Probability Theory I

Notes of a lecture held in winter 2019/2020
by Prof. Dr. Christoph Thäle

Nils Müller
Simon Damm

Bochum, April 5, 2020

In the accompanying lecture, we developed the mathematical foundations of probability theory. The topics of this course form the basis for all further lectures in probability and statistics held at Ruhr-Universität Bochum. Prerequisites are the Introduction to Probability and Statistics, Linear Algebra and Geometry I/II as well Analysis I-III. Since a particular focus has been on analytic methods, some basic knowledge in complex analysis is helpful but not strictly required. Prior experience with measure theory is also useful.

Contents

1 A Crash Course in Measure Theory

1.1 Elementary Definitions in Measure Theory

Definition 1.1 (σ-algebra, Measureable space) ▷ 08.10.2019
Let $\Omega \neq \emptyset$. $\mathcal{A} \subseteq 2^\Omega$ is called a σ-algebra over Ω if

1. $\Omega \in \mathcal{A}$,
2. $A \in \mathcal{A} \Rightarrow A^{\mathrm{C}} \in \mathcal{A}$ *and*
3. $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

$(\Omega, \mathcal{A})$ *is called a measurable space.*

Definition 1.2 (Measure, Measure space)
Let $(\Omega, \mathcal{A})$ be a measurable space, then a function $\mu : \mathcal{A} \to [0, \infty]$ is called a measure on $(\Omega, \mathcal{A})$ if

- $\mu(\emptyset) = 0$ *and*
- μ *is σ-additive, i.e. for all $A_1, A_2, \ldots \in \mathcal{A}$ pairwise disjoint,*
  we have $\mu(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i)$.

We call $(\Omega, \mathcal{A}, \mu)$ a measure space.

Definition 1.3 (Probability measure, Probability space, Event)
If μ is a measure on $(\Omega, \mathcal{A})$ with $\mu(\Omega) = 1$, we call it a probability measure, $(\Omega, \mathcal{A}, \mu)$ a probability space and $A \in \mathcal{A}$ an event.

Remark 1.1 *We write $\mathbb{P} := \mu$ if μ is a probability measure.*

Further Concepts Related to Measures

- *Generated σ-algebra:* Let $\mathcal{M} \subseteq 2^\Omega$, then

$$\sigma(\mathcal{M}) := \bigcap_{\substack{\mathcal{A}\ \sigma\text{-algebra} \\ \mathcal{M} \subseteq \mathcal{A}}} \mathcal{A}$$

  is called the generated σ-algebra of $\mathcal{M}$ over Ω.

- *Initial σ-algebra:* Let I be an arbitrary index set and for $i \in I$ let $f_i : \Omega \to \Omega_i$ be a function and $(\Omega_i, \mathcal{A}_i)$ be a measurable space. We call

$$\mathcal{I}((f_i)_{i \in I}) := \sigma\left(\bigcup_{i \in I} f_i^{-1}(\mathcal{A}_i)\right)$$

  the initial σ-algebra on Ω generated by $(f_i)_{i \in I}$.

- *Borel σ-algebra:* Let (T, τ) be a topological space, then then then Borel σ-algebra of T over τ is defined as $\mathcal{B}(T) := \sigma(\tau)$.

- *Finite measure:* Measure μ on Ω is finite if $\mu(\Omega) < \infty$.

- *σ-finite measure:* Ω can be covered by at most countably measurable sets in $\mathcal{A}$ with finite measure with respect to μ.

- *Dirac measure:* Let $\omega \in \Omega$, $\mathcal{A}$ a σ-algebra over Ω and $A \in \mathcal{A}$, then

$$\delta_\omega(A) := \mathbb{1}_A(\omega) := \begin{cases} 0, & \omega \notin A; \\ 1, & \omega \in A \end{cases}$$

  is called a Dirac measure on Ω.

· *Semiring:* $\mathcal{H} \subseteq 2^\Omega$ is called a semiring iff

1. $\emptyset \in \mathcal{H}$,
2. $\mathcal{H}$ is $\cap$-stable, i.e. $A, B \in \mathcal{H} \Rightarrow A \cap B \in \mathcal{H}$ and
3. $A, B \in \mathcal{H} \Rightarrow \exists C_1, \ldots, C_n \in \mathcal{H}$ pairwise disjoint with $A \setminus B = C_1 \cup \ldots \cup C_n$.

Example: $\Omega = \mathbb{R}, \mathcal{H} = \{(a, b] \mid a \leq b\}$.

· *Content:* $\mu : \mathcal{H} \to [0, \infty]$ such that $\mu(\emptyset) = 0$, μ finitely additive is called a content.

· *Pre-measure:* $\mu : \mathcal{H} \to [0, \infty]$ such that μ is a content and μ is σ-additive is called a pre-measure.

1.2 Important Theorems in Measure Theory

Theorem 1.1 (Uniqueness theorem for measures)
Let $(\Omega, \mathcal{A})$ be a measure space, $\mathcal{M} \subseteq 2^\Omega$ $\cap$-stable with $\sigma(\mathcal{M}) = \mathcal{A}$ and μ_1, μ_2 measures on $(\Omega, \mathcal{A})$ with

$$\mu_1(B) = \mu_2(B) \text{ for all } B \in \mathcal{M}.$$

If there is a sequence $(B_n)_{n \in \mathbb{N}}$ in $\mathcal{M}$ with $B_n \uparrow \Omega$ and $\mu_1(B_n) < \infty$ for all $n \in \mathbb{N}$, then $\mu_1 = \mu_2$.

Theorem 1.2 (Extension theorem for measures)
Let $\mathcal{H} \subseteq 2^\Omega$ be a semiring and $\mu : \mathcal{H} \to [0, \infty]$ be a pre-measure, then there is a measure $\tilde{\mu}$ on $(\Omega, \sigma(\mathcal{H}))$ with

$$\tilde{\mu}(A) = \mu(A) \text{ for all } A \in \mathcal{H}.$$

If μ is σ-finite, then $\tilde{\mu}$ is unique.

Definition 1.4 (Measure determining function)
A function $G : \mathbb{R} \to \mathbb{R}$ is called measure determining if

1. *G is increasing and*
2. *G is right continuous.*

If $G(-\infty) = 0, G(\infty) = 1$, then G is called a distribution function.

Theorem 1.3 *Let G be a measure determining function, then there exists a unique measure μ_G on $\mathcal{B}(\mathbb{R})$ with*

$$\mu([a, b]) = G(b) - G(a) \text{ , for all } a, b \in \mathbb{R} \text{ with } a \leq b.$$

If G is a distribution function, μ is a probability measure.

Example 1.1 *Let $G(x) = x$, then μ_G is called the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Definition 1.5 (Measurable function) ▷ 10.10.2019
Let $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ be measurable spaces. A function $f : \Omega \to \Omega'$ is called $\mathcal{A}$-$\mathcal{A}'$-measurable if

$$f^{-1}(A') \in \mathcal{A}, \text{ for all } A' \in \mathcal{A}'.$$

Remark 1.2 *If the respective measure spaces referred to are clear by context, we also say f is measurable.*

Theorem 1.4 *Let μ be a measure on Ω and $f : \Omega \to \Omega'$ be $\mathcal{A}$-$\mathcal{A}'$-measurable, then*

$$\mu^f(A') := \mu(f^{-1}(A'))$$

determines a measure on $(\Omega', \mathcal{A}')$. μ^f is called image measure of μ under f.

Remark 1.3 *We write μ^f or $\mu \circ f^{-1}$ for the image measure of μ under f.*

Definition 1.6 (Random variable, Distribution)
If $\mu = \mathbb{P}$ is a probability measure on $(\Omega, \mathcal{A})$, then a measurable function $X : (\Omega, \mathcal{A}, \mu) \to (\Omega', \mathcal{A}')$ is called a random variable with values in Ω'. We call $\mathbb{P}^X = \mathbb{P} \circ X^{-1}$ the distribution of X.

Remark 1.4 *To denote that ν is the distribution of a random variable X, we also write $X \sim \nu$. If X, Y are random variables with equal distributions, then we write $X \overset{\text{d}}{=} Y$ or $X \sim Y$.*

Important Facts about Measurable Functions

- Let $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ be measurable spaces and $f : \Omega \to \Omega'$. If

$$\mathcal{A}' = \sigma(\mathcal{M}') \text{ and } f^{-1}(\mathcal{M}') \subseteq \mathcal{A}\,,$$

  then f is $\mathcal{A}$-$\mathcal{A}'$-measurable.

- If additionally $\mathcal{A}'$ is generated by a topology, one speaks about Borel measurability.

- If $\Omega' = \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ we have to consider the σ-algebra

$$\mathcal{B}(\overline{\mathbb{R}}) = \{B \cup E \mid B \in \mathcal{B}(\mathbb{R}), E \subseteq \{\pm\infty\}\}\,.$$

1.3 Construction of an Integral for Measurable Functions

Definition 1.7 (Integral)
The integral is constructed in four steps. Let in the following $(\Omega, \mathcal{A}, \mu)$ *be a measure space and* $f : \Omega \to \overline{\mathbb{R}}$ *be* $\mathcal{A}$-$\mathcal{B}(\overline{\mathbb{R}})$*-measurable.*

1. *If* $f = \mathbb{1}_A$ *for* $A \in \mathcal{A}$, *then*

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega 1_A \,\mathrm{d}\mu := \mu(A)\,.$$

2. *If instead* f *is a non-negative step function, i.e.*

$$f \in \mathcal{E} := \{g : \Omega \to \overline{\mathbb{R}} \mid g \geq 0 \textit{ is } \mathcal{A}\text{-}\mathcal{B}(\overline{\mathbb{R}})\textit{-measurable, } |g(\Omega)| < \infty\}\,,$$

   then there exist $n \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_n \in \overline{\mathbb{R}}_{\geq 0}$ *and* $A_1, \ldots, A_n \in \mathcal{A}$ *such that*

$$f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}\,.$$

   In this case, we can define the integral of f *with respect to* μ *as*

$$\int_\Omega f \,\mathrm{d}\mu := \sum_{i=1}^{n} \alpha_i \int_\Omega \mathbb{1}_{A_i} \,\mathrm{d}\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i)\,.$$

3. *If instead* f *is a non-negative function, i.e.* $f(\Omega) \subseteq [0, \infty]$, *then as shown by Lebesgue there exists* $(u_n)_{n\in\mathbb{N}} \in \mathcal{E}$, *such that* $u_n \uparrow f$ *pointwise convergent and we define*

$$\int_\Omega f \,\mathrm{d}\mu := \lim_{n\to\infty} \int_\Omega u_n \,\mathrm{d}\mu\,.$$

   It can be shown that this definition is independent of the choice of sequence $(u_n)_{n\in\mathbb{N}}$.

4. *Otherwise decompose* $f = f_{\geq 0} - f_{\leq 0}$, *where*

$$f_{\geq 0}(\omega) := \max\{f(\omega), 0\} \textit{ and } f_{\leq 0} := -\min\{f(\omega), 0\}\,.$$

   We call f μ*-integrable if*

$$\int_\Omega f_{\geq 0} \,\mathrm{d}\mu < \infty \textit{ and } \int_\Omega f_{\leq 0} \,\mathrm{d}\mu < \infty\,.$$

   In that case, we define

$$\int_\Omega f \,\mathrm{d}\mu := \int_\Omega f_{\geq 0} \,\mathrm{d}\mu - \int_\Omega f_{\leq 0} \,\mathrm{d}\mu\,.$$

Remark 1.5 *If the respective space referred to is clear by context, we also write* $\int f \,\mathrm{d}\mu := \int_\Omega f \,\mathrm{d}\mu$ *and call it the* μ*-integral of* f.

Corollary 1.4.1 (Properties of the integral)
Let $f, g : \Omega \to \overline{\mathbb{R}}$ *be* μ*-integrable, then*

1. *for all* $a, b \in \mathbb{R}$*, we have* $\int_\Omega af + bg \,\mathrm{d}\mu = a\int_\Omega f \,\mathrm{d}\mu + b\int_\Omega g \,\mathrm{d}\mu$,
2. *from* $f \leq g$ *follows* $\int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega g \,\mathrm{d}\mu$ *and*
3. $|\int_\Omega f \,\mathrm{d}\mu| \leq \int_\Omega |f| \,\mathrm{d}\mu$.

