by monotone convergence. Therefore $\sum_n (S_n/n - \mu)^4 < \infty$ *a.s.* and thus $S_n/n \to \mu$ *a.s.* . □

Intuitively, the above result should also hold without the restrictive assumption on the fourth moment of the random variables. One goal of this chapter is in fact to prove the above statement with a much weaker assumption. For this purpose it is convenient to use a different approach, leading to ergodic theory which is introduced in the next two sections.

6.2 Measure-preserving transformations

Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be a σ-finite measure space.

Definition 6.1. Let $\theta : E \to E$ be measurable. $A \in \mathcal{E}$ is called *invariant* (under θ) if $\theta^{-1}(A) = A$. A measurable function $f : E \to F$ is called *invariant* (under θ) if $f = f \circ \theta$.

Remarks. (i) For general $A \subseteq E$ it is $\theta\big(\theta^{-1}(A)\big) \subseteq A \subseteq \theta^{-1}\big(\theta(A)\big)$ (where the first and second inclusion become equalities, if θ is surjective and injective, respectively) .
If A is invariant, $\theta(A) \subseteq A$ (motivating the definition), and also $A = \theta^{-1}\big(\theta(A)\big)$, since in addition to the general relation we have $\theta^{-1}\big(\theta(A)\big) \subseteq \theta^{-1}(A) = A$.

(ii) $\mathcal{E}_\theta := \{A \in \mathcal{E} : \theta^{-1}(A) = A\}$ is a σ-algebra since pre-images preserve set operations.

(iii) $A \in \mathcal{E}$ is invariant $\Leftrightarrow \mathbb{1}_A = \mathbb{1}_A \circ \theta$, since $\mathbb{1}_A \circ \theta = \mathbb{1}_{\theta^{-1}(A)}$.

(iv) $f : E \to F$ is invariant $\Leftrightarrow \forall B \in \mathcal{F} : f^{-1}(B) = \theta^{-1}\big(f^{-1}(B)\big)$
$\Leftrightarrow \forall B \in \mathcal{F} : f^{-1}(B) \in \mathcal{E}_\theta$, i.e. f is $\mathcal{E}_\theta$-measurable .

Definition 6.2. A measurable function $\theta : E \to E$ is called *measure-preserving* if

$$\mu\big(\theta^{-1}(A)\big) = \mu(A) \quad \text{for all } A \in \mathcal{E} .$$

Such θ is *ergodic* if $\mathcal{E}_\theta$ is *trivial*, i.e. contains only sets of measure 0 and their complements.

Examples. (i) The constant function $\theta(x) = c \in E$ is not measure preserving.
The identity $\theta(x) = x$ is measure preserving, but not ergodic, since $\mathcal{E}_\theta = \mathcal{E}$.

(i) *Translation map on the torus.* Take $E = [0, 1)^n$ with Lebesgue measure, for $a \in E$ set

$$\theta_a(x_1, \ldots, x_n) = (x_1 + a, \ldots, x_n + a) \quad \text{with addition modulo 1 .}$$

In problem 4.10 it is shown for $n = 1$ that θ is measure-preserving, and also ergodic if and only if a is irrational.

(ii) *Baker's map.* Take $E = (0, 1]$ with Lebesgue measure and set $\theta(x) = 2x - \lfloor 2x \rfloor$.
In problem 4.11 it is shown that θ is measure-preserving and ergodic.

Proposition 6.2. *If $f : E \to \mathbb{R}$ is integrable and $\theta : E \to E$ is measure-preserving, then $f \circ \theta$ is integrable and* $\displaystyle\int_E f\, d\mu = \int_E f \circ \theta\, d\mu$.

Proof. For $f = \mathbb{1}_A$, $A \in \mathcal{E}$ the statement reduces to $\mu(A) = \mu\big(\theta^{-1}(A)\big)$, which holds since μ is measure-preserving. This extends to simple functions by linearity, to non-negative measurable functions by monotone convergence and to integrable $f = f^+ - f^-$ again by linearity.□

Proposition 6.3. *If $\theta : E \to E$ is ergodic and $f : E \to \mathbb{R}$ is invariant, then $f = c$ $a.e.$ for some constant $c \in \mathbb{R}$.*

Proof. For all $A \in \mathcal{B}$, $\mu(f \in A) = 0$ or $\mu(f \in A^c) = 0$, since f is $\mathcal{E}_\theta$-measurable and θ is ergodic. Set $c := \inf\{a \in \mathbb{R} : \mu(f > a) = 0\}$. So $\mu(f \le a) = 0$ for all $a < c$ and $\mu(f \ge a) = 0$ for all $a > c$, and thus $f = c$ $a.e.$. □

Interpretation. $\theta : E \to E$ defines a dynamical system $x_n = x_n(x_0) = \theta^n(x_0) \in E$ with discrete time $n \in \mathbb{N}$ and initial condition $x_0 \in E$. The dynamics is defined on the abstract state space $(E, \mathcal{E}, \mu)$ and the *observables* are given by measurable functions $f : E \to \mathbb{R}$.
If μ is measure-preserving, then $\int_E f(x_n)\, \mu(dx_0) = \int_E f(x_0)\, \mu(dx_0)$ for all f by Proposition 6.2 and μ can be interpreted as a *stationary distribution* for the process $(x_n)_n$. If f is invariant then $f(x_n) = f(x_0)$ for all n and f is a conserved quantity, such as energy in a physical system. If there exists such a non-constant f, the state space can be partitioned in subsets $f^{-1}(y) \subseteq E$

for all $y \in f(E)$, which are non-communicating under the time evolution defined by θ. However if θ is ergodic, Proposition 6.3 implies that the only invariant functions are constant a.e.. So an ergodic dynamical system does not have conserved quantities which partition the state space into non-communicating classes of non-zero measure (compare to Markov chains).

For the rest of this section we consider the infinite product space

$$E = \mathbb{R}^{\mathbb{N}} = \big\{ x = (x_n)_{n\in\mathbb{N}} \,:\, x_n \in \mathbb{R},\, n \in \mathbb{N} \big\}$$

with σ-algebra $\mathcal{E} = \sigma(X_n \,:\, n \in \mathbb{N})$ generated by the *coordinate maps* $X_n : E \to \mathbb{R}$ with $X_n(x) = x_n$.

Remark. $\mathcal{E} = \sigma(\mathcal{C})$ generated by the π-system

$$\mathcal{C} = \Big\{ \bigotimes_{n\in\mathbb{N}} A_n \,:\, A_n \in \mathcal{B},\, A_n = \mathbb{R} \text{ for all but finitely many } n \Big\} ,$$

which consists of socalled *cylinder sets*, where only finitely many coordinates are specified.

Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of iidrv's with distribution m. With the Skorohod theorem they can be constructed on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$. $Y : \Omega \to \mathbb{R}^{\mathbb{N}}$ defined as $Y(\omega) = \big(Y_n(\omega)\big)_{n\in\mathbb{N}}$ is $\mathcal{A}/\mathcal{E}$-measurable and the distribution $\mu = \mathbb{P} \circ Y^{-1}$ of Y satisfies

$$\mu(A) = \prod_{n\in\mathbb{N}} m(A_n) \quad \text{for all cylinder sets } A = \bigotimes_{n\in\mathbb{N}} A_n \in \mathcal{C} .$$

Since $\mathcal{C}$ is a π-system generating $\mathcal{E}$, this is the unique measure on $(E, \mathcal{E})$ with this property. Therefore $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}^{\mathbb{N}}, \mathcal{E}, \mu)$ is a generic example of such a common probability space.

Definition 6.3. On the probability space $(\mathbb{R}^{\mathbb{N}}, \mathcal{E}, \mu)$ the coordinate maps $X_n : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ themselves are iidrv's with distribution m, and this is called the *canonical model* for such a sequence. The *shift map* $\theta : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is defined as $\theta(x_1, x_2, \ldots) = (x_2, x_3, \ldots)$.

Theorem 6.4. *The shift map θ is ergodic.*

Proof. θ is measurable and measure-preserving (see problem 4.9).
To see that θ is ergodic recall the tail σ-algebra

$$\mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n \quad \text{where} \quad \mathcal{T}_n = \sigma(X_m \,:\, m > n) \subseteq \mathcal{E} .$$

For $A = \bigotimes_{k\in\mathbb{N}} A_k \in \mathcal{C}$, $\theta^{-n}(A) = \big\{ x \in E \,:\, X_{n+k}(x) \in A_k \text{ for all } k \geq 1 \big\} \in \mathcal{T}_n$.
Since $\mathcal{T}_n$ is a σ-algebra, $\theta^{-n}(A) \in \mathcal{T}_n$ for all $A \in \mathcal{E}$. If $A \in \mathcal{E}_\theta = \big\{ B \in \mathcal{E} \,:\, \theta^{-1}(B) = B \big\}$ then $A = \theta^{-n}(A) \in \mathcal{T}_n$ for all $n \in \mathbb{N}$ and thus $A \in \bigcap_n \mathcal{T}_n = \mathcal{T}$ so that $\mathcal{E}_\theta \subseteq \mathcal{T}$. By Kolmogorov's 0-1-law $\mathcal{T}$ and thus $\mathcal{E}_\theta$ is trivial and θ is ergodic. □

6.3 Ergodic Theorems

In the following let $(E, \mathcal{E}, \mu)$ be a σ-finite measure space with a measure-preserving transformation $\theta : E \to E$. Let $f : E \to \mathbb{R}$ be integrable and define $S_n : E \to \mathbb{R}$ by $S_0 = 0$ and

$$S_n = S_n(f) = f + f \circ \theta + \ldots + f \circ \theta^{n-1} \quad \text{for } n \geq 1 .$$

Example. Let $(\mathbb{R}^{\mathbb{N}}, \mathcal{E}, \mu)$ be the canonical model for iidrv's $(X_n)_{n\in\mathbb{N}}$, $f = X_1 : E \to \mathbb{R}$ the first coordinate map and θ the shift map from the previous section. Then $S_n(X_1) = \sum_{i=1}^{n} X_i$.

