Ruprecht-Karls-Universität Heidelberg
Institut für Mathematik
Professor Dr. Jan JOHANNES

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Outline of the lecture course

PROBABILITY THEORY

Summer semester 2026

Preliminary version: July 13, 2026

If you find **errors in the outline**, please send a short note
by email to johannes@math.uni-heidelberg.de.

MΛTHEMΛTIKON, Im Neuenheimer Feld 205, 69120 Heidelberg
phone: +49 6221 54.14.190 – fax: +49 6221 54.14.101
email: johannes@math.uni-heidelberg.de
webpage: sip.math.uni-heidelberg.de

Table of contents

Chapter 1

Measure and integration theory

§01 Measure theory

§01.01 **Notation.** For $x, y \in \mathbb{R}$ we agree on the following notations $\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \in (-\infty, x]\}$ (integer part), $x \vee y = \max(x, y)$ (maximum), $x \wedge y = \min(x, y)$ (minimum), $x^+ = \max(x, 0)$ (positive part), $x^- = \max(-x, 0)$ (negative part) and $|x| = x^- + x^+$ (modulus).

(a) For $c \in \mathbb{R}$ and $\mathbb{A} \subseteq \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\} = [-\infty, \infty]$ we set $\mathbb{A}_{\geqslant c} := \mathbb{A} \cap [c, \infty]$, $\mathbb{A}_{\leqslant c} := \mathbb{A} \cap [-\infty, c]$, $\mathbb{A}_{>c} := \mathbb{A} \cap (c, \infty]$, $\mathbb{A}_{<c} := \mathbb{A} \cap [-\infty, c)$, $\mathbb{A}_{\neq c} := \mathbb{A} \setminus \{c\}$, and $\overline{\mathbb{A}} := \mathbb{A} \cup \{\pm\infty\}$.

(b) For $b \in \overline{\mathbb{R}}$ and $a \in \overline{\mathbb{R}}_{<b}$ we write $[\![a, b]\!] := [a, b] \cap \overline{\mathbb{Z}}$, $[\![a, b)\!) := [a, b) \cap \overline{\mathbb{Z}}$, $(\!(a, b]\!] := (a, b] \cap \overline{\mathbb{Z}}$, and $(\!(a, b)\!) := (a, b) \cap \overline{\mathbb{Z}}$. Moreover, let $[\![n]\!] := [\![1, n]\!]$ and $[\![n)\!) := [\![1, n)\!)$ for $n \in \mathbb{N} = \mathbb{Z}_{>0}$.

(c) $\Omega \neq \emptyset$ denotes a nonempty set, 2^Ω the set of all subsets of Ω, and $2^\Omega_{\neq\emptyset} := 2^\Omega \setminus \{\emptyset\}$. A set is called *countable* if it is at most countable infinite, meaning either finite or countably infinite. The *cardinality* of a set A is denoted by $|A|$, its complement in Ω by $A^c := \Omega \setminus A \in 2^\Omega$ and we agree on $\emptyset = \bigcup_{j \in \emptyset} A_j$. □

§01|01 Classes of sets

§01.02 **Definition.** A class of sets $\mathscr{E} \subseteq 2^\Omega$ is called

$\cap$-closed (*closed under intersections*) or a *π-system* if $A \cap B \in \mathscr{E}$ whenever $A, B \in \mathscr{E}$,

σ-$\cap$-closed (*closed under countable intersections*) if $\bigcap_{n \in \mathbb{N}} A_n \in \mathscr{E}$ for any sequence $(A_n)_{n \in \mathbb{N}}$ of sets in $\mathscr{E}$,

$\cup$-closed (*closed under unions*) if $A \cup B \in \mathscr{E}$ whenever $A, B \in \mathscr{E}$,

σ-$\cup$-closed (*closed under countable unions*) if $\bigcup_{n \in \mathbb{N}} A_n \in \mathscr{E}$ for any sequence $(A_n)_{n \in \mathbb{N}}$ of sets in $\mathscr{E}$,

$\setminus$-closed (*closed under differences*) if $A \setminus B \in \mathscr{E}$ whenever $A, B \in \mathscr{E}$, and

closed under complements if $A^c := \Omega \setminus A \in \mathscr{E}$ for any set $A \in \mathscr{E}$. □

§01.03 **Remark.**

(a) If $\mathscr{E} \subseteq 2^\Omega$ is closed under complements then de Morgan's rule (i.e. $(\bigcup A_i)^c = \bigcup A_i^c$) implies immediately the equivalences of $\cup$-closed and $\cap$-closed, as well as of σ-$\cup$-closed and σ-$\cap$-closed.