Definition 1.8 (Almost everywhere, Almost surely)
Let $(\Omega, \mathcal{A}, \mu)$ *be a measure space and let* E_ω *be a proposition for every* $\omega \in \Omega$*. We say* E *holds* μ*-almost everywhere (*μ*-a.e.) if there exists* $N \in \mathcal{A}$ *with* $\mu(N) = 0$ *and for all* $\omega \in \Omega \setminus N$ *we have* E_ω *. If* $\mu = \mathbb{P}$ *is a probability measure, we say that* E *holds* $\mathbb{P}$*-almost surely (*$\mathbb{P}$*-a.s.).*

Definition 1.9 (L^p **space)**
Let $(\Omega, \mathcal{A}, \mu)$ *be a measure space and let* $p \in [1, \infty)$*. The space* $L^p(\Omega, \mathcal{A}, \mu)$ *consists of equivalence classes of measurable functions* $f : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *with*

$$\int_\Omega |f|^p \,\mathrm{d}\mu < \infty .$$

In this space, two measurable functions with the above properties are considered equivalent if they are equal μ*-almost everywhere.*

Remark 1.6 *With respect to the previous definition, we define a norm on* $L^p(\Omega, \mathcal{A}, \mu)$ *as*

$$\|f\|_p := \left(\int_\Omega |f|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} .$$

Assuming instead that for $f : (\Omega, \mathcal{A}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *measurable,*

$$\|f\|_\infty := \sup\left\{ M \in [0, \infty) \,\middle|\, \mu(\{\omega \in \Omega \mid |f(\omega)| > M\}) = 0 \right\}$$

exists. We denote the set consisting of equivalence classes of functions fulfilling the previous property which are equal μ*-almost everywhere with* $L^\infty(\Omega, \mathcal{A}, \mu)$.

Definition 1.10 (Expectation of a random variable)
Let $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space and* $X : \Omega \to \mathbb{R}$ *a random variable. We say the expectation of* X *exists if* X *is* $\mathbb{P}$*-integrable. In that case, we define the expectation of* X *as*

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

1.4 Important Integral Theorems

Theorem 1.5 *Let* $(\Omega, \mathcal{A}, \mu)$ *be a measure space and* $f, g : \Omega \to \overline{\mathbb{R}}$ *be measurable with* $f = g$ *μ-almost everywhere, then* ▷ 15.10.2019

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu .$$

Theorem 1.6 (Markov's inequality)
Let $(\Omega, \mathcal{A}, \mu)$ *be a measure space,* $f : \Omega \to \mathbb{R}_{\geq 0}$ *be* μ*-integrable and* $t > 0$*, then*

$$\mu(f^{-1}[t, \infty)) \leq \frac{1}{t}\int_\Omega f \,\mathrm{d}\mu .$$

Corollary 1.6.1 *Let $f : \Omega \to \overline{\mathbb{R}}$ be μ-$\mathcal{B}(\overline{\mathbb{R}})$-measurable.*

1. *If $f \geq 0$, then $\int f \, \mathrm{d}\mu = 0 \Leftrightarrow f = 0$ μ-almost everywhere.*

2. *If $\int |f| \, \mathrm{d}\mu < \infty$, then $|f| < \infty$ μ-almost everywhere.*

Remark 1.7 *The former can be formulated in probabilistic terms. If X is a random variable with values in $\overline{\mathbb{R}}$, the following holds.*

1. *If $X \geq 0$, then $\mathbb{E}X = 0 \Leftrightarrow X = 0$ $\mathbb{P}$-almost surely.*

2. *If $\mathbb{E}(|X|) < \infty$, then $|X| < \infty$ $\mathbb{P}$-almost surely.*

Theorem 1.7 (Monotone convergence)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f_1, f_2, \ldots : \Omega \to [0, \infty]$ be a sequence of measurable functions. If $(f_n)_{n\in\mathbb{N}}$ is also pointwise non-decreasing μ-almost everywhere, then the integral exists and

$$\int \lim_{n\to\infty} f_n \, \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu \, .$$

Theorem 1.8 (Fatou's lemma)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f_1, f_2, \ldots : \Omega \to [0, \infty]$ be a sequence of measurable functions, then

$$\int \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu \, .$$

Theorem 1.9 (Lebesgue dominated convergence theorem)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f_1, f_2, \ldots : \Omega \to \overline{\mathbb{R}}$ be a sequence of measurable functions with pointwise $\lim_{n\to\infty} f_n = f$ μ-almost everywhere. If there exists a μ-integrable function $g : \Omega \to [0, \infty]$ with

$$|f_n| \leq g, \ \mu\text{-almost everywhere, for all } n \in \mathbb{N} \, ,$$

then f is μ-integrable and

$$\int f \, \mathrm{d}\mu = \lim_{n\to\infty} \int f_n \, \mathrm{d}\mu \, .$$

The Principle of Measure Theoretic Induction

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. If the goal is to show that an integral property E_f holds for all measurable functions $f : \Omega \to \overline{\mathbb{R}}$, one may proceed as follows.

1. Prove that E_f holds for non-negative step functions, i.e. $f \in \mathcal{E}$ and in particular for indicators $\mathbb{1}_A, A \in \mathcal{A}$.

2. Use the Lebesgue monotone convergence theorem to show that E_f holds for all non-negative measurable functions f.

3. Show that the property E_f holds for measurable functions f, by decomposing $f = f_{\geq 0} - f_{\leq 0}$.

Theorem 1.10 (Integration w.r.t. image measures)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $(\Omega', \mathcal{A}')$ a measurable space, $f : \Omega \to \Omega'$ and $h : \Omega' \to \overline{\mathbb{R}}$ both measurable. We observe the following properties.

1. *If $h \geq 0$, then*
$$\int_{\Omega'} h \, \mathrm{d}\mu^f = \int_{\Omega} h \circ f \, \mathrm{d}\mu \, .$$

2. *h is μ^f-integrable $\Leftrightarrow$ $h \circ f$ is μ-integrable. The formula above holds as well.*

1.5 Measures with Density

Definition 1.11 (Integration over subsets) ▷ 17.10.2019
Let $f : \Omega \to \overline{\mathbb{R}}$ be $\mathcal{A}$-$\mathcal{B}(\overline{\mathbb{R}})$-measurable and f non-negative or μ-integrable. For $A \in \mathcal{A}$, we define

$$\int_A f \, \mathrm{d}\mu := \int_\Omega f \cdot \mathbb{1}_A \, \mathrm{d}\mu$$

as the μ-integral of f over A.

Remark 1.8 *One needs to show that the former is well-defined, by showing, that $f \cdot \mathbb{1}_A$ is μ-integrable.*

Theorem 1.11 *In the above situation if $f \geq 0$ μ-almost everywhere,*

$$A \in \mathcal{A} \longmapsto f\mu(A) := \int_A f \, \mathrm{d}\mu$$

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

Definition 1.12 (Measures with density)
We call $f\mu$ the measure with density f with respect to μ.

Remark 1.9 *If $\Omega = \mathbb{R}^k$ for some $k \in \mathbb{N}$ and μ is the Lebesgue measure, f is called Lebesgue density.*

Theorem 1.12 (Uniqueness of densities)
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $f, g : \Omega \to \overline{\mathbb{R}}$ measurable and $f, g \geq 0$.

1. *If $f = g$ μ-almost everywhere, then $f\mu = g\mu$.*
2. *If f or g is μ-integrable, then the former holds in both directions.*

Theorem 1.13 *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $f\mu$ a measure with density.*

1. *If $\varphi : \Omega \to [0, \infty]$ measurable, then*

$$\int_\Omega \varphi \, \mathrm{d}f\mu = \int_\Omega \varphi \cdot f \, \mathrm{d}\mu \, .$$

2. *If $\varphi : \Omega \to \overline{\mathbb{R}}$ is measurable, then*

$$\varphi \text{ is } f\mu\text{-integrable} \iff \varphi \cdot f \text{ is } \mu\text{-integrable} \, .$$

If the latter holds, then the statement in 1. holds as well.

Definition 1.13 (Absolute continuity of measures)
Let $(\Omega, \mathcal{A})$ be a measurable space and μ, ν be associated measures. We say ν is absolutely continuous with respect to μ if for all $A \in \mathcal{A}$

$$\mu(A) = 0 \implies \nu(A) = 0$$

and we write $\nu \ll \mu$.

Theorem 1.14 (Radon-Nikodym)
Let $(\Omega, \mathcal{A})$ be a measurable space and μ, ν be associated measures. If μ is σ-finite, then

$$\nu \ll \mu \iff \nu \text{ has a density with respect to } \mu \, .$$

Remark 1.10

- *σ-finiteness of μ is necessary for the existence of a density with respect to ν.*
- *The density is μ-almost everywhere uniquely determined. We denote it by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.*
- *In the previous lecture we called a random variable continuous if $\mathbb{P}^X \ll \lambda$.*

Definition 1.14 (Singular measures)
Let $(\Omega, \mathcal{A})$ be a measurable space and μ, ν be associated measures.
We say μ and ν are singular if there exists $A \in \mathcal{A}$ such that $\mu(A) = 0$ and $\nu(\Omega \setminus A) = 0$. In this case we write $\mu \perp \nu$.

Theorem 1.15 (Lebesgue decomposition)
Let μ and ν be measures on $(\Omega, \mathcal{A})$ with ν being σ-finite. Then there are uniquely determined measures ν_a and ν_s with

$$\nu_a \ll \mu\,,\ \nu_s \perp \mu \text{ and } \nu_a + \nu_s = \nu\,.$$

Definition 1.15 (Product measure)
Let $(\Omega_j, \mathcal{A}_j, \mu_j)_{j=1,\dots,n}$ be measure spaces. Define

$$\Omega := \Omega_1 \times \ldots \times \Omega_n \text{ and } \mathcal{A} := \bigotimes_{j=1}^{n} A_j := \sigma\left(\{A_1 \times \ldots \times A_n \mid A_j \in \mathcal{A}_j\}\right)\,.$$

We call a measure μ on $(\Omega, \mathcal{A})$ product measure if

$$\mu(A_1 \times \ldots \times A_n) = \prod_{j=1}^{n} \mu_j(A_j)\,, \text{ for all } A_j \in \mathcal{A}_j\,.$$

Theorem 1.16 (Product measure)
In the former situation if $(\Omega_j, \mathcal{A}_j, \mu_j)_{j=1,\dots,n}$ are σ-finite measure spaces, then the product measure is unique.