Lemma 6.5. Maximal ergodic lemma
Let $S^* = \sup_{n\in\mathbb{N}} S_n : E \to \mathbb{R}$. *Then* $\displaystyle\int_{S^*>0} f \, d\mu \geq 0$.

Proof. Set $S_n^* = \max_{0\leq m\leq n} S_m$ and $A_n = \{S_n^* > 0\}$. Then, for $m = 1, \ldots, n$,

$$S_m = f + S_{m-1} \circ \theta \leq f + S_n^* \circ \theta .$$

On A_n we have $S_n^* = \max_{1\leq m\leq n} S_m$, so $S_n^* \leq f + S_n^* \circ \theta$.
On A_n^c we have $S_n^* = 0 \leq S_n^* \circ \theta$. So, integrating and adding, we obtain

$$\int_E S_n^* \, d\mu \leq \int_{A_n} f \, d\mu + \int_E S_n^* \circ \theta \, d\mu .$$

But S_n^* is integrable and θ is measure-preserving, so

$$\int_E S_n^* \circ \theta \, d\mu = \int_E S_n^* \, d\mu < \infty \quad \text{which implies} \quad \int_{A_n} f \, d\mu \geq 0 .$$

As $n \to \infty$, $A_n \nearrow \{S^* > 0\}$ so, by monotone convergence, $\displaystyle\int_{\{S^*>0\}} f \, d\mu \geq 0$. □

Theorem 6.6. Birkhoff's almost everywhere ergodic theorem
There exists $\bar{f} : E \to \overline{\mathbb{R}}$ *invariant, with* $\displaystyle\int_E |\bar{f}| \, d\mu \leq \int_E |f| \, d\mu$ *and* $\displaystyle\frac{S_n}{n} \to \bar{f}$ $a.e.$ *as* $n \to \infty$.

Proof. The functions $\liminf_n (S_n/n)$ and $\limsup_n (S_n/n)$ are invariant, since

$$\Big(\liminf \frac{S_n}{n}\Big) \circ \theta = \liminf \Big(\frac{S_n \circ \theta}{n}\Big) = \liminf \Big(\frac{S_{n+1} - f}{n}\Big) = \liminf \Big(\frac{S_{n+1}}{n+1}\Big) .$$

Therefore, for $a < b$,

$$D = D(a, b) = \big\{ \liminf_n (S_n/n) < a < b < \limsup_n (S_n/n) \big\} .$$

is an invariant event. We shall show that $\mu(D) = 0$. First, by invariance, we can restrict everything to D and thereby reduce to the case $D = E$. Note that either $b > 0$ or $a < 0$. We can interchange the two cases by replacing f by $-f$. Let us assume then that $b > 0$.
Let $B \in \mathcal{E}$ with $\mu(B) < \infty$, then $g = f - b \, \mathbb{1}_B$ is integrable and, for each $x \in D$, for some n,

$$S_n(g)(x) \geq S_n(f)(x) - n \, b > 0 .$$

Hence $S^*(g) > 0$ everywhere and, by the maximal ergodic lemma,

$$0 \le \int_D (f - b\,\mathbb{1}_B)\,d\mu = \int_D f\,d\mu - b\,\mu(B)\ .$$

Since μ is σ-finite, we can let $B \nearrow D$ to obtain $\quad b\,\mu(D) \le \int_D f\,d\mu\ .$
In particular, we see that $\mu(D) < \infty$. A similar argument applied to $-f$ and $-a$, this time with $B = D$, shows that $\quad (-a)\mu(D) \le \int_D (-f)\,d\mu\ .\quad$ Hence $\quad b\,\mu(D) \le \int_D f\,d\mu \le a\,\mu(D)\ .$
Since $a < b$ and the integral is finite, this forces $\mu(D) = 0$.
Back to general E. Set

$$\Delta = \left\{ \liminf_n (S_n/n) < \limsup_n (S_n/n) \right\}\ ,$$

then Δ is invariant. Also, $\Delta = \bigcup_{a,b\in\mathbb{Q},a<b} D(a,b)$, so $\mu(\Delta) = 0$. On the complement of Δ, S_n/n converges in $[-\infty,\infty]$, so we can define an invariant function $\bar{f} : E \to \overline{\mathbb{R}}$ by

$$\bar{f} = \begin{cases} \lim_n (S_n/n) & \text{on } \Delta^c \\ 0 & \text{on } \Delta \end{cases}\ .$$

Finally, we have $\int_E |f \circ \theta^n|\,d\mu = \int_E |f|\,d\mu$, so $\int_E |S_n|\,d\mu \le n \int_E |f|\,d\mu$ for all n. Hence, by Fatou's lemma,

$$\int_E |\bar{f}|\,d\mu = \int_E \liminf_n |S_n/n|\,d\mu \le \liminf_n \int_E |S_n/n|\,d\mu \le \int_E |f|\,d\mu\ . \qquad \square$$

Theorem 6.7. von Neumann's L^p ergodic theorem
Assume that $\mu(E) < \infty$ and $p \in [1,\infty)$ and let $\bar{f}$ be the invariant limit function of Theorem 6.6. Then, for $f \in L^p$, $S_n/n \to \bar{f}$ in L^p.

Proof. Since θ is measure-preserving we have

$$\|f\circ\theta^n\|_p = \left(\int_E |f|^p \circ \theta^n\,d\mu\right)^{1/p} = \|f\circ\theta^{n-1}\|_p = \ldots = \|f\|_p\ .$$

So, by Minkowski's inequality, $\qquad \|S_n(f)/n\|_p \le \|f\|_p\ .$
Given $\epsilon > 0$, choose $K < \infty$ so that $\|f - g\|_p < \epsilon/3$, where $g = (-K) \vee f \wedge K$.
By Birkhoff's theorem, $S_n(g)/n \to \bar{g}$ *a.e.* . We have $|S_n(g)/n| \le K$ for all n so, by bounded convergence ($\mu(E) < \infty$), there exists N such that, for $n \ge N$,

$$\|S_n(g)/n - \bar{g}\|_p < \epsilon/3\ .$$

By Fatou's lemma,

$$\|\bar{f} - \bar{g}\|_p^p = \int_E \liminf_n \left|\frac{S_n(f-g)}{n}\right|^p d\mu \le \liminf_n \int_E \left|\frac{S_n(f-g)}{n}\right|^p d\mu \le \|f-g\|_p^p\ .$$

Hence, for $n \ge N$,

$$\left\|\frac{S_n(f)}{n} - \bar{f}\right\|_p \le \left\|\frac{S_n(f-g)}{n}\right\|_p + \left\|\frac{S_n(g)}{n} - \bar{g}\right\|_p + \|\bar{g} - \bar{f}\|_p < 3\epsilon/3 = \epsilon\ . \qquad \square$$

Corollary 6.8. *Let $\mu(E) < \infty$, $f \in L^1$ and $\bar f$ be the invariant limit function of Theorem 6.6. Then* $\displaystyle\int_E \bar f\, d\mu = \int_E f\, d\mu$ *and if θ is ergodic,* $\displaystyle\bar f = \int f\, d\mu/\mu(E)$ *a.e.*.

Proof. $\displaystyle\Big|\int_E \frac{S_n}{n}\, d\mu - \int_E \bar f\, d\mu\Big| \leq \big\|S_n/n - \bar f\big\|_1 \to 0$ by Theorem 6.7. By the definition of S_n, $\displaystyle\int_E \frac{S_n}{n}\, d\mu = \int_E f\, d\mu$ for all $n \in \mathbb{N}$ since θ is measure preserving and the first statement follows.
If θ is ergodic, the invariant function $\bar f$ is constant a.e. by Proposition 6.3, and together with the first this implies the second statement. □

6.4 Limit theorems for sums of random variables

Theorem 6.9. Strong law of large numbers
Let $Y_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$ be iidrv's with $\mathbb{E}(Y_n) = \nu \in \mathbb{R}$ and $\mathbb{E}\big(|Y_n|\big) < \infty$, i.e. $Y_n \in L^1$. For $S_n = Y_1 + \ldots + Y_n$ we have

$$S_n/n \to \nu \ \ a.s.\,, \quad \text{as } n \to \infty\,.$$

Proof. Let $(\mathbb{R}^{\mathbb{N}}, \mathcal{E}, \mu)$ be the canonical model for the sequence $Y := (Y_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ with distribution μ as in Def. 6.3. Take $f = Y_1 \in L^1$ to be the first coordinate map. Note that

$$S_n = Y_1 + \ldots + Y_n = f + f \circ \theta + \ldots + f \circ \theta^{n-1}\,,$$

where $\theta : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is the shift map which is measure preserving and ergodic by Theorem 6.4. With $\mu(\mathbb{R}^{\mathbb{N}}) = 1$ we have by Theorem 6.6 and Corollary 6.8

$$S_n/n \stackrel{a.s.}{\longrightarrow} \mathbb{E}(f) = \mathbb{E}(Y_1) = \nu\,. \qquad \square$$

Remarks. (i) By Theorem 6.7 we also have convergence in L^1 in Theorem 6.9.