(b) Let $\mathscr{E} \subseteq 2^\Omega$ be $\setminus$-closed. Then $\mathscr{E}$ is $\cap$-closed. If in addition $\mathscr{E}$ is σ-$\cup$-closed, then $\mathscr{E}$ is σ-$\cap$-closed. Any countable (respectively finite) union of sets in $\mathscr{E}$ can be expressed as a countable (respectively finite) disjoint union of sets in $\mathscr{E}$. □

§01.04 **Definition.** A class of sets $\mathscr{E} \subseteq 2^\Omega$ is called

semiring if (i) $\emptyset \in \mathscr{E}$, (ii) for any two sets $A, B \in \mathscr{E}$ the difference set $A \setminus B$ is a finite union of mutually disjoints sets in $\mathscr{E}$, and (iii) $\mathscr{E}$ is $\cap$-closed;

ring, if (R1) $\varnothing \in \mathscr{E}$, (R2) $\mathscr{E}$ is $\backslash$-closed, and (R3) $\mathscr{E}$ is $\cup$-closed;
σ-ring, if $\mathscr{E}$ is a σ-$\cup$-closed ring;

algebra, if (A1) $\Omega \in \mathscr{E}$, (A2) $\mathscr{E}$ is $\backslash$-closed, and (A3) $\mathscr{E}$ is $\cup$-closed;
σ-algebra, if $\mathscr{E}$ is a σ-$\cup$-closed algebra;

Dynkin-system or *λ-system*, if (D1) $\Omega \in \mathscr{E}$, (D2) $\mathscr{E}$ is closed under complements, and (D3) $\biguplus_{n\in\mathbb{N}} A_n \in \mathscr{E}$ for any choice of countably many pairwise disjoint sets $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{E}$. □

§01.05 **Remark.**

(a) Sometimes the disjoint union of sets is denoted by the symbol $\biguplus$. Note that this is not a new operation but only stresses the fact that the sets involved are mutually disjoint.

(b) For any $\Omega \neq \varnothing$ the classes $\{\varnothing, \Omega\}$ and 2^Ω are trivial examples of algebras, σ-algebras and Dynkin systems. Trivial examples of semirings, rings and σ-rings are $\{\varnothing\}$ and 2^Ω.

(c) A (set-)ring $\mathscr{R}$ equipped with the symmetric difference Δ as addition and the intersection $\cap$ as multiplication forms an Abelian algebraic ring $(\mathscr{R}, \Delta, \cap)$.

(d) A class of sets $\mathscr{A} \subseteq 2^\Omega$ is an algebra if and only if $\Omega \in \mathscr{A}$, and $\mathscr{A}$ is closed under complements and $\cap$-closed.

(e) A class of sets $\mathscr{A} \subseteq 2^\Omega$ with $\Omega \in \mathscr{A}$, which is closed under complements and σ-$\cap$-closed is a σ-algebra.

(f) Let $\mathscr{D} \subseteq 2^\Omega$ be a Dynkin-system. The condition (D2), i.e. $\mathscr{D}$ is closed under complements, can be equivalently replaced by the apparently stronger condition (D2') $B\backslash A \in \mathscr{D}$ for any $A, B \in \mathscr{D}$ with $A \subseteq B$, since each Dynkin-system satisfies also (D2'). Indeed for $A, B \in \mathscr{D}$ with $A \subseteq B$ the sets A and B^c are mutually disjoint and $B\backslash A = (A \uplus B^c)^c \in \mathscr{D}$.