Remark 1.11 *We denote the product measure by $\mu = \bigotimes_{i=1}^{n} \mu_i = \mu_1 \otimes \ldots \otimes \mu_n$ and call $(\Omega, \mathcal{A}, \mu)$ the product of $(\Omega_j, \mathcal{A}_j, \mu_j)_{j=1,\dots,n}$.*

Theorem 1.17 (Tonelli)
In the above situation, let $(\Omega_1, \mathcal{A}_1, \mu_1), (\Omega_2, \mathcal{A}_2, \mu_2)$ be two σ-finite measure spaces and $f : \Omega \to [0, \infty]$ be a measurable function, then

$$\begin{aligned}\int_{\Omega} f \,\mathrm{d}(\mu_1 \otimes \mu_2) &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2)\,\mathrm{d}\mu_1(\omega_1) \right) \mathrm{d}\mu_2(\omega_2) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\,\mathrm{d}\mu_2(\omega_2) \right) \mathrm{d}\mu_1(\omega_1)\,.\end{aligned}$$

Theorem 1.18 (Fubini)
In the above situation let $(\Omega_1, \mathcal{A}_1, \mu_1), (\Omega_2, \mathcal{A}_2, \mu_2)$ be two σ-finite measure spaces and $f : \Omega \to \overline{\mathbb{R}}$ be $\mu_1 \otimes \mu_2$-integrable, then

1. *$f(\omega_1, \cdot)$ is μ_2-integrable for μ_1-almost every $\omega_1 \in \Omega_1$,*
2. *$f(\cdot, \omega_2)$ is μ_1-integrable for μ_2-almost every $\omega_2 \in \Omega_2$,*
3. *the μ_1-almost everywhere defined functions*

$$\omega_1 \longmapsto \int_{\Omega_2} f(\omega_1, \omega_2)\,\mathrm{d}\mu_2(\omega_2)$$

   is μ_1-integrable,

4. the μ_2-almost everywhere defined functions

$$\omega_2 \longmapsto \int_{\Omega_1} f(\omega_1, \omega_2)\, \mathrm{d}\mu_1(\omega_1)$$

is μ_2-integrable and

5.

$$\begin{aligned}\int_{\Omega} f \,\mathrm{d}(\mu_1 \otimes \mu_2) &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2)\, \mathrm{d}\mu_1(\omega_1) \right) \mathrm{d}\mu_2(\omega_2) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\, \mathrm{d}\mu_2(\omega_2) \right) \mathrm{d}\mu_1(\omega_1)\,.\end{aligned}$$

2 Random Variables

2.1 Independence

Definition 2.1 (Independence) ▷ 22.10.2019

We consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let I be an arbitrary index set. In the following we define three notions of independence.

1. *Independence of sets $A_i \in \mathcal{A}$ with respect to $\mathbb{P}$:*

$$(A_i)_{i\in I} \perp\!\!\!\perp :\Longleftrightarrow \forall J \subseteq I, |J| < \infty : \quad \mathbb{P}(\cap_{j\in J} A_j) = \prod_{j\in J} \mathbb{P}(A_j)$$

2. *Independence of families of events $\mathcal{F}_i \subseteq \mathcal{A}$ with respect to $\mathbb{P}$:*

$$(\mathcal{F}_i)_{i\in I} \perp\!\!\!\perp :\Longleftrightarrow \forall J \subseteq I, |J| < \infty : \quad \mathbb{P}(\cap_{j\in J} A_j) = \prod_{j\in J} \mathbb{P}(A_j)\,,\ \forall A_j \in \mathcal{F}_j$$

3. *Independence of random variables $X_i : (\Omega, \mathcal{A}) \to (E_i, \mathcal{E}_i)$ with respect to $\mathbb{P}$:*

$$(X_i)_{i\in I} \perp\!\!\!\perp :\Longleftrightarrow \big(\mathcal{I}(X_i) = X_i^{-1}(\mathcal{E}_i)\big)_{i\in I} \perp\!\!\!\perp$$

That is, the initial σ-algebras generated by $(X_i)_{i\in I}$ are independent with respect to $\mathbb{P}$ as families of events.

Lemma 2.1 *Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we have*

$$(A_i)_{i\in I} \in \mathcal{A} \perp\!\!\!\perp \Longleftrightarrow (\sigma(A_i))_{i\in I} \subseteq \mathcal{A} \perp\!\!\!\perp \Longleftrightarrow (\mathbb{1}_{A_i})_{i\in I} \text{ measurable} \perp\!\!\!\perp\ .$$

Theorem 2.1 *Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let $(\mathcal{F})_{i\in I} \subseteq \mathcal{A}$ be $\cap$-stable, then*

$$(\mathcal{F}_i)_{i\in I} \perp\!\!\!\perp \Longleftrightarrow (\sigma(\mathcal{F}_i))_{i\in I} \perp\!\!\!\perp\ .$$

Lemma 2.2 (1^{st} block lemma)

Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, let

$$\mathcal{F}_{i,k}, \subseteq \mathcal{A}\,,\ 1 \le k \le n(i)\,,\ 1 \le i \le m \in \mathbb{N}\,,$$

be independent $\cap$-stable families of events. Then the following σ-fields

$$\mathcal{G}_i := \sigma(\mathcal{F}_{i,1} \cup \ldots \cup \mathcal{F}_{i,n(i)})\,,\ 1 \le i \le m\,,$$

are independent.

Lemma 2.3 (2^{nd} block lemma)

Let

$$X_{i,k} : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})\,,\ 1 \le k \le n(i)\,,\ 1 \le i \le m \in \mathbb{N}\,,$$

be independent random variables and $f_i : E^{n(i)} \to \mathbb{R}$ measurable functions. Then the random variables

$$z_i = f_i(X_{i1}, \ldots, X_{in(i)})$$

are independent.

Theorem 2.2 *Let $X_1, \ldots, X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ be random variables. Then the following statements are equivalent.* ▷ 24.10.2019

1. *$X_1, \ldots, X_n$ are independent.*
2. *$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)\,, \quad \forall A_1, \ldots, A_n \in \mathcal{E}$*

3. $\mathbb{P}^{(X_1,\dots,X_n)} = \mathbb{P}^{X_1} \otimes \dots \otimes \mathbb{P}^{X_n}$
   (The joint distribution of the random vector is the product of its marginal distributions.)

Corollary 2.2.1 *Let $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathbb{R}^d$ be independent random variables and $h : \mathbb{R}^{2d} \to \mathbb{R}$ measurable. If $h \geq 0$ or if $h(X,Y)$ is $\mathbb{P}$-integrable, then*

$$\begin{aligned}\mathbb{E}h(X,Y) &= \int\int h(x,y)\,\mathbb{P}^X(\mathrm{d}x)\,\mathbb{P}^Y(\mathrm{d}y)\\ &= \mathbb{E}\int h(x,Y)\,\mathbb{P}^X(\mathrm{d}x)\\ &= \mathbb{E}\int h(X,y)\,\mathbb{P}^y(\mathrm{d}y)\,.\end{aligned}$$

Corollary 2.2.2 *Let $X, Y : \Omega \to \mathbb{R}^d$ be independent random variables and $f, g : \mathbb{R}^d \to \mathbb{R}$ be measurable functions such that $f, g \geq 0$ or $\mathbb{E}|f(X)|, \mathbb{E}|g(Y)| < \infty$, then*

$$\mathbb{E}(f(X)\cdot g(Y)) = (\mathbb{E}f(X))\cdot(\mathbb{E}g(Y))\,.$$

Corollary 2.2.3 *Let $X_1, \dots, X_n : \Omega \to \mathbb{R}^d$ be independent random variables and $f_1, \dots, f_n : \mathbb{R}^d \to \mathbb{R}$ measurable functions such that $f_1, \dots, f_n \geq 0$ or $\mathbb{E}|f_i(X_i)| < \infty$, for all $i = 1, \dots, n$, then*

$$\mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}\left(f_i(X_i)\right)\,.$$

Definition 2.2 (Convolution) ▷ 29.10.2019
Let μ, ν be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then the following measure defines the convolution of μ and ν

$$(\mu * \nu)(B) := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbb{1}_B(x+y)\,\mu(\mathrm{d}x)\nu(\mathrm{d}y)\,, \quad B \in \mathcal{B}(\mathbb{R}^d)\,.$$

Theorem 2.3 *Let $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be independent random variables, then*

1. $\mathbb{P}^{X+Y} = \mathbb{P}^X * \mathbb{P}^Y$ *and*
2. *if $\mathbb{P}^X \ll \lambda^d$ with density f and $\mathbb{P}^Y \ll \lambda^d$ with density g, then $\mathbb{P}^{X+Y} \ll \lambda^d$ with density*

$$h(z) = \int_{\mathbb{R}^d} f(z-x)\cdot g(x)\,\lambda^d(\mathrm{d}x)\,, \quad z \in \mathbb{R}^d\,.$$

Remark 2.1 *One can regard $\mu * \nu$ as the image measure of $\mu \otimes \nu$ under the map $(x,y) \mapsto x+y$.*

Definition 2.3 (Variance, Covariance, Uncorrelated)
Let $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be random variables and let X^2, Y^2 be $\mathbb{P}$-integrable. We define

1. *the variance of X as $\mathbb{V}(X) := \mathbb{E}\left((X - \mathbb{E}X)^2\right)$ and*
2. *the covariance of X and Y as $\mathrm{Cov}(X,Y) := \mathbb{E}\left((X-\mathbb{E}X)(Y-\mathbb{E}Y)\right)$.*

If $\mathrm{Cov}(X,Y) = 0$ we call X and Y uncorrelated.

Lemma 2.4 *Let $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be random variables, let X^2, Y^2 be $\mathbb{P}$-integrable and $a, b \in \mathbb{R}$, then*

1. $\mathbb{V}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2\,,$
2. $a \mapsto \mathbb{E}(X-a)^2$ *is minimized for* $a = \mathbb{E}X\,,$
3. $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)\,,$

4. $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$ *and* $\mathrm{Cov}(X,X) = \mathbb{V}(X)$,

5. $\mathbb{V}(X+Y) - \mathbb{V}(X-Y) = 4\mathrm{Cov}(X,Y)$ *and*

6. $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\mathrm{Cov}(X,Y)$.

Theorem 2.4 (Bianaymé's identity)
Let $X_1, \ldots, X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be pairwise uncorrelated random variables and let* $X_1^2, \ldots X_n^2$ *be* $\mathbb{P}$*-integrable, then*

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) \; .$$

2.2 Construction of Independent Random Variables

Problem 1 Given a measurable space $(E, \mathcal{E})$ and probability measure $\mathbb{P}$, is there always a random variable X taking values in E and having distribution $\mathbb{P}$? ▷ 31.10.2019

Problem 2 Given a measurable space $(E, \mathcal{E})$ and probability measures $\mathbb{P}_1, \mathbb{P}_2$, are there always random variables $X \sim \mathbb{P}_1, Y \sim \mathbb{P}_2$ taking values in E with $X \perp\!\!\!\perp Y$?

Problem 3 Given $(E, \mathcal{E})$ and probability measures $(\mathbb{P}_i)_{i \in I}$, where I is an arbitrary index set, are there always random variables $X_i \sim \mathbb{P}_i$ taking values in E with $(X_i)_{i \in I} \perp\!\!\!\perp$?

Solution to problem 1: Let $(E, \mathcal{E}, \mathbb{P})$ be a measure space. Define a random variable

$$X : (E, \mathcal{E}, \mathbb{P}) \to (E, \mathcal{E}), \omega \longmapsto \omega \quad \text{that is} \quad X := \mathrm{id} \, .$$

It follows that $\mathbb{P}^X = \mathbb{P} \circ \mathrm{id}^{-1} = \mathbb{P}$.

Solution to problem 2: By applying the solution to problem 1, we can find probability spaces $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$, $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ and random variables

$$X : (\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \to (E, \mathcal{E}) \quad \text{with} \quad \mathbb{P}_1^X = \mathbb{P}_1 \, ,$$
$$Y : (\Omega_2, \mathcal{A}_2, \mathbb{P}_2) \to (E, \mathcal{E}) \quad \text{with} \quad \mathbb{P}_2^Y = \mathbb{P}_2 \, .$$

Define $\Omega := \Omega_1 \times \Omega_2, \mathcal{A} := \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P} := \mathbb{P}_1 \otimes \mathbb{P}_2$ and new random variables

$$\tilde{X} : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), \omega = (\omega_1, \omega_2) \mapsto X(\omega_1) \, ,$$
$$\tilde{Y} : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), \omega = (\omega_1, \omega_2) \mapsto Y(\omega_2) \, .$$

They are well defined, i.e. measurable, as X, Y are measurable. To show independence and by theorem 2.2 it is sufficient to prove that for $B_1, B_2 \in \mathcal{E}$

$$\mathbb{P}(\tilde{X} \in B_1, \tilde{Y} \in B_2) = \mathbb{P}_1 \otimes \mathbb{P}_2(\underbrace{\{\tilde{X} \in B_1\} \cap \{\tilde{Y} \in B_2\}}_{=\{X \in B_1\} \cap \{Y \in B_2\}}) = \mathbb{P}_1(\underbrace{X \in B_1}_{=\{X \in B_1\}})\mathbb{P}_2(\underbrace{Y \in B_2}_{=\{Y \in B_2\}}) \, .$$

Therefore $\mathbb{P}^{\tilde{X}} = \mathbb{P}_1^X$, $\mathbb{P}^{\tilde{Y}} = \mathbb{P}_2^Y$ and $\tilde{X} \perp\!\!\!\perp \tilde{Y}$.