(ii) Theorem 6.9 is stronger than Theorem 6.1 where we needed $\mathbb{E}(Y_n^4) \leq M$ for all $n \in \mathbb{N}$. But here the Y_n have to be identically distributed for θ to be measure preserving.

(iii) With Thm 2.10 and Prop 2.11 the strong implies the **weak law of large numbers**:
$S_n/n \to \nu\ \ a.s. \quad\Rightarrow\quad S_n/n \to \nu$ in probability $\quad\Leftrightarrow\quad S_n/n \Rightarrow \nu \quad$ as $n \to \infty$.

Theorem 6.10. Lévy's convergence theorem for characteristic functions
Let X_n, $n \in \mathbb{N}$ and X be random variables in $\mathbb{R}$ with characteristic functions $\phi_X(u) = \mathbb{E}\big(e^{iuX}\big)$ and $\phi_{X_n}(u)$. Then

$$\phi_{X_n}(u) \to \phi_X(u) \ \text{ for all } u \in \mathbb{R} \quad\Leftrightarrow\quad X_n \to X \ \text{ in distribution}\,.$$

Proof. '$\Leftarrow$': By Theorem 3.9, $X_n \stackrel{D}{\longrightarrow} X \quad\Leftrightarrow\quad \mathbb{E}\big(f(X_n)\big) \to \mathbb{E}\big(f(X)\big)$ for all $f \in C_b(\mathbb{R})$ and $e^{iux} = \cos(ux) + i\,\sin(ux)$ is bounded and continuous for all $u \in \mathbb{R}$.
'$\Rightarrow$' is more involved, see e.g. Billingsley, Probability and Measure (3rd ed.), Thm 26.3. □

Theorem 6.11. Central limit theorem
Let $(X_n)_{n\in\mathbb{N}}$ *be a sequence of iidrv's with mean* 0 *and variance* 1*. Set* $S_n = X_1 + \ldots + X_n$*. Then* $\mathbb{P}(S_n/\sqrt{n} \le .) \Rightarrow \mathcal{N}(0,1)$ *, i.e. for all* $a < b$*,*

$$\mathbb{P}\big(S_n/\sqrt{n} \in [a,b]\big) \to \int_a^b \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy \quad \textit{as } n \to \infty \ .$$

Proof. Set $\phi(u) = \mathbb{E}\big(e^{iuX_1}\big)$. Since $\mathbb{E}(X_1^2) < \infty$, we can differentiate $\mathbb{E}\big(e^{iuX_1}\big)$ twice under the expectation, to show that (see problem 4.2(b))

$$\phi(0) = 1 \ , \quad \phi'(0) = 0 \ , \quad \phi''(0) = -1 \ .$$

Hence, by Taylor's theorem, $\phi(u) = 1 - u^2/2 + o(u^2)$ as $u \to 0$.
So, for the characteristic function ϕ_n of $S_n/\sqrt{n}$,

$$\phi_n(u) = \mathbb{E}\big(e^{iu(X_1+\ldots+X_n)/\sqrt{n}}\big) = \Big(\mathbb{E}\big(e^{i(u/\sqrt{n})X_1}\big)\Big)^n = \big(1 - u^2/(2n) + o(u^2/n)\big)^n \ .$$

The complex logarithm satisfies $\log(1+z) = z + o\big(|z|\big)$ as $z \to 0$, so, for each $u \in \mathbb{R}$,

$$\log \phi_n(u) = n \log\big(1 - u^2/(2n) + o(u^2/n)\big) = -u^2/2 + o(1) \ , \qquad \text{as } n \to \infty \ .$$

Hence $\phi_n(u) \to e^{-u^2/2}$ for all u. But $e^{-u^2/2}$ is the characteristic function of the $\mathcal{N}(0,1)$ distribution, so Lévy's convergence theorem completes the proof. □

Remarks. (i) This is only the simplest version of the central limit theorem. It holds in more general cases, e.g. for non-independent or not identically distributed r.v.s.

(ii) Problem 4.5 indicates that S_n/n can also converge to other (socalled *stable*) distributions than the Gaussian.

Appendix

A Example sheets

A.1 Example sheet 1 – Set systems and measures

1.1 Let E be a set.

(a) Let $\mathcal{F} \subseteq \mathcal{P}(E)$ be the set of all finite sets and their complements (called *cofinite* sets). Show that $\mathcal{F}$ is an algebra.

(b) Let $\mathcal{G} \subseteq \mathcal{P}(E)$ be the set of all countable sets and their complements. Show that $\mathcal{G}$ is a σ-algebra.

(c) Give a simple example of E and σ-algebras $\mathcal{E}_1$, $\mathcal{E}_2$, such that $\mathcal{E}_1 \cup \mathcal{E}_2$ is not a σ-algebra.

1.2 A non-empty set A in a σ-algebra $\mathcal{E}$ is called an *atom*, if there is no proper subset $B \subseteq A$ such that $B \in \mathcal{E}$. Let $A_1, \ldots, A_N$ be non-emtpy subsets of a set E.

(a) If the A_n are mutually disjoint and $\bigcup_n A_n = E$, how many elements does $\sigma\big(\{A_1, \ldots, A_N\}\big)$ have and what are its atoms?

(b) Show that in general $\sigma\big(\{A_1, \ldots, A_N\}\big)$ consists of only finitely many sets.

1.3 Show that the following families of subsets of $\mathbb{R}$ generate the same σ-algebra $\mathcal{B}$:
(i) $\{(a, b) : a < b\}$, (ii) $\{(a, b] : a < b\}$, (iii) $\{(-\infty, b] : b \in \mathbb{R}\}$.

1.4 A σ-algebra is called *separable* if it can be generated by a countable family of sets. Show that the Borel σ-algebra $\mathcal{B}$ of $\mathbb{R}$ is separable.

1.5 For which σ-algebras on $\mathbb{R}$ are the following set-functions measures:

$$\mu_1(A) = \begin{cases} 0 \ , \text{ if } A = \emptyset \\ 1 \ , \text{ if } A \neq \emptyset \end{cases}, \qquad \mu_2(A) = \begin{cases} 0 \ , \text{ if } A = \emptyset \\ \infty \ , \text{ if } A \neq \emptyset \end{cases}, \qquad \mu_3(A) = \begin{cases} 0 \ , \text{ if } A \text{ is finite} \\ 1 \ , \text{ if } A^c \text{ is finite} \end{cases} ?$$

1.6 Let $\mathcal{E}$ be a ring on E and $\mu : \mathcal{E} \to [0, \infty]$ an additive set function. Show that:

(a) If μ is continuous from below at all $A \in \mathcal{E}$ it is also countably additive.

(b) If μ is countably additive it is also countably **sub**additive.

1.7 Let $(E, \mathcal{E}, \mu)$ be a measure space. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of sets in $\mathcal{E}$, and define

$$\liminf A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_m \ , \qquad \limsup A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m \ .$$

(a) Show that $\mu(\liminf A_n) \leq \liminf_{n\to\infty} \mu(A_n)$.

(b) Show that $\mu(\limsup A_n) \geq \limsup_{n\to\infty} \mu(A_n)$ if $\mu(E) < \infty$.
Give an example with $\mu(E) = \infty$ when this inequality fails.

1.8 (a) Show that a π-system which is also a d-system is a σ-algebra.

(b) Give an example of a d-system that is not a σ-algebra.

1.9 (a) Find a Borel set that cannot be written as a countable union of intervals.

(b) Let $B \in \mathcal{B}$ be a Borel set with $\lambda(B) < \infty$, where λ is the Lebesgue measure. Show that, for every $\epsilon > 0$, there exists a finite union of disjoint intervals $A = (a_1, b_1] \cup \ldots \cup (a_n, b_n]$ such that $\lambda(A \triangle B) < \epsilon$, where $A \triangle B = (A \setminus B) \cup (B \setminus A)$.
(Hint: First consider all bounded sets B for which the conlusion holds and show that they form a d-system.)

1.10 Completion

Let $(E, \mathcal{E}, \mu)$ be a measure space. A subset N of E is called *null* if $N \subseteq B$ for some $B \in \mathcal{E}$ with $\mu(B) = 0$. Write $\mathcal{N}$ for the set of all null sets.