(g) Every σ-algebra also is a Dynkin-system. The converse does not apply because (D3) is required only for mutually disjoint sets. For example let $\Omega = \{1, 2, 3, 4\}$ and $\mathscr{D} = \{\varnothing, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \Omega\}$. Then $\mathscr{D}$ is a Dynkin-system but is not an algebra. □

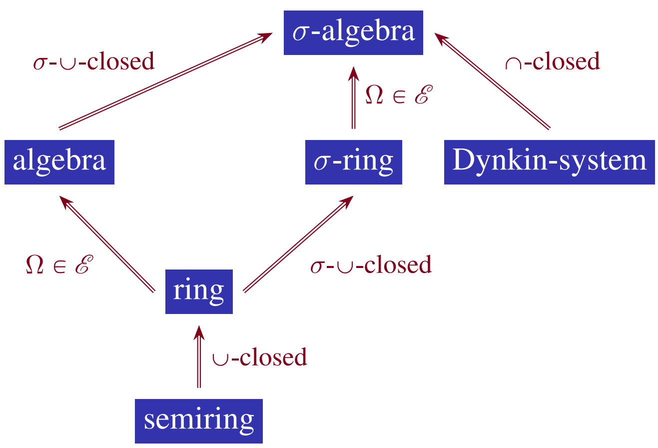
§01.06 **Illustration.**

(i) Every σ-algebra also is a Dynkin-system, an algebra and a σ-ring.

(ii) Every σ-ring is a ring, and every ring is a semiring.

(iii) Every algebra is a ring. An algebra on a finite set Ω is a σ-algebra.

Fig. 01 [§01] Inclusions between classes of sets $\mathscr{E} \subseteq 2^\Omega$.



The Figure 01 [§01] was created based on Klenke (2020, Fig. 1.1, p.7). □

§01.07 **Lemma**. *A Dynkin-system $\mathscr{D} \subseteq 2^\Omega$ is $\cap$-closed if and only if it is a σ-algebra.*

§01.08 **Proof** of Lemma §01.07. In the lecture course EWS. □

§01.09 **Lemma**. *Let $\mathscr{E} \subseteq 2^\Omega$ be a class of sets. Then*

$$\sigma(\mathscr{E}) := \bigcap \{\mathscr{A} : \mathscr{A} \subseteq 2^\Omega \text{ is a } \sigma\text{-algebra and } \mathscr{E} \subseteq \mathscr{A}\} \qquad \text{and}$$

$$\delta(\mathscr{E}) := \bigcap \{\mathscr{D} : \mathscr{D} \subseteq 2^\Omega \text{ is a Dynkin-system and } \mathscr{E} \subseteq \mathscr{D}\}$$

is the smallest σ-algebra, respectively, Dynkin-system on Ω containing $\mathscr{E}$. $\mathscr{E}$ is called generator, and $\sigma(\mathscr{E})$ and $\delta(\mathscr{E})$ is called the σ-algebra and the Dynkin-system generated by $\mathscr{E}$, respectively.

§01.10 **Proof** of Lemma §01.09. In the lecture course EWS. □

§01.11 **π-λ-Theorem**. *Let $\mathscr{E} \subseteq 2^\Omega$ be $\cap$-closed. Then $\sigma(\mathscr{E}) = \delta(\mathscr{E})$ and also $\sigma(\mathscr{E}) \subseteq \mathscr{D}$ for any Dynkin-system $\mathscr{D} \subseteq 2^\Omega$ with $\mathscr{E} \subseteq \mathscr{D}$.*

§01.12 **Proof** of Theorem §01.11. In the lecture course EWS. □

§01.13 **Definition**. Let $\mathscr{E} \subseteq 2^\Omega$ and $A \in 2^\Omega_{\neq\emptyset}$ be an arbitrary class of subsets, respectively, nonempty subset of Ω. The class $\mathscr{E}_A := \mathscr{E}\big|_A := \mathscr{E} \cap A := \{B \cap A : B \in \mathscr{E}\} \subseteq 2^\Omega$ of subsets of Ω is called *trace* of $\mathscr{E}$ on A or *restriction* of $\mathscr{E}$ to A. □

§01.14 **Remark**. If $\mathscr{E}$ is a semiring, (σ-)ring or (σ-)algebra then $\mathscr{E}_A$ is a class of sets of the same type as $\mathscr{E}$, however, on A instead of Ω. For a Dynkin-system this generally does not apply. Moreover, we have $\sigma(\mathscr{E})\big|_A = \sigma(\mathscr{E}|_A)$. □

§01.15 **Reminder**.