Definition 2.4 (Generalized product)
Let $I \neq \emptyset$ *be an index set and* $(\Omega_i, \mathcal{A}, \mathbb{P}_i)_{i \in I}$ *probability spaces. For* $\emptyset \neq K \subseteq I$ *define*

$$\Omega_K := \bigtimes_{k \in K} \Omega_k := \left\{ f : K \to \bigcup_{k \in K} \Omega_k \mid f(k) \in \Omega_k \, , \; \textit{for all } k \in K \right\}$$

as the generalized product of $(\Omega_k)_{k \in K}$*. Further we define*

· *the* i*-th coordinate of* $f \in \Omega_K$ *as* $f(i), i \in K$,

· *the i-th coordinate projection* $\pi_i : \Omega_I \to \Omega_i, f \mapsto f(i), i \in I$,

· *the restricted i-th coordinate projection* $\pi_i^K : \Omega_K \to \Omega_i, f \mapsto f(i),\ K \subseteq I, i \in K$,

· *the restriction to* Ω_K *as* $\pi_K^J : \Omega_J \to \Omega_K, f \mapsto f_{|K}, K \subseteq J,\ \pi_K := \pi_K^I$ *and*

· *the set of non-empty finite index subsets of* I *as* $\kappa := \kappa(I) := \{K \subseteq I \mid 0 < |K| < \infty\}$.

Definition 2.5 (Generalized product σ-field)
Let $I \neq \emptyset$ *be an index set and let* $(\Omega_i, \mathcal{A}_i)_{i\in I}$ *be measurable spaces, then*

$$\mathcal{A}_I := \bigotimes_{i\in I} \mathcal{A}_i := \mathcal{I}\left((\pi_i)_{i\in I}\right) ,$$

that is, the generalized product σ*-field of* $(\Omega_i, \mathcal{A}_i)_{i\in I}$ *is defined as the smallest* σ*-field on* Ω_I *such that all coordinate projections* $\pi_i : \Omega_I \to (\Omega_i, \mathcal{A}_i)$ *are measurable, i.e. their initial* σ*-algebra.*

Remark 2.2 *The projection maps* $\pi_H : \Omega_I \to \Omega_H, H \in \kappa$ *are* $\mathcal{A}_I$*-*$\mathcal{A}_H$*-measurable. It is possible to verify using the generator of* $\mathcal{A}_H$ *given by so called cylinder sets of the form*

$$\bigcap_{i\in H} (\pi_i^H)^{-1}(A_i) , \quad A_i \in \mathcal{A}_i \text{ for all } i \in H .$$

Theorem 2.5 (Kolmogorov existence theorem)
Let $I \neq \emptyset$ *be an index set and* $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)_{i\in I}$ *probability spaces. Then there exists a unique probability measure*

$$\mathbb{P}_I = \bigotimes_{i\in I} \mathbb{P}_i \quad \text{on} \quad (\Omega_I, \mathcal{A}_I)$$

with the property that for all $H \in \kappa$

$$\mathbb{P}_I \circ \pi_H^{-1} = \mathbb{P}_I^{\pi_H} = \bigotimes_{i\in H} \mathbb{P}_i .$$

That is, it exists a unique measure where every finite restriction has an image measure that is equal to the finite product measure.

Definition 2.6 (Generalized product measure, Generalized Product)
The measure $\mathbb{P}_I$ *defined above is called the generalized product measure on* $(\Omega_I, \mathcal{A}_I)$ *and the measure space* $(\Omega_I, \mathcal{A}_I, \mathbb{P}_I)$ *is called the generalized product of* $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)_{i\in I}$.

Corollary 2.5.1 *Let* $I \neq \emptyset$ *be an index set and let* $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), i \in I$ *be random variables, then*

$$X_I : (\Omega, \mathcal{A}) \longrightarrow (E_I, \mathcal{E}_I),\ \omega \longmapsto (i \mapsto X_i(\omega))_{i\in I} \quad \text{is measurable}$$

and

$$(X_i)_{i\in I} \ \perp\!\!\!\perp \quad \Longleftrightarrow \quad \mathbb{P}^{X_I} = \bigotimes_{i\in I} \mathbb{P}^{X_i} ,$$

where $\mathbb{P}^{X_I}$ *and* $\bigotimes_{i\in I} \mathbb{P}^{X_i}$ *are both measures on* $(E_I, \mathcal{E}_I)$.

Solution to problem 3: Let $I \neq \emptyset$ be an index set, let $(E, \mathcal{E})$ be a measure space and $(\mathbb{P}_i)_{i\in I}$ corresponding measures. For each $j \in I$ define a random variable

$$X_j : \left(E_I, \mathcal{E}_I, \bigotimes_{i\in I} \mathbb{P}_i\right) \longrightarrow (E, \mathcal{E}) ,\ (\omega : I \to E) \longmapsto \omega(j) ,$$

which is equal the projection map π_j. It exists with respect to its unique measure due to the Kolmogorov existence theorem. To these we apply the previous corollary as follows.

As per the previous corollary define

$$X_I : (E_I, \mathcal{E}_I) \longrightarrow (E_I, \mathcal{E}_I),\ (\omega : I \to E) \longmapsto (j \mapsto X_j(\omega)) = (\omega : I \to E)\,,$$

which is equal to the identity map. Therefore, clearly for all $j \in I$

$$\mathbb{P}^{X_j} = \left(\bigotimes_{i\in I} \mathbb{P}_i\right) \circ \pi_j^{-1} \circ = \mathbb{P}_j \quad \text{and} \quad \mathbb{P}^{X_I} = \left(\bigotimes_{i\in I} \mathbb{P}_i\right) \circ \mathrm{id}_{E_I} = \bigotimes_{i\in I} \mathbb{P}_i = \bigotimes_{i\in I} \mathbb{P}^{X_i}\,.$$

That is, $X_j \sim \mathbb{P}_j$ for all $j \in I$ and $(X_j)_{j\in I} \perp\!\!\!\perp$ as they fulfill the right of the equality in the previous corollary.

Lemma 2.5 (3rd block lemma) ▷ 05.11.2019
Let $I \neq \emptyset$ be an index set, let $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\Omega_i, \mathcal{A}_i), i \in I$ be independent random variables, let $I = \bigcup_{k\in K} I_k$ be a partition of I and

$$f_k : \left(\bigtimes_{i\in I_k} \Omega_i, \bigotimes_{i\in I_k} \mathcal{A}\right) \longrightarrow (E_k, \mathcal{E}_k), k \in K$$

are measurable functions. Then $(f_k((X_i)_{i\in I_k}))_{k\in K}$ are independent.

2.3 0-1 laws

Lemma 2.6 (Borel-Cantelli lemma)
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}$, then

1. $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty \implies \mathbb{P}(\limsup_{n\to\infty} A_n) = 0$ *and*
2. $(A_n)_{n\in\mathbb{N}}$ *are pairwise independent and* $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty \implies \mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Definition 2.7 (Terminal σ-algebra)

1. *Let $(\Omega, \mathcal{A})$ a measurable space and $(\mathcal{A}_n)_{n\in\mathbb{N}} \subseteq \mathcal{A}$ an a sequence of subsets. We define*

$$\tau_k\left((\mathcal{A}_n)_{n\in\mathbb{N}}\right) := \sigma\left(\bigcup_{m=k}^{\infty} \mathcal{A}_m\right).$$

We call

$$\tau_\infty\left((\mathcal{A}_n)_{n\in\mathbb{N}}\right) := \bigcap_{k=1}^{\infty} \tau_k\left((\mathcal{A}_n)_{n\in\mathbb{N}}\right)$$

the terminal σ-algebra associated with $(\mathcal{A}_n)_{n\in\mathbb{N}}$.

2. *If we are instead given measurable functions $X_n : (\Omega, \mathcal{A}) \to (E, \mathcal{E}), n \in \mathbb{N}$, we define*

$$\tau_k\left((X_n)_{n\in\mathbb{N}}\right) := \mathcal{I}\left((X_m)_{m=k}^{\infty}\right)$$

and call

$$\tau_\infty\left((X_n)_{n\in\mathbb{N}}\right) := \bigcap_{k=1}^{\infty} \tau_k\left((X_n)_{n\in\mathbb{N}}\right)$$

the terminal σ-algebra associated with $(X_n)_{n\in\mathbb{N}}$.

Remark 2.3 *Setting $\mathcal{A}_n := X_n^{-1}(\mathcal{E}), n \in \mathbb{N}$, one could equivalently define $\tau_k((X_n)_{n\in\mathbb{N}}) := \tau_k((\mathcal{A}_n)_{n\in\mathbb{N}})$.*

Theorem 2.6 (Kolmogorov's 0-1-law)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}$ be independent random variables and let $\tau_\infty((X_n)_{n\in\mathbb{N}})$ be their associated terminal σ-algebra, then

$$A \in \tau_\infty\left((X_n)_{n\in\mathbb{N}}\right) \implies \mathbb{P}(A) \in \{0, 1\}.$$

Corollary 2.6.1 (Borel's 0-1-law) *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(A_n)_{n\in\mathbb{N}} \in \mathcal{A}$ be independent, then $\mathbb{P}(\limsup_{n\to\infty} A_n) \in \{0, 1\}$.*

Corollary 2.6.2 *Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}$ be independent random variables, let $\tau_\infty((X_n)_{n\in\mathbb{N}})$ be their associated terminal σ-algebra and let $Y : \Omega \to \overline{\mathbb{R}}$ be a $\tau_\infty((X_n)_{n\in\mathbb{N}})$-$\mathcal{B}(\overline{\mathbb{R}})$-measurable random variable. Then Y is constant $\mathbb{P}$-almost surely.* ▷ 07.11.2019

Definition 2.8 (Finite permutation, Symmetric event) ▷ 12.11.2019

- *We call a map $\psi : \mathbb{N} \to \mathbb{N}$ such that $\{k \in \mathbb{N} \mid \psi(k) \neq k\} < \infty$ a finite permutation and $\Psi := \{\psi : \mathbb{N} \to \mathbb{N} \mid \psi$ is a finite permutation$\}$.*
- *Let*

$$T_\psi : (E_\mathbb{N}, \mathcal{E}_\mathbb{N}) \longrightarrow (E_\mathbb{N}, \mathcal{E}_\mathbb{N}), \ (x_n)_{n\in\mathbb{N}} \longmapsto (x_{\psi(n)})_{n\in\mathbb{N}}, \psi \in \Psi,$$

we call an event $A \in \mathcal{E}_\mathbb{N}$ symmetric

$$:\iff T_\psi(A) = A, \ \forall \psi \in \Psi$$

and $\mathcal{E}_{\text{sym}} := \{A \in \mathcal{E}_\mathbb{N} \mid A$ symmetric$\}$.

Lemma 2.7 *Let $(E_\mathbb{N}, \mathcal{E}_\mathbb{N})$ be a generalized product space, then the set of symmetric events $\mathcal{E}_{\text{sym}} \subseteq \mathcal{E}_\mathbb{N}$ is a σ-algebra over $E_\mathbb{N}$.* ▷ 14.11.2019

Lemma 2.8 *Let $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \ldots$ be σ-fields over Ω, let $A \in \mathcal{A} := \sigma(\bigcup_{i=1}^\infty \mathcal{A}_i)$ and let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{A})$. Then, there exists $A_i \in \mathcal{A}_i, \forall i \in \mathbb{N}$, such that*

$$\mathbb{P}((A \setminus A_i) \cup (A_i \setminus A)) \xrightarrow{i\to\infty} 0.$$

Theorem 2.7 (Hewitt-Savage 0-1-law)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}$ be a family independent and identically distributed random variables and define

$$X_\mathbb{N} : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E_\mathbb{N}, \mathcal{E}_\mathbb{N}), \ \omega \longmapsto (X_n(\omega))_{n\in\mathbb{N}},$$

with values in the associated generalized product. Then,

$$\mathbb{P}^{X_\mathbb{N}}(A) \in \{0, 1\}, \ \forall A \in \mathcal{E}_{\text{sym}}, \ \text{where } \mathcal{E}_{\text{sym}} \subseteq \mathcal{E}_\mathbb{N}.$$

That is, symmetric events fulfill a 0-1-law with respect to the image measure of the random variables $(X_n)_{n\in\mathbb{N}}$.