(a) Prove that the family of subsets $\mathcal{C} = \big\{A \cup N \,:\, A \in \mathcal{E},\ N \in \mathcal{N}\big\}$ is a σ-algebra.

(b) Show that the measure μ may be extended to a measure μ' on $\mathcal{C}$ with $\mu'(A \cup N) = \mu(A)$.

The σ-algebra $\mathcal{C}$ is called the *completion* of $\mathcal{E}$ with respect to μ.

1.11 Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events in the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, i.e. $A_n \in \mathcal{A}$ for all n. Show that the A_n, $n \in \mathbb{N}$, are independent if and only if the σ-algebras which they generate, $\mathcal{A}_n = \{\emptyset, A_n, A_n^c, \Omega\}$, are independent.

1.12 (a) Let μ_F be the Lebesgue-Stieltjes measure on $\mathbb{R}$ associated with the distribution function F. Show that F is continuous at x if and only if $\mu_F\big(\{x\}\big) = 0$.

(b) Let $(F_n)_{n\in\mathbb{N}}$, be a sequence of distribution functions on $\mathbb{R}$ such that $F(x) = \lim_{n\to\infty} F_n(x)$ exists for all $x \in \mathbb{R}$. Show that F need not be a distribution function.

1.13 Cantor set

Let $C_0 = [0, 1]$, and let $C_1, C_2, \ldots$ be constructed iteratively by deletion of middle-thirds.

Thus $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$ and so on.

The set $C = \lim_{n\to\infty} C_n = \bigcap_{n\in\mathbb{N}} C_n$ is called the *Cantor set*.

Let F_n be the distribution function of the uniform probability measure concentrated on C_n.

(a) Show that C is uncountable and has Lebesgue measure 0.

(b) Show that the limit $F(x) = \lim_{n\to\infty} F_n(x)$ exists for all $x \in [0, 1]$.
(Hint: Establish a recursion relation for $F_n(x)$ and use the contraction mapping theorem.)

(c) Show that F is continuous on $[0, 1]$ with $F(0) = 0$, $F(1) = 1$.

(d) Show that F is differentiable except on a set of measure 0, and that $F'(x) = 0$ wherever F is differentiable.

1.14 Riemann zeta function

The *Riemann zeta function* is given by $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$, $s > 1$.

Let $s{>}1$ and $\mathbb{P}_f : \mathcal{P}(\mathbb{N}) \to [0, 1]$ be the probability measure with mass function $f(n) = n^{-s}/\zeta(s)$.
For $p \in \{1, 2, \ldots\}$ let $A_p = \{n \in \mathbb{N} \,:\, p \mid n\}$ (p divides n).

(a) Show that the events $\{A_p \,:\, p \text{ prime}\}$ are independent. Deduce Euler's formula

$$\frac{1}{\zeta(s)} = \prod_{p \text{ prime}} \Big(1 - \frac{1}{p^s}\Big) .$$

(b) Show that $\mathbb{P}_f\big(\{n \in \mathbb{N} \,:\, n \text{ is square-free}\}\big) = \dfrac{1}{\zeta(2s)}$.

A.2 Example sheet 2 – Measurable functions and integration

Unless otherwise specified, let $(E, \mathcal{E})$, $(F, \mathcal{F})$ be measurable spaces and $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.

2.1 Let $f : E \to F$ be any function (not necessarily measurable).

(a) Show that $f^{-1}\big(\sigma(\mathcal{A})\big) = \sigma\big(f^{-1}(\mathcal{A})\big)$ for all $\mathcal{A} \subseteq \mathcal{P}(F)$.

(b) Let f be $\mathcal{E}/\mathcal{F}$-measurable. Under which circumstances is $f(\mathcal{E}) \subseteq \mathcal{P}(F)$ a σ-algebra?

(c) Take $(E, \mathcal{E}) = (F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$. Find the σ-algebras $\sigma(f_i)$ generated by the functions $f_1(x) = x\ ,\quad f_2(x) = x^2\ ,\quad f_3(x) = |x|\ ,\quad f_4(x) = \mathbb{1}_{\mathbb{Q}}(x)\ .$

2.2 Let $f_n : E \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$ be $\mathcal{E}/\overline{\mathcal{B}}$-measurable functions. Show that also the following functions are measurable, whenever they are well defined:

(a) $f_1 + f_2$ (b) $\inf_{n\in\mathbb{N}} f_n$ (c) $\sup_{n\in\mathbb{N}} f_n$ (d) $\liminf_{n\to\infty} f_n$ (e) $\limsup_{n\to\infty} f_n$.

(f) Deduce further that: $\big\{x \in E\ :\ f_n(x) \text{ converges as } n \to \infty\big\} \in \mathcal{E}$

2.3 Let $f : E \to \mathbb{R}^d$ be written in the form $f(x) = \big(f_1(x), \ldots, f_d(x)\big)$. Show that f is measurable w.r.t. $\mathcal{E}$ and $\mathcal{B}(\mathbb{R}^d)$ if and only if each $f_i : E \to \mathbb{R}$ is measurable w.r.t. $\mathcal{E}$ and $\mathcal{B}$.

2.4 Skorohod representation theorem
Let $F_n : \mathbb{R} \to [0, 1]$, $n \in \mathbb{N}$ be probability distribution functions. Consider the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ where $\Omega = (0, 1]$, $\mathcal{A} = \mathcal{B}((0, 1])$ are the Borel sets on $(0, 1]$ and $\mathbb{P}$ is the restriction of Lebesgue measure to $\mathcal{A}$. For each n define $X_n : (0, 1] \to \mathbb{R}\ ,\quad X_n(\omega) = \inf\big\{x\ :\ \omega \le F_n(x)\big\}$.

(a) Show that the X_n are random variables with distributions F_n. Are the X_n independent?

(b)* Suppose $F(x)$ is a probability distribution function such that $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}$ at which F is continuous. Let $X : (0, 1] \to \mathbb{R}$ be a random variable with distribution F defined analogously to the X_n. Show that $X_n \to X$ $a.s.$.

2.5 Let $X_1, X_2, \ldots$ be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$.

(a) Show that X_1 and X_2 are independent if and only if

$$\mathbb{P}\big(X_1 \le x,\ X_2 \le y\big) = \mathbb{P}(X_1 \le x)\,\mathbb{P}(X_2 \le y) \quad \text{for all } x, y \in \mathbb{R}\ . \qquad (*)$$

(b) Suppose $(*)$ holds for all pairs X_i, X_j, $i \ne j$. Is this sufficient for the $(X_n)_{n\in\mathbb{N}}$ to be independent? Justify your answer.

(c) Let X_1, X_2 be independent and identically distributed. Show that $X_1 = X_2$ almost surely implies that X_1 and X_2 are almost surely constant.

2.6 Let $X_1, X_2, \ldots$ be random variables with $X_n \xrightarrow{D} X$. Show that then also $h(X_n) \xrightarrow{D} h(X)$ for all continuous functions $h : \mathbb{R} \to \mathbb{R}$. (Hint: Use the Skorohod representation theorem)

2.7 Let $X_1, X_2, \ldots$ be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{T}$ the tail σ-algebra of $(X_n)_{n\in\mathbb{N}}$.
For each $n \in \mathbb{N}$ let $S_n = X_1 + \ldots + X_n$. Which of the following are tail events in $\mathcal{T}$,

$$\{X_n \le 0\ ev.\}\ ,\quad \{S_n \le 0\ i.o.\}\ ,\quad \big\{\liminf_{n\to\infty} S_n \le 0\big\}\ ,\quad \big\{\lim_{n\to\infty} S_n \text{ exists}\big\}\ ?$$

2.8 Let $X_1, X_2, \ldots$ be random variables with $\quad X_n = \begin{cases} n^2-1 & \text{with probability } 1/n^2 \\ -1 & \text{with probability } 1-1/n^2 \end{cases}$.

Show that $\mathbb{E}\Big(\dfrac{X_1+\cdots+X_n}{n}\Big) = 0$ for each n , but $\dfrac{X_1+\cdots+X_n}{n} \to -1$ almost surely.

2.9 Let $X, X_1, X_2, \ldots$ be random variables on $(\Omega, \mathcal{A}, \mathbb{P})$.

(a) Show that $\big\{\omega \in \Omega \,:\, X_n(\omega) \to X(\omega)\big\} \in \mathcal{A}$.

(b) Show that $X_n \to X$ almost surely $\Leftrightarrow$ $\sup_{m \geq n} |X_m - X| \to 0$ in probability .

2.10 Let $X_1, X_2, \ldots$ be independent random variables with distribution $\mathcal{N}(0,1)$. Prove that

$$\limsup_{n\to\infty} \big(X_n/\sqrt{2\log n}\big) = 1 \qquad \text{a.s.}$$

(Hint: Consider the events $A_n = \big\{X_n > \alpha\sqrt{2\log n}\big\}$ for $\alpha \in (0,\infty)$.)

2.11 Show that, as $n \to \infty$,

(a) $\displaystyle\int_0^\infty \sin(e^x)/\big(1+nx^2\big)\,dx \to 0$, (b) $\displaystyle\int_0^1 (n\cos x)/\big(1+n^2x^{3/2}\big)\,dx \to 0$.