(a) Let $\mathcal{S}$ be a metric (or topological) space and $\mathscr{O}$ the class of open subsets in $\mathcal{S}$. The σ-algebra $\boxed{\mathscr{B}_{\mathcal{S}}} := \sigma(\mathscr{O})$ that is generated by the open sets $\mathscr{O}$ is called the *Borel σ-algebra* on $\mathcal{S}$. The elements of $\mathscr{B}_{\mathcal{S}}$ are called *Borel sets* or *Borel measurable sets*.

(b) In many cases, we are interested in the Borel σ-algebra $\boxed{\mathscr{B}^n} := \mathscr{B}_{\mathbb{R}^n}$ over $\mathbb{R}^n$, where $\mathbb{R}^n$ is equipped with the Euclidean distance $d(x,y) = \|x-y\| = \sqrt{\sum_{i\in[\![n]\!]}(x_i-y_i)^2}$ for $x = (x_i)_{i\in[\![n]\!]}, y = (y_i)_{i\in[\![n]\!]} \in \mathbb{R}^n$.

(c) For $a = (a_i)_{i\in[\![n]\!]}, b = (b_i)_{i\in[\![n]\!]} \in \overline{\mathbb{R}}^n$ we write $a < b$, if $a_i < b_i$ for all $i \in [\![n]\!]$. For $a < b$, define the open *rectangle* as the Cartesian product $(a,b) := \bigtimes_{i\in[\![n]\!]}(a_i,b_i) := (a_1,b_1) \times (a_2,b_2) \times \cdots \times (a_n,b_n)$. Analogously, we define $[a,b]$, $(a,b]$ and $[a,b)$. Moreover, we set $(-\infty,b) := \bigtimes_{i\in[\![n]\!]}(-\infty,b_i)$ and $(-\infty,b] := \bigtimes_{i\in[\![n]\!]}(-\infty,b_i]$.

(d) The Borel σ-algebra $\mathscr{B}^n$ is generated by any of the classes of sets:
(i) $\mathscr{E}_1 := \{A \subseteq \mathbb{R}^n : A \text{ is closed}\}$; (ii) $\mathscr{E}_2 := \{A \subseteq \mathbb{R}^n : A \text{ is compact}\}$;
(iii) $\mathscr{E}_3 := \{(a,b) : a,b \in \mathbb{Q}^n, a < b\}$; (iv) $\mathscr{E}_4 := \{[a,b] : a,b \in \mathbb{Q}^n, a < b\}$;
(v) $\mathscr{E}_5 := \{(a,b] : a,b \in \mathbb{Q}^n, a < b\}$; (vi) $\mathscr{E}_6 := \{[a,b) : a,b \in \mathbb{Q}^n, a < b\}$;
(vii) $\mathscr{E}_7 := \{(-\infty,b] : b \in \mathbb{Q}^n\}$; (viii) $\mathscr{E}_8 := \{(-\infty,b) : b \in \mathbb{Q}^n\}$;
(ix) $\mathscr{E}_9 := \{(a,\infty) : a \in \mathbb{Q}^n\}$ and (x) $\mathscr{E}_{10} := \{[a,\infty) : a \in \mathbb{Q}^n\}$. (Exercise).

(e) We denote by $\boxed{\overline{\mathscr{B}}} := \mathscr{B}_{\overline{\mathbb{R}}}$ the Borel σ-algebra over the extension $\overline{\mathbb{R}} := [-\infty,\infty]$ of the real line by the points $\{\pm\infty\}$ where in $\overline{\mathbb{R}}$ the sets $\{-\infty\}$ and $\{\infty\}$ are closed, and $\mathbb{R}$ is open. In particular, $\boxed{\mathscr{B}} := \mathscr{B}_{\mathbb{R}} = \overline{\mathscr{B}} \cap \mathbb{R}$ is the Borel σ-algebra over $\mathbb{R}$. For $c \in \mathbb{R}$ and