2.4 Strong Law of Large Numbers

Theorem 2.8 *Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})), n \in \mathbb{N}$ be independent and identically distributed random variables with $\mathbb{E}|X_1| = \infty$. Then for $S_n := X_1 + \ldots + X_n, n \in \mathbb{N}$, we have*

1. *$\mathbb{P}(\limsup_{n\to\infty}\{|X_n| \geq n\}) = 1$ and*
2. *$\mathbb{P}(\{\lim_{n\to\infty} \frac{S_n}{n}$ exists and is finite$\}) = 0$.*

Definition 2.9 (Almost sure convergence) ▷ 19.11.2019

Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ *be random variables, we write*

$$X_n \xrightarrow{\text{a.s.}} X :\Longleftrightarrow \lim_{n\to\infty} X_n = X \ \mathbb{P}\text{-almost surely} :\Longleftrightarrow \mathbb{P}\left(\left\{\lim_{n\to\infty} X_n = X\right\}\right) = 1$$

and say $(X_n)_{n\in\mathbb{N}}$ *converge* $\mathbb{P}$*-almost surely to* X*.*

Lemma 2.9 *Let* $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ *be random variables, then*

$$X_n \xrightarrow{\text{a.s.}} X \iff \mathbb{P}\left(\limsup_{n\to\infty} \{|X_n - X| > \varepsilon\}\right) = 0\,, \ \text{for all } \varepsilon > 0\,.$$

Lemma 2.10 (Cesàro's lemma)

Let $(a_n)_{n\in\mathbb{N}} \in (0,\infty)$ *such that* $a_n \uparrow \infty$ *and let* $(v_k)_{k\in\mathbb{N}} \in \mathbb{R}$ *be convergent with* $\lim_{k\to\infty} v_k = v_\infty$*. Then*

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} (a_k - a_{k-1}) v_k = v_\infty\,,\ a_0 := 0\,.$$

Lemma 2.11 *Let* $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})), n \in \mathbb{N}$ *be independent and identically distributed random variables with* $\mathbb{E}|X_1| < \infty$*. Define*

$$Y_n := X_n \cdot \mathbb{1}_{\{|X_n| \le n\}}\,,\ T_n := Y_1 + \ldots + Y_n\,.$$

It follows, that

1. $\lim_{n\to\infty} \mathbb{E}Y_n = \mathbb{E}X_1\,,$
2. $\mathbb{P}(\limsup_{n\to\infty}\{X_n \neq Y_n\}) = 0\,,$
3. $\sum_{n=1}^{\infty} \frac{\mathbb{V}Y_n}{n^2} \le 2\mathbb{E}|X_1|$ *and*
4. $\sum_{n=1}^{\infty} \frac{\mathbb{V}T_{\lfloor \alpha^n \rfloor}}{\lfloor \alpha^n \rfloor^2} \le \frac{2\alpha}{\alpha-1}\mathbb{E}|X_1|$ *for all* $\alpha > 1\,.$

Theorem 2.9 (Strong law of large numbers) ▷ 21.11.2019

Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})), n \in \mathbb{N}$ *be independent and identically distributed random variables with* $\mathbb{E}|X_1| < \infty$*. Then*

$$\frac{X_1 + \ldots + X_n}{n} \quad \xrightarrow{\text{a.s.}} \quad \mathbb{E}X_1\,.$$

Definition 2.10 (Empirical distribution function)

Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})), n \in \mathbb{N}$ *be independent and identically distributed random variables. Define the empirical distribution function associated with* $(X_n)_{n\in\mathbb{N}}$ *as*

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(X_i)\,,\ n \in \mathbb{N}\,.$$

Corollary 2.9.1 (Glivenko-Cantelli theorem)

Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})), n \in \mathbb{N}$ *be independent and identically distributed random variables with distribution function* F *and let* F_n *be the associated empirical distribution function. Then*

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| = 0 \quad \mathbb{P}\text{-almost surely.}$$

3 Characteristic Functions

Definition 3.1 (Characteristic function) ▷ 26.11.2019
Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be a random variable. We define

$$\varphi_X : \mathbb{R}^d \longrightarrow \mathbb{C}\,, t \longmapsto \mathbb{E}\left(e^{i\langle t,X\rangle}\right)$$

as the characteristic function of the random variable X.

Theorem 3.1 *Let $G \sim \mathcal{N}(0,1)$, then $\varphi_G(t) = e^{-\frac{t^2}{2}}, t \in \mathbb{R}$.*

Theorem 3.2 *Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable and $a, b \in \mathbb{R}$. Then*

$$\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)\,,\ t \in \mathbb{R}\,.$$

Corollary 3.2.1 $X \sim \mathcal{N}(\mu, \sigma^2) \implies \varphi_X(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}, t \in \mathbb{R}$.

Theorem 3.3 (Properties of characteristic functions)
Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be a random variable with characteristic function φ_X. Then

1. *φ_X is uniformly continuous,*
2. $\|\varphi_X\|_\infty \leq 1$,
3. *$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$ for all $t \in \mathbb{R}^d$,*
4. *$t \longmapsto \operatorname{Re}\varphi_X(t) = \mathbb{E}\cos\langle t, X\rangle$ is the characteristic function of $\varepsilon \cdot X$, where $\varepsilon : (\Omega, \mathcal{A}, \mathbb{P}) \to \{-1, 1\}$, $\mathbb{P}(\varepsilon = \pm 1) = \frac{1}{2}$ and $\varepsilon \perp\!\!\!\perp X$,*
5. *$|\varphi_X(t)|^2$ is the characteristic function of $X - \tilde{X}$, where $X, \tilde{X}$ are idenpendent and identically distributed,*
6. *let*
$$T_X = \Sigma X + m\,,\ \Sigma \in \mathbb{R}^{n\times d} \text{ and } m \in \mathbb{R}^n\,,$$
it follows that
$$\varphi_{T_X}(t) = e^{i\langle t,m\rangle} \cdot \varphi_X(\Sigma^{\mathrm{T}} t)\,,\ t \in \mathbb{R}^d\,,$$
7. *let $d = 1$ and $\mathbb{E}|X|^n < \infty$ for some $n \in \mathbb{N}$, it follows that φ_X is n-times continuously differentiable and*
$$\varphi_X^{(k)}(0) = i^k \cdot \mathbb{E}X^k\,,\ \text{for all } k \in \{0, \ldots, n\} \quad \text{and}$$
8. *let $d = 1$, $t_1, \ldots, t_n \in \mathbb{R}$ and $A_\varphi := (a_{ij})_{ij=1,\ldots,n}$, where $a_{ij} := \varphi_X(t_i - t_j)$, it follows that A_φ is Hermitian and positive semidefinite.*

Theorem 3.4 (Bodiner's theorem) ▷ 28.11.2019
Let $\varphi : \mathbb{R} \to \mathbb{C}$, then the following statements are equivalent.

1. *φ is the characteristic function of a $\mathbb{R}$-valued random variable.*
2. *φ is continuous, $\varphi(0) = 1$ and $A_\varphi := (a_{ij})_{ij=1,\ldots,n}$ is positive semidefinite, where $a_{ij} := \varphi_X(t_i - t_j)$, $t_1, \ldots, t_n \in \mathbb{R}$.*

Theorem 3.5 (Lévy's theorem)
Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable and $[a, b] \subset \mathbb{R}$, then

$$\frac{1}{2}\mathbb{P}(X = a) + \frac{1}{2}\mathbb{P}(X = b) + \mathbb{P}(a < X < b) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\,\mathrm{d}t\,.$$

Corollary 3.5.1 *Let* $X : (\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $Y : (\Omega_2, \mathcal{A}_2, \mathbb{P}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables with* $\varphi_X = \varphi_Y$, *then* $X \overset{\mathrm{d}}{=} Y$.

Corollary 3.5.2 (Kac's theorem)
Let $X, Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables, then*

$$X \perp\!\!\!\perp Y \iff \varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t)\,,\ s,t \in \mathbb{R}\,.$$

Theorem 3.6 (Lévy's inequality)
Let $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *be a random variable and* $r > 0$. *Then*

$$\mathbb{P}\left(\max_{1\le k\le d} |X_k| \ge r\right) \le 7\left(\frac{r}{2}\right)^d \int_{-\frac{1}{r}}^{\frac{1}{r}} \cdots \int_{-\frac{1}{r}}^{\frac{1}{r}} (1 - \operatorname{Re}\varphi_X(t))\,\mathrm{d}t_1 \ldots \mathrm{d}t_d\,.$$

4 Convergence of Random Variables

Definition 4.1 (Modes of convergence) ▷ 03.12.2019
Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables. We define the following modes of convergence.*

1. *Convergence* ***in distribution****:*

$$X_n \xrightarrow{\text{d}} X \;:\Longleftrightarrow\; \forall f \in \mathcal{C}_b(\mathbb{R}) : \lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$$

Let now $\Omega = \Omega_1 = \ldots = \Omega_n$ *,* $\mathcal{A} = \mathcal{A}_1 = \ldots = \mathcal{A}_n$ *and* $\mathbb{P} = \mathbb{P}_1 = \ldots = \mathbb{P}_n$ *.*

2. *Convergence* ***almost surely****:*

$$X_n \xrightarrow{\text{a.s.}} X \;:\Longleftrightarrow\; \mathbb{P}(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

3. *Convergence* ***in*** L^p***/in*** $\mathbf{p^{th}}$ ***mean****,* $1 < p \in \mathbb{R}$ *:*

$$X_n \xrightarrow{L^p} X \;:\Longleftrightarrow\; X, X_n \in L^p \text{ and } \lim_{n\to\infty} \|X_n - X\|_p = 0$$

4. *Convergence* ***in probability****:*

$$X_n \xrightarrow{\mathbb{P}} X \;:\Longleftrightarrow\; \forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

Remark 4.1 *For a real valued random variable* Z*, let* $\|Z\|_p := (\mathbb{E}|Z|^p)^{1/p}$ *,* $p \in \mathbb{N}$ *. Further, let* $\mathcal{C}_b(\mathbb{R})$ *denote the set of continuous and bounded functions on* $\mathbb{R}$ *to* $\mathbb{R}$*.*

Lemma 4.1 (Uniqueness of limits) *Let* $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*,* $Y : (E, \mathcal{E}, \mathbb{W}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables. It follows that*

1. $X_n \xrightarrow{\text{d}} X$ *and* $X_n \xrightarrow{\text{d}} Y \implies X \overset{d}{=} Y$ *.*

Let now $\Omega = E = \Omega_1 = \ldots = \Omega_n$ *,* $\mathcal{A} = \mathcal{E} = \mathcal{A}_1 = \ldots = \mathcal{A}_n$ *and* $\mathbb{P} = \mathbb{W} = \mathbb{P}_1 = \ldots = \mathbb{P}_n$ *. Then*

2. $X_n \xrightarrow{\text{a.s.}} X$ *and* $X_n \xrightarrow{\text{a.s.}} Y \implies X = Y$ $\mathbb{P}$*-almost surely,*

3. $X_n \xrightarrow{\mathbb{P}} X$ *and* $X_n \xrightarrow{\mathbb{P}} Y \implies X = Y$ $\mathbb{P}$*-almost surely,*

4. $X_n \xrightarrow{L^p} X$ *and* $X_n \xrightarrow{L^p} Y \implies X = Y$ $\mathbb{P}$*-almost surely.*

Theorem 4.1 (Relating the modes of convergence)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables. It follows that*