2.12 Let $u,\, v : \mathbb{R} \to \mathbb{R}$ be differentiable on $[a,b]$ with continuous derivatives u' and v'.
Show that for $a < b$

$$\int_a^b u(x)\,v'(x)\,dx = \big[u(b)\,v(b) - u(a)\,v(a)\big] - \int_a^b u'(x)\,v(x)\,dx \,.$$

2.13 Let $\phi : [a,b] \to \mathbb{R}$ be continuously differentiable and strictly increasing. Show that for all continuous functions g on $\big[\phi(a), \phi(b)\big]$

$$\int_{\phi(a)}^{\phi(b)} g(y)\,dy = \int_a^b g\big(\phi(x)\big)\,\phi'(x)\,dx \,.$$

2.14 Show that the function $f(x) = x^{-1}\sin x$ is not Lebesgue integrable over $[1,\infty)$ but that

$$\lim_{y\to\infty}\int_0^y f(x)\,dx = \frac{\pi}{2}\,. \qquad \text{(use e.g. Fubini's theorem and } x^{-1} = \textstyle\int_0^\infty e^{-xt}\,dt)$$

2.15 (a) Let μ be a measure on $(E,\mathcal{E})$ and $f : E \to [0,\infty)$ be $\mathcal{E}/\mathcal{B}$-measurable with $\int_E f\,d\mu < \infty$. Define $\nu(A) = \int_A f\,d\mu$ for each $A \in \mathcal{E}$. Show that ν is a measure on $(E,\mathcal{E})$ and that

$$\int_E g\,d\nu = \int_E f\,g\,d\mu \quad \text{for all integrable } g : E \to \mathbb{R}\,.$$

(b) Let μ be a σ-finite measure on $(E,\mathcal{E})$. Show that for all $\mathcal{E}/\mathcal{B}$-measurable $g : E \to [0,\infty)$

$$\int_E g\,d\mu = \int_0^\infty \mu(g \geq \lambda)\,d\lambda\,.$$

A.3 Example sheet 3 – Convergence, Fubini, L^p-spaces

Unless otherwise specified, let $(E, \mathcal{E}, \mu)$ be a measure space and $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.

3.1 Let μ be the Lebesgue measure on $\big(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)\big)$.

(a) For $f(x,y) = \frac{x^2-y^2}{(x^2+y^2)^2}$ calculate the iterated Lebesgue integrals

$\int_0^1 \int_0^1 f(x,y)\, dx\, dy$ and $\int_0^1 \int_0^1 f(x,y)\, dy\, dx$.

What does the result tell about the double integral $\int_{(0,1)^2} f\, d\mu$?

(b) Show that for $f(x,y) = \begin{cases} \frac{x\,y}{(x^2+y^2)^2} & , (x,y) \neq (0,0) \\ 0 & , (x,y) = (0,0) \end{cases}$ the iterated integrals

$\int_{-1}^1 \int_{-1}^1 f(x,y)\, dx\, dy$ and $\int_{-1}^1 \int_{-1}^1 f(x,y)\, dy\, dx$

coincide, but that the double integral $\int_{(-1,1)^2} f\, d\mu$ does not exist.

(c) Let ν be the counting measure on $(\mathbb{R}, \mathcal{B})$, i.e. $\nu(A)$ is equal to the number of elements in A whenever A is finite, and $\nu(A) = \infty$ otherwise. Denote by $\Delta = \big\{(x,y) \in (0,1)^2 \ : \ x = y\big\}$ the diagonal in $(0,1)^2$ and calculate the iterated integrals

$\int_0^1 \int_0^1 \mathbb{1}_\Delta(x,y)\, dx\, \nu(dy)$ and $\int_0^1 \int_0^1 \mathbb{1}_\Delta(x,y)\, \nu(dy)\, dx$.

Does the result contradict Fubini's theorem?

3.2 (a) Are the following statements equivalent? (Justify your answer.)

(i) f is continuous almost everywhere, (ii) $f = g\ a.e.$ for a continuous function g.

(b) Let $X_n \sim U\big([-1/n, 1/n]\big)$ be uniform random variables on $[-1/n, 1/n]$ for $n \in \mathbb{N}$.

Do the X_n converge, and if yes in what sense?

3.3 Prove that the space $L^\infty(E, \mathcal{E}, \mu)$ is complete.

3.4 Let $p \in [1, \infty]$ and let $f_n, f \in L^p(E, \mathcal{E}, \mu)$ for $n \in \mathbb{N}$. Show that:

$$f_n \to f \ \text{ in } L^p \quad \Rightarrow \quad f_n \to f \ \text{ in measure}, \quad \text{but the converse is not true} .$$

3.5 Read hand-out 2 carefully. Find examples which show that the reverse implications, concerning the concepts of convergence on page 1, are in general false. How does the picture change if the measure space $(\Omega, \mathcal{A}, \mathbb{P})$ is not finite?

3.6 Let X be a random variable in $\mathbb{R}$ and let $1 \leq p < q < \infty$. Show that

$$\mathbb{E}\big(|X|^p\big) = \int_0^\infty p\, \lambda^{p-1}\, \mathbb{P}\big(|X| \geq \lambda\big)\, d\lambda$$

and deduce: $X \in L^q(\mathbb{P}) \ \Rightarrow \ \mathbb{P}\big(|X| \geq \lambda\big) = O(\lambda^{-q}) \ \Rightarrow \ X \in L^p(\mathbb{P})$.

Remark on questions 3.7(a) and 3.8(a): Start with an indicator function and extend your argument to the general case, analogous to the proof of Lemma 3.14(ii).

3.7 A *stepfunction* $g : \mathbb{R} \to \mathbb{R}$ is any finite linear combination of indicator functions of finite intervals.

(a) Show that the set of stepfunctions $\mathcal{I}$ is dense in $L^p(\mathbb{R})$ for all $p \in [1, \infty)$,

i.e. for all $f \in L^p(\mathbb{R})$ and every $\epsilon > 0$ there exists $g \in \mathcal{I}$ such that $\|f - g\|_p < \epsilon$.

(Hint: Use the result of question 1.9.)

(b) Using (a), argue that the set of continuous functions $C(\mathbb{R})$ is dense in $L^p(\mathbb{R})$, $p \in [1, \infty)$.

3.8 (a) Show that, if X and Y are independent random variables, then $\|X\,Y\|_1 = \|X\|_1\,\|Y\|_1$, but that the converse is in general not true.

(b) Show that, if X and Y are independent and integrable, then $\mathbb{E}(X\,Y) = \mathbb{E}(X)\,\mathbb{E}(Y)$.

3.9 Let $V_1 \subseteq V_2 \subseteq \ldots$ be an increasing sequence of closed subspaces of $L^2 = L^2(E, \mathcal{E}, \mu)$.
For $f \in L^2$, denote by f_n the orthogonal projection of f on V_n. Show that f_n converges in L^2.

3.10 Given a countable family of disjoint events $(G_i)_{i\in I}$, $G_i \in \mathcal{A}$, with $\bigcup_{i\in I} G_i = \Omega$.
Set $\mathcal{G} = \sigma\big(G_n : n \in \mathbb{N}\big)$ and $V = L^2(\Omega, \mathcal{G}, \mathbb{P})$.
Show that, for $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, the conditional expectation $\mathbb{E}(X\,|\,\mathcal{G})$ is a version of the orthogonal projection of X on V.

3.11 (a) Find a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ which is not bounded in L^1, but satisfies the other condition for uniform integrability, i.e.

$$\forall\, \epsilon > 0\ \exists\, \delta > 0\ \forall\, A \in \mathcal{A}\ \forall\, i \in I\ :\ \mathbb{P}(A) < \delta\ \Rightarrow\ \mathbb{E}\big(|X_i|\,\mathbb{1}_A\big) < \epsilon\ .$$

(b) Find a uniformly integrable sequence of random variables $(X_n)_{n\in\mathbb{N}}$ such that

$$X_n \to 0\ \ a.s.\quad \text{and}\quad \mathbb{E}\big(\sup_n |X_n|\big) = \infty\ .$$

3.12 Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of identically distributed r.v.s in $L^2(\mathbb{P})$. Show that, as $n \to \infty$,

(a) for all $\epsilon > 0$, $n\,\mathbb{P}\big(|X_1| > \epsilon\sqrt{n}\big) \to 0$,

(b) $n^{-1/2} \max_{k\le n} |X_k| \to 0$ in probability ,

(c) $n^{-1/2} \max_{k\le n} |X_k| \to 0$ in L^1 .

3.13 The *moment generating function* M_X of a real-valued random variable X is defined by

$$M_X(\theta) = \mathbb{E}\big(e^{\theta X}\big)\ ,\quad \theta \in \mathbb{R}.$$

(a) Show that the maximal domain of definition $I = \big\{\theta \in \mathbb{R}\ :\ M_X(\theta) < \infty\big\}$ is an interval and find examples for $I = \mathbb{R}$, $\{0\}$ and $(-\infty, 1)$.

Assume for simplicity that $X \ge 0$ from now on.