σ-algebra $\mathscr{A} \subseteq 2^{\overline{\mathbb{R}}}$ we write $\mathscr{A}_{\geqslant c} := \mathscr{A} \cap \overline{\mathbb{R}}_{\geqslant c}$, $\mathscr{A}_{>c} := \mathscr{A} \cap \overline{\mathbb{R}}_{>c}$, $\mathscr{A}_{\leqslant c} := \mathscr{A} \cap \overline{\mathbb{R}}_{\leqslant c}$, and $\mathscr{A}_{<c} := \mathscr{A} \cap \overline{\mathbb{R}}_{<c}$ □

§01|02 Set functions

§01.16 **Definition**. Let $\mathscr{E} \subseteq 2^{\Omega}$ and let $\mu : \mathscr{E} \to \overline{\mathbb{R}}_{\geqslant 0} = [0, \infty]$ be a set function. We say that μ is

monotone if $\mu(A) \leqslant \mu(B)$ for any two sets $A, B \in \mathscr{E}$ with $A \subseteq B$,

additive if $\mu(\biguplus_{j\in[\![n]\!]} A_j) = \sum_{j\in[\![n]\!]} \mu(A_j)$ for any choice of finitely many mutually disjoint sets $A_j \in \mathscr{E}$, $j \in [\![n]\!]$, with $\biguplus_{j\in[\![n]\!]} A_j \in \mathscr{E}$,

σ-additive if $\mu(\biguplus_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ for any choice of countably many mutually disjoint sets $A_j \in \mathscr{E}$, $j \in \mathbb{N}$, with $\biguplus_{j\in\mathbb{N}} A_j \in \mathscr{E}$,

subadditive if $\mu(A) \leqslant \sum_{i\in[\![n]\!]} \mu(A_j)$ for any choice of finitely many sets $A, A_j \in \mathscr{E}$, $j \in [\![n]\!]$, with $A \subseteq \bigcup_{j\in[\![n]\!]} A_j$,

σ-subadditive if $\mu(A) \leqslant \sum_{j\in\mathbb{N}} \mu(A_j)$ for any choice of countably many sets $A, A_j \in \mathscr{E}$, $j \in \mathbb{N}$, with $A \subseteq \bigcup_{j\in\mathbb{N}} A_j$. □

§01.17 **Definition**. Let $\mathscr{E} \subseteq 2^{\Omega}$ be a semiring. A set function $\mu : \mathscr{E} \to \overline{\mathbb{R}}_{\geqslant 0}$ with $\mu(\emptyset) = 0$ is called a

content if μ is additive,

premeasure if μ is σ-additive,

measure if μ is a premeasure and $\mathscr{E}$ is a σ-algebra, and

probability measure if μ is a measure and $\mu(\Omega) = 1$.

We denote by $\mathfrak{M}(\mathscr{E})$ the set of all premeasures on $(\Omega, \mathscr{E})$. A content μ on $\mathscr{E}$ is called

finite if $\mu(A) \in \mathbb{R}_{\geqslant 0}$ for every $A \in \mathscr{E}$ and

σ-finite if there exists a sequence of sets $(\mathcal{E}_j)_{j\in\mathbb{N}}$ in $\mathscr{E}$ such that $\Omega = \bigcup_{j\in\mathbb{N}} \mathcal{E}_j$ and $\mu(\mathcal{E}_j) \in \mathbb{R}_{\geqslant 0}$ for all $j \in \mathbb{N}$.

We denote by $\mathfrak{M}_{\mathrm{fi}}(\mathscr{E})$ and $\mathfrak{M}_{\sigma}(\mathscr{E})$ the set of all finite, respectively, σ-finite premeasures on $(\Omega, \mathscr{E})$. Moreover, for a σ-algebra $\mathscr{A} \subseteq 2^{\Omega}$ we denote by $\mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ the set of all probability measures on $(\Omega, \mathscr{A})$.

§01.18 **Example**.