1. $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\text{d}} X$ *and* $f(X_n) \xrightarrow{L^1} f(X)$ *, for all* $f \in \mathcal{C}_b(\mathbb{R})$ *,*

2. $X_n \xrightarrow{L^p} X$ *for some* $p > 1 \implies X_n \xrightarrow{L^1} X$ *,*

3. $X_n \xrightarrow{L^1} X \implies X_n \xrightarrow{\mathbb{P}} X$ *and*

4. $X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\mathbb{P}} X$ *.*

Lemma 4.2 (Fast convergence) ▷ 05.12.2019
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables. Assume that* $X_n \xrightarrow{\mathbb{P}} X$ *and there exist* $(\varepsilon_n)_{n\in\mathbb{N}} \in \mathbb{R}\,,\ \varepsilon_n \downarrow 0$ *such that*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon_n) < \infty\,.$$

Then $X_n \xrightarrow{\text{a.s.}} X$ *.*

Corollary 4.1.1 *Let* $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables. Assume* $X_n \xrightarrow{\mathbb{P}} X$ *then there is a monotone sequence* $(n_k)_{k\in\mathbb{N}} \in \mathbb{N}$ *such that* $X_{n_k} \xrightarrow{\text{a.s.}} X$ *.*

Lemma 4.3 *Let* $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables, where* $X = b \in \mathbb{R}$ $\mathbb{P}$*-almost surely. Then*

$$X_n \xrightarrow{\mathbb{P}} X \iff X_n \xrightarrow{\text{d}} X\,.$$

Remark 4.2 *In general, the implications in the below diagram can not be reversed, which can be proved by finding counterexamples.*

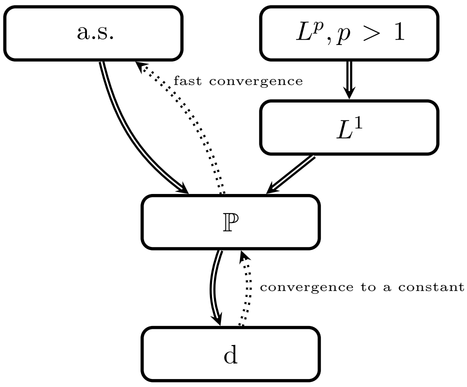

Figure 1: Hierarchy of modes of convergences

Theorem 4.2 *Let* $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be random variables, then* ▷ 10.12.2019

$$X_n \xrightarrow{\text{d}} X \iff F_{X_n}(t) \xrightarrow{n\to\infty} F_X(t)$$

for all continuity points t *of* F_X*, where* F_{X_n}, F_X *are the associated distribution functions.*

Lemma 4.4 (Tightness)
Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)), n \in \mathbb{N}$ *be random variables. Suppose that the sequence of characteristic functions of* $(X_n)_{n\in\mathbb{N}}$ *evaluated at* t*,* $(\varphi_{X_n}(t))_{n\in\mathbb{N}}$*, has a limit* $\varphi(t)$ *for all* $t \in \mathbb{R}^d$*. Then*

$$\forall \varepsilon > 0 : \exists r(\varepsilon) : \forall R > r(\varepsilon) : \sup_{n\in\mathbb{N}} \mathbb{P}(\|X_n\| > R) \leq \varepsilon\,.$$

Theorem 4.3 *Let* $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *be random variables. Then*

$$X_n \xrightarrow{\text{d}} X \iff \varphi_{X_n}(t) \xrightarrow{n\to\infty} \varphi_X(t) \text{ for all } t \in \mathbb{R}^d\,.$$

Corollary 4.3.1 (Cramér-Wold theorem) ▷ 12.12.2019

Let $X_n : (\Omega_n, \mathcal{A}_n, \mathbb{P}_n) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)), n \in \mathbb{N}$ *and* $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *be random variables. Then*

$$X_n \xrightarrow{\mathrm{d}} X \iff \langle X_n, t\rangle \xrightarrow{\mathrm{d}} \langle X, t\rangle \text{ for all } t \in \mathbb{R}^d\,.$$

Corollary 4.3.2 (Slutsky's lemma)

Let $X_n, Y_n, X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *be random variables with* $X_n \xrightarrow{\mathrm{d}} X,\ X_n - Y_n \xrightarrow{\mathbb{P}} 0$. *Then* $Y_n \xrightarrow{\mathrm{d}} X$.

5 Central Limit Theorems

5.1 A Look at Several Central Limit Theorems

Definition 5.1 (Central limit theorem)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *be random variables and* $\mathcal{S}_n := X_1 + \ldots + X_n$ *. Then* $(X_n)_{n\in\mathbb{N}}$ *satisfies a central limit theorem*

$$:\Longleftrightarrow \exists (a_n)_{n\in\mathbb{N}}, (s_n)_{n\in\mathbb{N}} \in \mathbb{R} : \quad \frac{\mathcal{S}_n - a_n}{s_n} \xrightarrow{\mathrm{d}} G \sim \mathcal{N}(0,1)\,. \tag{CLT}$$

Theorem 5.1 *Let* $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be independent and identically distributed,* $\mathbb{E}X_1 = 0$, $\mathbb{V}X_1 = \sigma^2 \in (0,\infty)$ *and* $\mathcal{S}_n := X_1 + \cdots + X_n$. *Then* $(X_n)_{n\in\mathbb{N}}$ *satisfies a central limit theorem with* $a_n = 0, s_n = \sigma\sqrt{n}, n \in \mathbb{N}$ *, that is*

$$\frac{\mathcal{S}_n}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} G \sim \mathcal{N}(0,1)\,.$$

Theorem 5.2 (Lindeberg (1922) & Lévy (1925, 1937)) ▷ 17.12.2019
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *be independent random variables. Assume* $\mathbb{E}X_n = 0, \mathbb{V}X_n =: \sigma_n^2 \in (0,\infty)$ *and denote the distribution of* X_n *by* μ_n *. Define* $s_n^2 := \sigma_1^2 + \cdots + \sigma_n^2$ *and suppose that the Lindeberg condition*

$$\forall \varepsilon > 0 : \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\{|x|>\varepsilon s_k\}} x^2\, \mu_k(\mathrm{d}x) = 0 \tag{L}$$

holds. Let $\mathcal{S}_n := X_1 + \cdots + X_n$*, then*

$$\frac{\mathcal{S}_n}{s_n} \xrightarrow{\mathrm{d}} G \sim \mathcal{N}(0,1)\,.$$

Remark 5.1 *In the above setting one can consider the classical Lindeberg condition*

$$\forall \varepsilon > 0 : \lim_{n\to\infty} \sum_{k=1}^{n} \int_{\{|x|>\varepsilon s_n\}} x^2\, \mu_k(\mathrm{d}x) = 0\,, \tag{L'}$$

which is equivalent to the Lindeberg condition (L).

Lemma 5.1 (Feller (1935))
In the above setting the Lindeberg condition implies the Feller condition

$$\lim_{n\to\infty} \max_{1\le k\le n} \frac{\sigma_k}{s_n} = 0\,. \tag{F}$$

Corollary 5.2.1 (Lyapunov (1901)) ▷ 19.12.2019
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *be independent random variables with* $\mathbb{E}X_n = 0, \mathbb{V}X_n =: \sigma_n^2$. *Define* $s_n^2 := \sigma_1^2 + \cdots + \sigma_n^2$ *and suppose that the Lyapunov condition*

$$\exists \delta > 0 : \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}|X_k|^{2+\delta} = 0 \tag{LY}$$

holds, then (CLT) *holds.*

Remark 5.2 (LY) *is usually easier to check than* (L)*, but it is stronger than* (L). *In practice* (LY) *is checked with* $\delta = 1$.

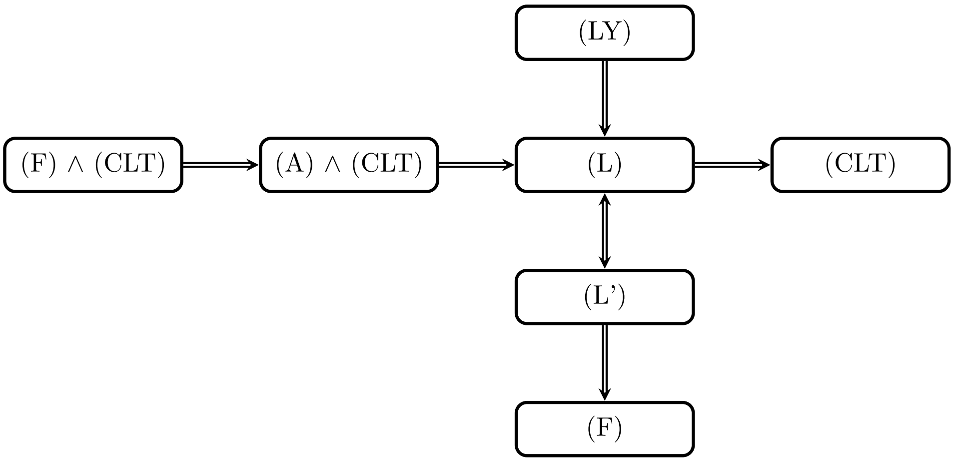

Figure 2: Relations of limit conditions

Theorem 5.3 (Feller (1935))
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ be independent random variables with $\mathbb{E}X_n = 0, \mathbb{V}X_n < \infty, \forall n \in \mathbb{N}$. Then, $(\mathrm{F}) \wedge (\mathrm{CLT}) \Leftrightarrow (\mathrm{L})$.

Remark 5.3 *The results remains valid for triangular arrays of random variables.*

$$\begin{array}{lllllll}
X_{1,1} & X_{1,2} & \dots & X_{1,k(1)} & & & \\
X_{2,1} & X_{2,2} & \dots & \dots & X_{2,k(2)} & & \\
\vdots & \vdots & & & & \ddots & \\
X_{n,1} & X_{n,2} & \dots & \dots & \dots & \dots & X_{n,k(n)} \\
\vdots & \vdots & & & & & \ddots
\end{array}$$

We assume independence within each line, but not of the lines. Roughly, there will be the following correspondence.

classical	triangular arrays
$X_j \sim \mu_j$	$X_{n,j} \sim \mu_{n,j}$
X_j independent	$(X_{n,j})_{j=1,\dots,k(n)}$ independent
$\sigma_j^2 = \mathbb{V}X_j$	$\sigma_{n,j}^2 = \mathbb{V}X_{n,j}$
$s_n^2 = \sum_{j=1}^n \sigma_j^2$	$s_n^2 = \sum_{j=1}^{k(n)} \sigma_{n,j}^2$
$\mathcal{S}_n = \sum_{j=1}^n X_j$	$\mathcal{S}_n = \sum_{j=1}^{k(n)} X_{n,j}$
(L) ∨ (L')	$\forall \varepsilon > 0 : \frac{1}{s_n^2} \sum_{j=1}^{k(n)} \int_{\lvert X_j \rvert > \varepsilon(s_j \vee s_n)} x^2 \mu_{n,j}(\mathrm{d}x) \to 0$
(F)	$\max_{1 \le j \le k(n)} \frac{\sigma_{n,j}}{s_n} \to 0$
(A)	$\forall \varepsilon > 0 : \max_{1 \le j \le k(n)} \mathbb{P}(\lvert X_{n,j} \rvert > \varepsilon s_n) \to 0$
(CLT)	$\frac{\mathcal{S}_n - \mathbb{E}\mathcal{S}_n}{s_n} \underset{n\to\infty}{\overset{\mathrm{d}}{\to}} G \sim \mathcal{N}(0,1)$

Theorem 5.4 (Goncharov (1944)) ▷ 07.01.2020
Let $\mathcal{S}_n$ be the number of cycles of a uniform random permutation on n elements. Then

$$\frac{\mathcal{S}_n - \log n}{\sqrt{\log n}} \overset{\mathrm{d}}{\to} G \sim \mathcal{N}(0,1)\,.$$

5.2 Poisson Limit Theorem

Theorem 5.5 (Poisson limit theorem)
Let $X_{1,n}, \dots, X_{n,n} : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}), n \in \mathbb{N}$ *be independent random variables such that*

$$\mathbb{P}(X_{k,n} = 1) = p_{k,n} \quad \textit{and} \quad \mathbb{P}(X_{k,n} = 0) = 1 - p_{k,n} \quad \textit{for all } k, n \in \mathbb{N}, 1 \le k \le n\,.$$

Suppose that $\max_{1 \le k \le n} p_{k,n} \xrightarrow{n\to\infty} 0$ *and* $\sum_{k=1}^{n} p_{k,n} \xrightarrow{n\to\infty} \lambda$*. Then*

$$\mathcal{S}_n = X_{1,n} + \cdots + X_{n,n} \xrightarrow{\mathrm{d}} Z \sim Po(\lambda)\,.$$

5.3 Weak Law of Large Numbers

Theorem 5.6 (Weak law of large numbers)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{R}$ *be independent and identically distributed random variables with* $\mathbb{E}X_1 < \infty$*, in the improper Riemann sense, and* $\mathcal{S}_n := X_1 + \cdots + X_n$ *. Then* $\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}X_1$.