(b) Show that if I contains a neighbourhood of 0 then X has finite moments of all orders given by $\mathbb{E}(X^n) = \Big(\dfrac{d}{d\theta}\Big)^n\Big|_{\theta=0} M_X(\theta)$.

(c) Find a necessary and sufficient condition on the sequence of moments $m_n = \mathbb{E}(X^n)$ for I to contain a neighbourhood of 0.

A.4 Example sheet 4 – Characteristic functions, Gaussian rv's, ergodic theory

4.1 Let μ_1, μ_2 be finite measures on $(\mathbb{R}, \mathcal{B})$ such that $\displaystyle\int_{\mathbb{R}} g\, d\mu_1 = \int_{\mathbb{R}} g\, d\mu_2$ for all bounded continuous functions $g : \mathbb{R} \to \mathbb{R}$. Show that $\mu_1 = \mu_2$.

4.2 Let μ be a finite measure on $(\mathbb{R}, \mathcal{B})$ with Fourier transform $\hat{\mu}$. Show the following:

(a) $\hat{\mu}$ is a bounded continuous function.

(b) If $\displaystyle\int_{\mathbb{R}} |x|^k\, \mu(dx) < \infty$, then $\hat{\mu}$ has a k-th continuous derivative, which at 0 is given by

$$\hat{\mu}^{(k)}(0) = i^k \int_{\mathbb{R}} x^k\, \mu(dx) \ .$$

4.3 Let X be a real-valued random variable with characteristic function ϕ_X.

(a) Show that $\phi_X(u) \in \mathbb{R}$ for all $u \in \mathbb{R}$ if and only if $-X \sim X$, i.e. $\mu_{-X} = \mu_X$.

(b) Suppose that $|\phi_X(u)| = 1$ for all $|u| < \epsilon$ with some $\epsilon > 0$. Show that X is a.s. constant. (Hint: Take an independent copy X' of X, calculate $\phi_{X-X'}$ to see that $X = X'$ $a.s.$.)

4.4 By considering characteristic functions or otherwise, show that there do not exist iidrv's X, Y such that $X - Y$ is uniformly distributed on $[-1, 1]$.

4.5 The Cauchy distribution has density function $\displaystyle f(x) = \frac{1}{\pi(1 + x^2)}$, $x \in \mathbb{R}$.

(a) Show that the corresponding characteristic function is given by $\phi(u) = e^{-|u|}$.

(b) Show also that, if $X_1, \ldots, X_n$ are independent Cauchy random variables, then $(X_1 + \cdots + X_n)/n$ is also Cauchy.
Comment on this in the light of the strong law of large numbers and the central limit theorem.

4.6 Let $X, Y \sim \mathcal{N}(0, 1)$ and $Z \sim \mathcal{N}(0, \sigma^2)$ be independent Gaussian random variables. Calculate the characteristic function of $\eta = XY - Z$.

4.7 Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Prove Proposition 5.3, i.e. show that

(a) $\mathbb{E}(X) = \mu$,

(b) $\mathrm{var}(X) = \sigma^2$,

(c) $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$,

(d) $\phi_X(u) = e^{iu\mu - u^2\sigma^2/2}$.

4.8 Let $X_1, \ldots, X_n$ be independent $\mathcal{N}(0, 1)$ random variables. Show that

$$\Big(\overline{X}, \sum_{m=1}^{n} (X_m - \overline{X})^2\Big) \quad \text{and} \quad \Big(X_n/\sqrt{n}, \sum_{m=1}^{n-1} X_m^2\Big)$$

have the same distribution, where $\overline{X} = (X_1 + \cdots + X_n)/n$.

4.9 Show that the shift map θ of Definition 6.3 is measurable and measure-preserving.

4.10 Let $E = [0, 1)$ with Lebesgue measure. For $a \in E$ consider the mapping

$$\theta_a : E \to E\ , \quad \theta_a(x) = (x + a) \bmod 1\ .$$

(a) Show that θ_a is measure-preserving.

(b) Show that θ_a is not ergodic when a is rational.

(c) Show that θ_a is ergodic when a is irrational.
(Hint: Consider $a_n = \int_E f(x)\, e^{2\pi i n x}\, dx$ to show that every invariant function is constant.)

(d) Let $f : E \to \mathbb{R}$ be integrable. Determine for each $a \in E$ the limit function

$$\bar{f} = \lim_{n\to\infty} \big(f + f \circ \theta_a + \ldots + f \circ \theta_a^{n-1}\big)/n \, .$$

4.11 Show that $\theta(x) = 2x \mod 1$ is a measure-preserving transformation on $E = [0, 1)$ with Lebesgue measure, and that θ is ergodic. Find $\bar{f}$ for each integrable function f.
(Hint: Consider the binary expansion $x = 0.x_1x_2x_3\ldots$ and use that $X_n(x) = x_n$ are iidrvs with $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = \frac{1}{2}$, which is proved on hand-out 2.)

4.12 Call a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ on a common probability space *stationary* if for each $n, k \in \mathbb{N}$ the random vectors $(X_1, \ldots, X_n)$ and $(X_{k+1}, \ldots, X_{k+n})$ have the same distribution, i.e. for $A_1, \ldots, A_n \in \mathcal{B}$,

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \mathbb{P}(X_{k+1} \in A_1, \ldots, X_{k+n} \in A_n) \, .$$

Show that, if $(X_n)_{n\in\mathbb{N}}$ is a stationary sequence and $X_1 \in L^p$, for some $p \in [1, \infty)$, then

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to X \;\; a.s. \text{ and in } L^p \, ,$$

for some random variable $X \in L^p$, and find $\mathbb{E}(X)$.

4.13 Find a sequence $(X_n)_{n\in\mathbb{N}}$ of independent random variables with $\mathbb{E}\big(|X_n|\big) < \infty$ and $\mathbb{E}(X_n) = 0$ for all $n \in \mathbb{N}$, such that $(X_1 + \ldots + X_n)/n$ does not almost surely converge to 0.

4.14 Let $(X_n)_{n\in\mathbb{N}}$ be independent random variables with $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 1) = \frac{1}{2}$, and define

$$U_n = X_1X_2 + X_2X_3 + \ldots + X_{2n}X_{2n+1} \, .$$

Show that $U_n/n \to c \;\; a.s.$ for some $c \in \mathbb{R}$, and determine c.

B Hand-outs

B.1 Hand-out 1 – Proof of Carathéodory's extension theorem

Theorem 1.4. Carathéodory's extension theorem
Let $\mathcal{E}$ be a ring on E and $\mu : \mathcal{E} \to [0,\infty]$ be a countably additive set function. Then there exists a measure μ' on $\big(E, \sigma(\mathcal{E})\big)$ such that $\quad \mu'(A) = \mu(A) \quad$ *for all $A \in \mathcal{E}$.*

Proof. For any $B \subseteq E$, define the *outer measure* $\qquad \mu^*(B) = \inf \sum_n \mu(A_n)$,

where the infimum is taken over all sequences $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that $B \subseteq \bigcup_n A_n$ and is taken to be ∞ if there is no such sequence. Note that μ^* is increasing and $\mu^*(\emptyset) = 0$. Let us say that $A \subseteq E$ is μ^*-*measurable* if, for all $B \subseteq E$,

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) .$$

Write $\mathcal{M}$ for the set of all μ^*-measurable sets. *We shall show that $\mathcal{M}$ is a σ-algebra containing $\mathcal{E}$ and that μ^* is a measure on $\mathcal{M}$, extending μ.* This will prove the theorem.

Step I. We show that μ^* is countably subadditive.
Suppose that $B \subseteq \bigcup_n B_n$. If $\mu^*(B_n) < \infty$ for all n, then, given $\epsilon > 0$, there exist sequences $(A_{nm})_{m\in\mathbb{N}}$ in $\mathcal{E}$, with

$$B_n \subseteq \bigcup_m A_{nm}, \quad \mu^*(B_n) + \epsilon/2^n \geq \sum_m \mu(A_{nm}) .$$

Then $\quad B \subseteq \bigcup_n \bigcup_m A_{nm} \quad$ and thus $\quad \mu^*(B) \leq \sum_n \sum_m \mu(A_{nm}) \leq \sum_n \mu^*(B_n) + \epsilon$.

Hence, in any case $\quad \mu^*(B) \leq \sum_n \mu^*(B_n)$.

Step II. We show that μ^* extends μ.
Since $\mathcal{E}$ is a ring and μ is countably additive, μ is countably subadditive. Hence, for $A \in \mathcal{E}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{E}$ with $A \subseteq \bigcup_n A_n$, we have $\qquad \mu(A) \leq \sum_n \mu(A_n)$.

On taking the infimum over all such sequences, we see that $\mu(A) \leq \mu^*(A)$. On the other hand, it is obvious that $\mu^*(A) \leq \mu(A)$ for $A \in \mathcal{E}$.