Remark 5.4 *If additionally* $\mathbb{V}X_n < \infty$ *for all* $n \in \mathbb{N}$*, the random variables* $(X_n)_{n\in\mathbb{N}}$ *need not be independent nor identically distributed, but only pairwise uncorrelated for*

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}X_k) \xrightarrow{\mathbb{P}} 0$$

to hold.

6 Conditional Expectations

6.1 Construction and Elementary Properties

Definition 6.1 (Conditional expected value) ▷ 09.01.2020
Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}), F \in \mathcal{A}$ with $\mathbb{P}(F) > 0$. We define the conditional expected value as

$$\mathbb{E}(X \mid F) := \int X \,\mathrm{d}\mathbb{P}_F \,.$$

Lemma 6.1 *Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P}), F \in \mathcal{A}$ with $\mathbb{P}(F) > 0$. Then*

$$\mathbb{E}(X \mid F) = \frac{\mathbb{E}(X \cdot \mathbb{1}_F)}{\mathbb{P}(F)} \,.$$

Definition 6.2 (Conditional expectation)
Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra over Ω. The conditional expectation of X with respect to $\mathcal{F}$ is a $\mathcal{F}$-measurable random variable $X^{\mathcal{F}} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for which either of

$$\forall F \in \mathcal{F} : \int_F X \,\mathrm{d}\mathbb{P} = \int_F X^{\mathcal{F}} \,\mathrm{d}\mathbb{P}_{|\mathcal{F}} \iff \mathbb{E}\left(X \mathbb{1}_F\right) = \mathbb{E}\left(X^{\mathcal{F}} \mathbb{1}_F\right) \tag{CE}$$

holds.

Remark 6.1 *In the above situation, we will sometimes denote a random variable $X^{\mathcal{F}}$ fulfilling either condition* (CE) *as*

$$\mathbb{E}(X \mid \mathcal{F}) := X^{\mathcal{F}} \,.$$

Further, if $X = \mathbb{1}_A$ for $A \in \mathcal{A}$, then

$$\mathbb{P}(A \mid \mathcal{F}) := \mathbb{E}(\mathbb{1}_A \mid \mathcal{F}) = X^{\mathcal{F}} \,.$$

Theorem 6.1 (Existence & uniqueness of conditional expectations)
Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra over Ω.

a) If $X^{\mathcal{F}}, Y^{\mathcal{F}} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfy

$$\mathbb{E}[X\mathbb{1}_F] = \mathbb{E}[X^{\mathcal{F}}\mathbb{1}_F] = \mathbb{E}[Y^{\mathcal{F}}\mathbb{1}_F] \,, \ \forall F \in \mathcal{F} \,,$$

then $X^{\mathcal{F}} = Y^{\mathcal{F}}$ $\mathbb{P}$-almost surely.

b) A random variable $X^{\mathcal{F}} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ fulfilling (CE) *exists.*

Remark 6.2 *When working with conditional expectations, we usually apply the following method.* ▷ 14.01.2020

1. *Check the defining equation* (CE) *to find a* <u>*candidate*</u> *for the conditional expectation.*
2. *Conclude by* <u>*uniqueness*</u>*, that this candidate is indeed the conditional expectation.*

Theorem 6.2 *Let $X, Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, $a, b, c \in \mathbb{R}$ and let $\mathcal{F} \subseteq \mathcal{A}$ be a σ-algebra over Ω. Then*

1. $X \geq 0 \quad \Rightarrow \quad \mathbb{E}\left(X \mid \mathcal{F}\right) \geq 0$,
2. $X \equiv c \quad \Rightarrow \quad \mathbb{E}\left(X \mid \mathcal{F}\right) \equiv c$,
3. $\mathcal{F} = \{\emptyset, \Omega\} \quad \Rightarrow \quad \mathbb{E}\left(X \mid \mathcal{F}\right) = \mathbb{E}X$,
4. $\mathbb{E}\left(\mathbb{E}\left(X \mid \mathcal{F}\right)\right) = \mathbb{E}X$,
5. $\mathbb{E}\left(a \cdot X + b\dot{Y} \mid \mathcal{F}\right) = a \cdot \mathbb{E}\left(X \mid \mathcal{F}\right) + b \cdot \mathbb{E}\left(Y \mid \mathcal{F}\right)$ *and*
6. $X \geq Y \quad \Rightarrow \quad \mathbb{E}\left(X \mid \mathcal{F}\right) \geq \mathbb{E}\left(Y \mid \mathcal{F}\right)$.

Theorem 6.3 (Convergence theorems)
Let $X, X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ be random variables and $\mathcal{F} \subseteq \mathcal{A}$ a σ-algebra over Ω.

- *Conditional Fatou's lemma: If $X_n \geq 0$, $X_n \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ for all $n \in \mathbb{N}$ and $\liminf_{n\to\infty} \mathbb{E}X_n < \infty$, then $\mathbb{P}$-almost surely*

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n \mid \mathcal{F}\right) \leq \liminf_{n\to\infty} \mathbb{E}\left(X_n \mid \mathcal{F}\right) .$$

- *Conditional dominated convergence: If $X_n \xrightarrow{\text{a.s.}} X$ and there is $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ such that for all $n \in \mathbb{N}$, we have $|X_n| \leq Y$, then $\mathbb{P}$-almost surely*

$$\lim_{n\to\infty} \mathbb{E}\left(X_n \mid \mathcal{F}\right) = \mathbb{E}\left(\lim_{n\to\infty} X_n \mid \mathcal{F}\right) = \mathbb{E}\left(X \mid \mathcal{F}\right) .$$

- *Conditional monotone convergence: If $X_n \geq 0, X_n \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ for all $n \in \mathbb{N}$, $X_n \uparrow X$ and $\sup_{n\in\mathbb{N}} \mathbb{E}X_n < \infty$, then $\mathbb{P}$-almost surely*

$$\mathbb{E}\left(X_n \mid \mathcal{F}\right) \uparrow \mathbb{E}\left(X \mid \mathcal{F}\right) .$$

- *Conditional Jensen inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ convex and measurable and $\mathbb{E}|\varphi(X)| < \infty$, then almost surely*

$$\varphi(\mathbb{E}\left(X \mid \mathcal{F}\right)) \leq \mathbb{E}\left(\varphi(X) \mid \mathcal{F}\right) .$$

Theorem 6.4 *Let $X, Y, Z : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be random variables, $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F} \subseteq \mathcal{A}$ a σ-algebra over Ω, then*

1. pull-out: $Z \in L^\infty(\mathcal{F}, \mathbb{P}) \Rightarrow \mathbb{E}\left(X \cdot Z \mid \mathcal{F}\right) = Z \cdot \mathbb{E}\left(X \mid \mathcal{F}\right)$,
2. pull-out: Z $\mathcal{F}$-mb, $\mathbb{E}|X \cdot Z| < \infty \Rightarrow \mathbb{E}\left(XZ \mid \mathcal{F}\right) = Z \cdot \mathbb{E}\left(X \mid \mathcal{F}\right)$,
3. tower: $\mathcal{G} \subseteq \mathcal{F} \subseteq \mathcal{A}$ σ-fields $\Rightarrow \mathbb{E}\left(\mathbb{E}\left(X \mid \mathcal{F}\right) \mid \mathcal{G}\right) = \mathbb{E}\left(X \mid \mathcal{G}\right)$,
4. projection: $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \Rightarrow$
   $\mathbb{E}(X \mid \mathcal{F}) \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, $\mathbb{E}\left[\left(X - \mathbb{E}\left(X \mid \mathcal{F}\right)\right)^2\right] \leq \mathbb{E}\left(\left(X - Y\right)^2\right)$ *and*
5. monotonicity: $X \leq Y \Rightarrow \mathbb{E}\left(X \mid \mathcal{F}\right) \leq \mathbb{E}\left(Y \mid \mathcal{F}\right)$.

Remark 6.3 *In the above situation, there is a geometric meaning to (4.). $L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a Hilbert space, where $\langle u, v\rangle := \mathbb{E}\left(uv\right)$ with $L^2(\Omega, \mathcal{F}, \mathbb{P})$ being a linear subspace. Thus, as per (4.), $\mathbb{E}\left(X \mid \mathcal{F}\right)$ is the orthogonal projection of X onto $L^2(\Omega, \mathcal{F}, \mathbb{P})$.*

6.2 Conditional Expectation and Independence

Theorem 6.5 *$X, Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, $\mathcal{F}, \mathcal{G}$ σ-algebras over Ω and $g : (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^d)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ a bounded and measurable function. Then* ▷ 16.01.2020

1. $\mathcal{I}(X) \perp\!\!\!\perp \mathcal{F} \Rightarrow \mathbb{E}\left(X \mid \mathcal{F}\right) = \mathbb{E}X$,
2. $\mathcal{I}(X), \mathcal{G} \perp\!\!\!\perp \mathcal{F} \Rightarrow \mathbb{E}\left(X \mid \mathcal{F}, \mathcal{G}\right) = \mathbb{E}\left(X \mid \mathcal{G}\right)$ *and*
3. $X \perp\!\!\!\perp Y$ and Y $\mathcal{F}$-measurable $\Rightarrow \mathbb{E}\left(g(X, Y) \mid \mathcal{F}\right) = \mathbb{E}g(X, t)|_{t=Y}$.

6.3 Conditioning on Random Variables

Remark 6.4 *Let $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ be measurable and $\mathcal{F}, \mathcal{G} \subseteq \mathcal{A}$ σ-algebras over Ω. We will use abbreviations such as $X \perp\!\!\!\perp \mathcal{F}$ instead of $\mathcal{I}(X) \perp\!\!\!\perp \mathcal{F}$, and similarly $X, \mathcal{G} \perp\!\!\!\perp \mathcal{F}$ instead of $\sigma(\mathcal{I}(X) \cup \mathcal{G}) \perp\!\!\!\perp \mathcal{F}$.*

Remark 6.5 *Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ measurable. We use the abbreviation $\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \mathcal{I}(Y))$.*

Lemma 6.2 (Factorization Lemma)

$$\left.\begin{array}{l} Y : (\Omega, \mathcal{A}) \xrightarrow{mb.} (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \\ Z : (\Omega, \mathcal{I}(Y)) \xrightarrow{mb.} (\mathbb{R}, \mathcal{B}(\mathbb{R})) \end{array}\right\} \Longrightarrow \begin{array}{c} \exists\, g : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ measurable} \\ \textit{such that } Z = g(Y) \end{array}$$

Definition 6.3 *Let $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $Y : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then we denote by*

$$\mathbb{E}\left[X \mid Y = y\right] := g(y) \in \mathbb{R} \longleftarrow y \in \mathbb{R}^d$$

the measurable function $g : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that satisfies $\mathbb{E}\left(X \mid Y\right) = g(Y)$ by the previous lemma.