Step III. We show that $\mathcal{M}$ contains $\mathcal{E}$.
Let $A \in \mathcal{E}$ and $B \subseteq E$. We have to show that

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) .$$

By subadditivity of μ^*, it is enough to show that

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c) .$$

If $\mu^*(B) = \infty$, this is trivial, so let us assume that $\mu^*(B) < \infty$. Then, given $\epsilon > 0$, we can find a sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that

$$B \subseteq \bigcup_n A_n , \quad \mu^*(B) + \epsilon \geq \sum_n \mu(A_n) .$$

Then $\quad B \cap A \subseteq \bigcup_n (A_n \cap A)\ , \quad B \cap A^c \subseteq \bigcup_n (A_n \cap A^c)\ ,$ so that

$$\mu^*(B \cap A) + \mu^*(B \cap A^c) \le \sum_n \mu(A_n \cap A) + \sum_n \mu(A_n \cap A^c) = \sum_n \mu(A_n) \le \mu^*(B) + \epsilon\ .$$

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

Step IV. We show that $\mathcal{M}$ is an algebra.
Clearly $E \in \mathcal{M}$ and $A^c \in \mathcal{E}$ whenever $A \in \mathcal{E}$. Suppose that $A_1, A_2 \in \mathcal{M}$ and $B \subseteq E$. Then

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

Hence $A_1 \cap A_2 \in \mathcal{M}$.

Step V. We show that $\mathcal{M}$ is a σ-algebra and that μ^* is a measure on $\mathcal{M}$.
We already know that $\mathcal{M}$ is an algebra, so it suffices to show that, for any sequence of disjoint sets $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{M}$, for $A = \bigcup_n A_n$ we have

$$A \in \mathcal{M}\ , \quad \mu^*(A) = \sum_n \mu^*(A_n).$$

So, take any $B \subseteq E$, then

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) = \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\ &= \ldots = \sum_{i=1}^n \mu^*(B \cap A_i) + \mu^*(B \cap A_1^c \cap \ldots \cap A_n^c)\ . \end{aligned}$$

Note that $\mu^*(B \cap A_1^c \cap \ldots \cap A_n^c) \ge \mu^*(B \cap A^c)$ for all n. Hence, on letting $n \to \infty$ and using countable subadditivity, we get

$$\mu^*(B) \ge \sum_{n=1}^\infty \mu^*(B \cap A_n) + \mu^*(B \cap A^c) \ge \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

The reverse inequality holds by subadditivity, so we have equality. Hence $A \in \mathcal{M}$ and, setting $B = A$, we get $\quad \mu^*(A) = \sum_{n=1}^\infty \mu^*(A_n)\ .$ □

B.2 Hand-out 2 – Convergence of random variables

Let $X, X_n : \Omega \to \mathbb{R}$ be random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with distributions μ, μ_n. There are several concepts of convergence of random variables, which are summarised in the following:

(i) $X_n \to X$ *everywhere* or *pointwise* if $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$ as $n \to \infty$.

(ii) $X_n \xrightarrow{a.s.} X$ *almost surely* (*a.s.*) if $\mathbb{P}(X_n \not\to X) = 0$.

(iii) $X_n \xrightarrow{P} X$ *in probability* if $\forall\, \epsilon > 0 \ : \ \mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

(iv) $X_n \xrightarrow{L^p} X$ *in L^p* for $p \in [1, \infty]$, if $\|X_n - X\|_p \to 0$ as $n \to \infty$.

(v) $X_n \xrightarrow{D} X$ in *distribution* or in *law* if $\mathbb{P}(X_n \le x) \to \mathbb{P}(X \le x)$ as $n \to \infty$, for all continuity points of $\mathbb{P}(X \le x)$. Since equivalent to (vi), this is often also called *weak convergence*.

(vi) $\mu_n \Rightarrow \mu$ *weakly* if $\displaystyle\int_{\mathbb{R}} f\, d\mu_n \to \int_{\mathbb{R}} f\, d\mu$ for all $f \in C_b(\mathbb{R}, \mathbb{R})$.

The following implications hold $(q \ge p \ge 1)$:

$$
\begin{array}{ccccccccc}
X_n \to X & \Rightarrow & X_n \xrightarrow{a.s.} X & & & & & & \\
 & & & \searrow & & & & & \\
 & & & & X_n \xrightarrow{P} X & \Rightarrow & X_n \xrightarrow{D} X & \Leftrightarrow & \mu_n \Rightarrow \mu \\
 & & & \nearrow & & & & & \\
X_n \xrightarrow{L^q} X & \Rightarrow & X_n \xrightarrow{L^p} X & & & & & &
\end{array}
$$

Proofs are given in Theorem 2.10, Proposition 2.11 (see below), Theorem 3.9 (see below), Corollary 5.3 and example sheet question 3.2.

Proof. of Proposition 2.11: $X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{D} X$
Suppose $X_n \xrightarrow{P} X$ and write $F_n(x) = \mathbb{P}(X_n \le x)$, $F(x) = \mathbb{P}(X \le x)$ for the distr. fcts.
If $\epsilon > 0$, $F_n(x) = \mathbb{P}\big(X_n \le x,\ X \le x+\epsilon\big) + \mathbb{P}\big(X_n \le x,\ X > x+\epsilon\big) \le F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$.
Similarly, $F(x-\epsilon) = \mathbb{P}\big(X \le x-\epsilon,\ X_n \le x\big) + \mathbb{P}\big(X \le x-\epsilon,\ X_n > x\big) \le F_n(x) + \mathbb{P}(|X_n - X| > \epsilon)$.
Thus $F(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \le F_n(x) \le F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$, and as $n \to \infty$

$$F(x-\epsilon) \le \liminf_n F_n(x) \le \limsup_n F_n(x) \le F(x+\epsilon) \quad \text{for all } \epsilon > 0 \, .$$

If F is continuous at x, $F(x-\epsilon) \nearrow F(x)$ and $F(x+\epsilon) \searrow F(x)$ as $\epsilon \to 0$, proving the result. □

Proof. of Theorem 3.9: $X_n \xrightarrow{D} X \quad \Leftrightarrow \quad \mu_n \Rightarrow \mu$
Suppose $X_n \xrightarrow{D} X$. Then by the Skorohod theorem 2.12 there exist $Y \sim X$ and $Y_n \sim X_n$ on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that, $f(Y_n) \to f(Y)\ a.e.$ since $f \in C_b(\mathbb{R}, \mathbb{R})$. Thus

$$\int_{\mathbb{R}} f\, d\mu_n = \int_{\Omega} f(Y_n)\, d\mathbb{P} \to \int_{\Omega} f(Y)\, d\mathbb{P} = \int_{\mathbb{R}} f\, d\mu \quad \text{and} \quad \mu_n \Rightarrow \mu \quad \text{by bounded convergence .}$$

Suppose $\mu_n \to \mu$ and let y be a continuity point of F_X.
For $\delta > 0$, approximate $\mathbb{1}_{(-\infty,y]}$ by $\quad f_\delta(x) = \begin{cases} \mathbb{1}_{(-\infty,y]}(x) & , x \notin (y, y+\delta) \\ 1 + (y-x)/\delta & , x \in (y, y+\delta) \end{cases}$ such that

$$\Big| \int_{\mathbb{R}} (\mathbb{1}_{(-\infty,y]} - f_\delta)\, d\mu \Big| \leq \Big| \int_{\mathbb{R}} g_\delta\, d\mu \Big| \quad \text{where} \quad g_\delta(x) = \begin{cases} 1+(x-y)/\delta & , x \notin (y-\delta, y) \\ 1+(y-x)/\delta & , x \in [y, y+\delta) \\ 0 & , \text{otherwise} \end{cases} .$$

The same inequality holds for μ_n for all $n \in \mathbb{N}$. Then as $n \to \infty$

$$\big|F_{X_n}(y) - F_X(y)\big| = \Big| \int_{\mathbb{R}} \mathbb{1}_{(-\infty,y]}\, d\mu_n - \int_{\mathbb{R}} \mathbb{1}_{(-\infty,y]}\, d\mu \Big| \leq$$
$$\leq \Big| \int_{\mathbb{R}} g_\delta\, d\mu_n \Big| + \Big| \int_{\mathbb{R}} g_\delta\, d\mu \Big| + \Big| \int_{\mathbb{R}} f_\delta\, d\mu_n - \int_{\mathbb{R}} f_\delta\, d\mu \Big| \to 2 \Big| \int_{\mathbb{R}} g_\delta\, d\mu \Big| ,$$

since $f_\delta, g_\delta \in C_b(\mathbb{R}, \mathbb{R})$. Now, $\quad \big| \int_{\mathbb{R}} g_\delta\, d\mu \big| \leq \mu\big((y-\delta, y+\delta)\big) \to 0$ as $\delta \to 0$, $\quad$ since $\mu\big(\{y\}\big) = 0$, so $X_n \xrightarrow{D} X$. □

Skorohod representation theorem
For all probability distribution functions $F_1, F_2, \ldots : \mathbb{R} \to [0,1]$ there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and random variables $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ such that X_n has distribution function F_n.

- *(a) The X_n can be chosen to be independent.*
- *(b) If $F_n \to F$ for all continuity points of the probability distribution function F, then the X_n can also be chosen such that $X_n \to X$ $a.s.$ with $X : \Omega \to \mathbb{R}$ having distribution function F.*

Proof. Consider the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ where $\Omega = (0,1]$, $\mathcal{A} = \mathcal{B}\big((0,1]\big)$ and $\mathbb{P}$ is the restriction of Lebesgue measure to $\mathcal{A}$. For each $n \in \mathbb{N}$ define $G_n : (0,1] \to \mathbb{R}$, $\quad G_n(\omega) = \inf\big\{x \,:\, \omega \leq F_n(x)\big\}$.
(b) In problem 2.4 it is shown that $X_n = G_n$ are random variables with distribution functions F_n and that $X_n \to X$ $a.s.$ under the assumption in (b), where $X(\omega) = G(\omega)$ is defined analogously.
(a) Each $\omega \in \Omega$ has a unique binary expansion $\omega = 0.\omega_1\omega_2\omega_3\ldots$, where we forbid infinite sequences of 0's. The **Rademacher functions** $R_n : \Omega \to \{0,1\}$ are defined as $R_n(\omega) = \omega_n$. Note that

$$R_1 = \mathbb{1}_{(\frac{1}{2},1]} , \quad R_2 = \mathbb{1}_{(\frac{1}{4},\frac{1}{2}]} + \mathbb{1}_{(\frac{3}{4},1]} , \quad R_3 = \mathbb{1}_{(\frac{1}{8},\frac{1}{4}]} + \mathbb{1}_{(\frac{3}{8},\frac{1}{2}]} + \mathbb{1}_{(\frac{5}{8},\frac{3}{4}]} + \mathbb{1}_{(\frac{7}{8},1]} , \quad \ldots$$

thus in general $\quad R_n = \mathbb{1}_{A_n} \quad$ where $\quad A_n = \bigcup_{k=1}^{2^{n-1}} I_{n,k} \quad$ and $\quad I_{n,k} = \Big(\dfrac{2k-1}{2^n}, \dfrac{2k}{2^n}\Big]$.
With problem 1.11 the R_n are independent if and only if the A_n are. To see this, take $n_1 < \ldots < n_L$ for some $L \in \mathbb{N}$ and we see that $A_{n_1}, \ldots, A_{n_L}$ are independent by induction, using that

$$\mathbb{P}(I_{n_l,k} \cap A_{n_{l+1}}) = \tfrac{1}{2}\mathbb{P}(I_{n_l,k}) = \mathbb{P}(I_{n_l,k})\,\mathbb{P}(A_{n_{l+1}}) \quad \text{for all } k = 1, \ldots, 2^{n_l - 1} ,$$

and thus $\qquad \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_l} \cap A_{n_{l+1}}) = \mathbb{P}(A_{n_1} \cap \ldots \cap A_{n_l})\,\mathbb{P}(A_{n_{l+1}})$.

Now choose a bijection $\quad m : \mathbb{N}^2 \to \mathbb{N} \quad$ and set $\quad Y_{k,n} = R_{m(k,n)} \quad$ and $\quad Y_n = \sum_{k=1}^{\infty} 2^{-k} Y_{k,n}$.

Then $Y_1, Y_2, \ldots$ are independent and $\quad \mathbb{P}\big(i\,2^{-k} < Y_n \leq (i+1)\,2^{-k}\big) = 2^{-k} \quad$ for all n, k, i .
Thus $\mathbb{P}(Y_n \leq x) = x$ for all $x \in (0,1]$. So $X_n = G_n(Y_n)$ are independent random variables with distribution F_n, which can be shown analogous to (b). □

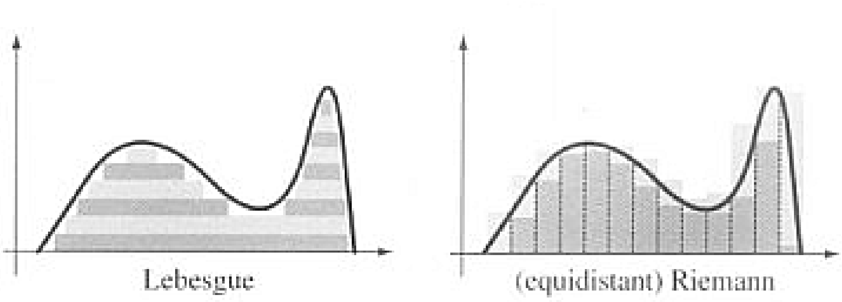
B.3 Hand-out 3 – Connection between Lebesgue and Riemann integration

Definition. $f : [a,b] \to \mathbb{R}$ is *Riemann integrable* (R-integrable) with integral $R \in \mathbb{R}$, if

$$\forall\, \epsilon > 0 \;\; \exists\, \delta > 0 \;\; \forall\, x_j \in I_j \; : \; \Big| R - \sum_{j=1}^{n} f(x_j)\, |I_j| \Big| < \epsilon\,, \tag{B.1}$$

for some finite partition $\{I_1, \ldots, I_n\}$ of $[a,b]$ into subintervals of lengths $|I_j| < \delta$.

This corresponds to an approximation of f by step functions $\sum_{j=1}^{n} f(x_j)\, \mathbb{1}_{I_j}$, a special case of simple functions which are constant on intervals.



The picture is taken from R.L. Schilling, *Measures, Integrals and Martingales,* CUP 2005. He writes:

> ... the Riemann sums partition the domain of the function without taking into account the shape of the function, thus slicing up the area under the function *vertically*. Lebesgue's approach is exactly the opposit: the domain is partitioned according to the values of the function at hand, leading to a *horizontal* decomposition of the area.

Theorem. Lebesgue's integrability criterium
$f : [a,b] \to \mathbb{R}$ is R-integrable if and only if f is bounded on $[a,b]$ and continuous almost everywhere, i.e. the set of points in $[a,b]$ where f is not continuous has Lebesgue measure 0.

Corollary. *Suppose $f : [a,b] \to \mathbb{R}$ is R-integrable. Then it is also Lebesgue integrable (L-integrable) and the values of both integrals coincide.*

Proof. For a partition $\{I_1, \ldots, I_n\}$ define the step functions

$$\bar{g}_n = \sum_{j=1}^{n} \sup\{f(x) : x \in I_j\}\, \mathbb{1}_{I_j}\;, \quad \underline{g}_n = \sum_{j=1}^{n} \inf\{f(x) : x \in I_j\}\, \mathbb{1}_{I_j}\;.$$

Thus $\underline{g}_n \leq f \leq \bar{g}_n$ and if f is continuous $a.e.$, $\underline{g}_n, \bar{g}_n \to f \;\; a.e.$ as $n \to \infty$ and $|I_j| \to 0$.
Since f is bounded, it follows by dominated convergence for the L-integrals

$$\int_a^b \underline{g}_n(x)\, dx \to \int_a^b f(x)\, dx\;, \quad \int_a^b \bar{g}_n(x)\, dx \to \int_a^b f(x)\, dx\;,$$

so the sum in (B.1) converges and $R = \int_a^b f(x)\,dx$.
On the other hand, if the sum in (B.1) converges, then $\int_a^b \big|\bar{g}_n(x) - \underline{g}_n(x)\big|\,dx \to 0$ and thus for the limit functions $\underline{g} = \bar{g}$ $a.e.$, and f is continuous $a.e.$ since $\underline{g} \leq f \leq \bar{g}$.
If f was not bounded, one could choose x_j in (B.1) such that the sum does not converge. □

On the other hand, not every L-integrable function is also R-integrable. The standard example is $f = \mathbb{1}_{[0,1]\cap\mathbb{Q}}$, which can be made R-integrable by changing it on a set of L-measure 0.
This might suggest that for every L-integrable f there exists an R-integrable g with $f = g$ $a.e.$.

This is not true as demonstrated by the following example:
Let $\{r_1, r_2 \ldots\}$ be an enumeration of the rationals in $(0,1)$. For small $\epsilon > 0$ and each $n \in \mathbb{N}$ choose an open interval $I_n \subseteq (0,1)$ with $r_n \in I_n$ and L-measure $\mu(I_n) < \epsilon\, 2^{-n}$. Put $A = \bigcup_n I_n$. Then A is dense in $(0,1)$ with $0 < \mu(A) < \epsilon$ and thus for any non-degenerate subinterval I of $(0,1)$, $\mu(A \cap I) > 0$.
Take $f = \mathbb{1}_A$ and suppose that $f = g$ a.e.. Let $\{I_j\}$ be some decomposition of $(0,1)$ into subintervals. Since for each j, $\mu\big(I_j \cap A \cap \{f = g\}\big) = \mu(I_j \cap A) > 0$, $g(x_j) = f(x_j) = 1$ for some $x_j \in I_j \cap A$, and thus

$$\sum_{j=1}^{n} g(x_j)\,\mu(I_j) = 1 > \mu(A) \ . \tag{B.2}$$

If g were R-integrable, its integral would have to coincide with the L-integral $\int_0^1 f\,d\mu = \mu(A)$, which is in contradiction to (B.2).

B.4 Hand-out 4 – Ergodic theorems

Hand-out 4 contains statements and proofs of Lemma 6.5 and Theorems 6.6 and 6.7, which can be found in Section 6.3.