Remark 6.6 *In the discrete setting, that is $\operatorname{im} Y$ is countable, we are able to evaluate the expression as*

$$\mathbb{E}\left[X \mid Y = y\right] = \begin{cases} \frac{\mathbb{E}(X \mathbb{1}_y(Y))}{\mathbb{P}(Y=y)} & \text{if } \mathbb{P}(Y = y) > 0 \\ 0 & \text{otherwise} \end{cases}.$$

What about the continuous setting (conditioning on events with probability 0)?

Definition 6.4 (Conditional density)
Let $(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ be a random variable with an image measure that has a density $f_{(X,Y)}$ with respect to the Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. We define the conditional density of X given Y as

$$f_{X|Y}(x, y) := \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & \text{if } f_Y(y) > 0 \\ 0 & \text{otherwise} \end{cases}, \ \textit{where } x, y \in \mathbb{R}.$$

Remark 6.7 *Recall that in the setting of the previous definition, the following holds $\mathbb{P}$-almost surely*

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, \mathrm{d}y\,, \ \textit{where } y \in \mathbb{R}.$$

Theorem 6.6 *Let $(X, Y) : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ be a random variable with an image measure that has a density $f_{(X,Y)}$ with respect to the Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ and let $h : (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable such that $\mathbb{E}|h(X, Y)| < \infty$. Then* ▷ 21.01.2020

$$\mathbb{E}[h(X, Y) \mid Y = y] = \int_{\mathbb{R}} h(x, y) f_{X|Y}(x \mid y)\, \mathrm{d}x\,, \ \textit{for } \mathbb{P}^Y\textit{-almost all } y \in \mathbb{R}.$$

6.4 Introduction to Martingales

Definition 6.5 (Discrete stochastic process)
The family of random variables $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}$ is called a stochastic process.

Definition 6.6 (Discrete filtration)
Let $(\Omega, \mathcal{A})$ be a measurable space. A sequence $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ of sub σ-fields of $\mathcal{A}$ over Ω is called a filtration over $(\Omega, \mathcal{A})$.

Definition 6.7 (Adapted process)
A stochastic process $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}$ is adapted to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$ if for all $n \in \mathbb{N}$, X_n is $\mathcal{F}_n$-$\mathcal{E}$-measurable.

Definition 6.8 (Discrete martingale)
Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathcal{F}_n)_{n\in\mathbb{N}}$ a filtration over $(\Omega, \mathcal{A})$. The random variables $(X_n)_{n\in\mathbb{N}}$ are called a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ $:\Longleftrightarrow$

1. $X_n \in L^1(\Omega, \mathcal{F}_n, \mathbb{P})$ *and*

2. $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$ $\mathbb{P}$*-almost surely for all* $n \in \mathbb{N}$.

Remark 6.8 *In the above situation, if instead of* (2.) *only*

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \geq X_n \ \mathbb{P}\text{-almost surely for all } n \in \mathbb{N}$$

holds, then we refer to it as a submartingale and if only

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \leq X_n \ \mathbb{P}\text{-almost surely for all } n \in \mathbb{N}$$

holds, we call it a supermartingale.

Definition 6.9 (Predictable process) ▷ 23.01.2020
A stochastic process $C_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}_{\geq 2}$ *is called predictable with respect to a filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *over* $(\Omega, \mathcal{A})$ *if* C_{n+1} *is* $\mathcal{F}_n$-$\mathcal{E}$*-measurable for all* $n \in \mathbb{N}$.

Definition 6.10 (Martingale transform)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *be a martingale with respect to filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *over* $(\Omega, \mathcal{A})$ *and let* $C_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}_{\geq 2}$ *be a predictable process with respect to* $(\mathcal{F}_n)_{n\in\mathbb{N}}$*. We define*

$$(C \cdot X)_n := \sum_{k=1}^{n} C_{k+1}(X_{k+1} - X_k)$$

and call $((C \cdot X)_n)_{n\in\mathbb{N}}$ *the martingale transform of* $(X_n)_{n\in\mathbb{N}}$ *by* $(C_n)_{n\in\mathbb{N}_{\geq 2}}$.

Lemma 6.3 *For martingales* $(X_n)_{n\in\mathbb{N}}$ *we have* $\mathbb{E}X_n = \mathbb{E}X_1$ *for all* $n \in \mathbb{N}$.

Theorem 6.7 *Let* $C_{n+1}, X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ *be stochastic processes, where* $(C_n)_{n\in\mathbb{N}_{\geq 2}}$ *is predictable and* $(X_n)_{n\in\mathbb{N}}$ *is a martingale with respect to a filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *over* $(\Omega, \mathcal{A})$*. Then* $((C \cdot X)_n)_{n\in\mathbb{N}}$ *is a martingale with respect to* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *with* $\mathbb{E}(C \cdot X)_n = 0$ *for all* $n \in \mathbb{N}$.

6.5 Martingales and Stopping Times

Definition 6.11 (Stopping time)
Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *be a filtration over a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. A random variable* $\tau : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ *such that*

$$\{\tau \leq n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}$$

is called a stopping time with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Remark 6.9 *In the above setting,* τ *is a stopping time* $\iff$ $\{\tau = n\} \in \mathcal{F}_n$ *for all* $n \in \mathbb{N}$.

Definition 6.12 (Stopped process)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E}), n \in \mathbb{N}$ *be a process adapted to filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *over* $(\Omega, \mathcal{A})$ *and let* $\tau : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ *be a stopping time with respect to* $(\mathcal{F}_n)_{n\in\mathbb{N}}$*. We call*

$$(X_n^\tau)_{n\in\mathbb{N}}, \text{ where } X_n^\tau(\omega) := X_{\min\{\tau(\omega), n\}}(\omega), \text{ for all } \omega \in \Omega, n \in \mathbb{N}$$

the process $(X_n)_{n\in\mathbb{N}}$ *stopped at time* τ.

Theorem 6.8 (Elementary stopping theorem)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be a martingale with respect to* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *over* $(\Omega, \mathcal{A})$ *and* $\tau : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ *be a stopping time with respect to* $(\mathcal{F}_n)_{n\in\mathbb{N}}$*. Then the stopped process* $(X_n^\tau)_{n\in\mathbb{N}}$ *is again a martingale with* $\mathbb{E}X_n^\tau = \mathbb{E}X_1$ *for all* $n \in \mathbb{N}$.

Definition 6.13 (σ-algebra of the τ-past) ▷ 28.01.2020
Let $\tau : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ be a stopping time with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$. We call

$$\mathcal{A}_\tau := \{A \in \mathcal{A} \mid A \cap \{\tau \leq n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\}$$

the σ-algebra of the τ-past.

Theorem 6.9 *Let $\tau : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\leq\infty})$ be $\mathbb{P}$-almost surely finite and a stopping time with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$. Let further $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ be a stochastic process adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Then the random variable*

$$X_\tau : \Omega \to \mathbb{R}\,,\ \omega \mapsto X_\tau(\omega) := \begin{cases} X_{\tau(\omega)} & , \text{ if } \tau(\omega) < \infty \\ 0 & , \text{ otherwise} \end{cases}$$

is $\mathcal{A}_\tau$-$\mathcal{B}(\mathbb{R})$-measurable.

Theorem 6.10 (Doob's stopping theorem)
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$ and $\tau : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{N}_{\leq\infty}, 2^{\mathbb{N}_{\leq\infty}})$ be a stopping time with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Then X_τ is $\mathbb{P}_{|\mathcal{A}_\tau}$-integrable and $\mathbb{E}X_\tau = \mathbb{E}X_0$ if either of

1. *τ is $\mathbb{P}$-almost surely bounded,*
2. *τ is $\mathbb{P}$-almost surely finite and $(X_n)_{n\in\mathbb{N}}$ is $\mathbb{P}$-almost surely bounded or*
3. *$\mathbb{E}\tau < \infty$ and $(|X_{n+1} - X_n|)_{n\in\mathbb{N}}$ is $\mathbb{P}$-almost surely bounded*

is satisfied.

Remark 6.10 *If in the above situation $(X_n)_{n\in\mathbb{N}}$ is only a supermartingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and* (1.) - (3.) *or*

4. *τ is $\mathbb{P}$-almost surely finite and $(X_n)_{n\in\mathbb{N}}$ is $\mathbb{P}$-almost surely non-negative*

then X_τ is $\mathbb{P}_{|\mathcal{A}_\tau}$-integrable and instead $\mathbb{E}X_\tau \leq \mathbb{E}X_0$.

6.6 Martingales and Convergence

Definition 6.14 (L^p-boundedness)
Let $(X_n)_{n\in\mathbb{N}} \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ and let $p \in [1, \infty)$. We say

$$(X_n)_{n\in\mathbb{N}} \text{ is } L^p\text{-bounded } :\iff \exists K \in \mathbb{R} : \forall n \in \mathbb{N} : \|X_n\|_p < K\,.$$

Lemma 6.4 (Upcrossing lemma) ▷ 30.01.2020
Let $X_n : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ be a supermartingale with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$ and let $a, b \in \mathbb{R}, a < b$. Then

$$\forall N \geq 0 : \quad (b-a)\mathbb{E}U_N[a,b] \leq -\mathbb{E}\min\{(X_N - a), 0\}\,,$$

where

$$U_N[a,b] := \max\left\{k \in \mathbb{N}_0 \,\middle|\, \begin{array}{c} \exists 0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq N \in \mathbb{N} \\ \text{s.t. } X_{s_i} < a \text{ and } X_{t_i} > b, \forall i \in \{1, \ldots, k\} \end{array}\right\}.$$

Lemma 6.5 *Let $(X_n)_{n\in\mathbb{N}} \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ be a L^1-bounded supermartingale with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega, \mathcal{A})$. Then the monotone limit*

$$U[a,b] := \lim_{N\to\infty} U_N[a,b]$$

satisfies

$$(b-a)\mathbb{E}U[a,b] \ \leq\ |a| + \sup_{n\geq 0}\mathbb{E}|X_n| \ <\ \infty\,.$$

In particular,

$$\mathbb{P}(U[a,b] = \infty) = 0\,.$$

Theorem 6.11 (Doob's martingale convergence theorem)
Let $(X_n)_{n\in\mathbb{N}} \in L^1(\Omega,\mathcal{A},\mathbb{P})$ be a L^1-bounded supermartingale with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega,\mathcal{A})$. Then a random variable $X \in L^1(\Omega,\mathcal{A},\mathbb{P})$ exists such that $X_n \xrightarrow{\text{a.s.}} X$.

Remark 6.11 *Let $(X_n)_{n\in\mathbb{N}} \in L^1(\Omega,\mathcal{A},\mathbb{P})$ be any supermartingale with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega,\mathcal{A})$. If $(X_n)_{n\in\mathbb{N}}$ non-negative $\mathbb{P}$-almost surely then $(X_n)_{n\in\mathbb{N}}$ is already L^1-bounded.*

Definition 6.15 *Random variables $X_n : (\Omega,\mathcal{A},\mathbb{P}) \to (\mathbb{R},\mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ are called $\mathbb{P}$-uniformly integrable if*

$$\forall \varepsilon > 0 : \exists K > 0 : \quad \mathbb{E}(\mathbb{1}_{\{|X_n|>K\}}|X_n|) \leq \varepsilon\,,\ \forall n \in \mathbb{N}\,.$$

Remark 6.12 *In the above situation if $(X_n)_{n\in\mathbb{N}} \in L^p(\Omega,\mathcal{A},\mathbb{P})$ for some $p > 1$, they are $\mathbb{P}$-uniformly integrable as well.*

Theorem 6.12 *Let $X_n : (\Omega,\mathcal{A},\mathbb{P}) \to (\mathbb{R},\mathcal{B}(\mathbb{R})), n \in \mathbb{N}$ be a $\mathbb{P}$-uniformly integrable martingale with respect to filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ over $(\Omega,\mathcal{A})$. Then a random variable $X \in L^1(\Omega,\mathcal{A},\mathbb{P})$ exists such that*

$$X_n \xrightarrow{\text{a.s. \& } L^1} X \quad \textit{and} \quad \forall n \in \mathbb{N} : \mathbb{E}(X \mid \mathcal{F}_n) = X_n \ \ \mathbb{P}\textit{-almost surely}.$$