(a) For $A \in 2^{\Omega}$ we denote by

$$\boxed{\mathbb{1}_A} : \Omega \to \{0, 1\} \quad \text{mit} \quad \mathbb{1}_A^{-1}(\{1\}) := A \quad \text{and} \quad \mathbb{1}_A^{-1}(\{0\}) := A^{\mathrm{c}}$$

the *indicator function* on A. For any σ-algebra $\mathscr{A} \subseteq 2^{\Omega}$ and $\omega \in \Omega$ the set function

$$\boxed{\delta_\omega} : \mathscr{A} \ni A \mapsto \delta_\omega(A) := \mathbb{1}_A(\omega) \to \{0, 1\}$$

is a probability measure on $\mathscr{A}$. $\delta_\omega \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ is called the *Dirac measure* at the point ω.

(b) Let $\Omega \neq \emptyset$ be countably infinite and let $\mathscr{E} := \{A \in 2^{\Omega} \colon (|A| \wedge |A^{\mathrm{c}}|) \in \mathbb{Z}_{\geqslant 0}\}$. Then $\mathscr{E}$ is an algebra. The set function $\nu : \mathscr{E} \to \{0, \infty\}$ is given by $\nu(A) = 0$ for $A \in \mathscr{E}$ with $|A| \in \mathbb{Z}_{\geqslant 0}$ and $\nu(A) = \infty$ for $|A^{\mathrm{c}}| \in \mathbb{Z}_{\geqslant 0}$. Then ν is a content, but it is not a premeasure. Indeed, ν is not σ-additive, since $\nu(\Omega) = \infty$ and $\sum_{\omega\in\Omega} \nu(\{\omega\}) = 0$.

(c) Let $\Omega \neq \emptyset$ be countable and let $\mathbb{p} : \Omega \to \mathbb{R}_{\geqslant 0}$. Then $\mu : 2^\Omega \to \overline{\mathbb{R}}_{\geqslant 0}$ with $A \mapsto \mu(A) := \sum_{\omega \in \Omega} \mathbb{p}(\omega)\delta_\omega(A)$ is a σ-finite measure on 2^Ω, i.e. $\mu \in \mathfrak{M}_\sigma(2^\Omega)$. We call $\mathbb{p}$ the *mass function* of μ. The number $\mathbb{p}(\omega)$ is called the mass of μ at point ω. Remember, if in addition $\mathbb{p}$ satisfies $\sum_{\omega\in\Omega} \mathbb{p}(\omega) = 1$ then $\mu \in \mathfrak{M}_{\mathrm{pr}}(2^\Omega)$ is a discrete probability measure. If $\mathbb{p}(\omega) = 1$ for every $\omega \in \Omega$, then $\zeta_\Omega := \sum_{\omega\in\Omega} \delta_\omega$ is called *counting measure* on Ω. Evidently, if Ω is finite, then so is $\mu \in \mathfrak{M}_{\mathrm{fi}}(2^\Omega)$. If $\Omega \subseteq \mathbb{R}$ then for each $\omega \in \Omega$ the dirac measure $\delta_\omega \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{B})$, and hence $\mu, \zeta_\Omega \in \mathfrak{M}_\sigma(\mathscr{B})$ are also called discrete measures on $(\mathbb{R}, \mathscr{B})$.

(d) For arbitrary measures $\mu, \nu \in \mathfrak{M}(\mathscr{A})$ the set function

$$\nu + \mu : \mathscr{A} \ni A \mapsto (\nu + \mu)(A) := \nu(A) + \mu(A) \in \overline{\mathbb{R}}_{\geqslant 0}$$

is a measure. □

§01.19 **Lemma**. *Let $\mathscr{E}$ be a semiring and let μ be a content on $\mathscr{E}$. Then the following statements hold.*

(i) *If $\mathscr{E}$ is a ring, then $\mu(A \cup B) = \mu(A) + \mu(B\backslash A)$ and $\mu(B) = \mu(A \cap B) + \mu(B\backslash A)$, hence $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ for any two sets $A, B \in \mathscr{E}$.*

(ii) *μ is monotone. If $\mathscr{E}$ is a ring, then $\mu(B) = \mu(A) + \mu(B\backslash A)$ for any two sets $A, B \in \mathscr{E}$ with $A \subseteq B$.*

(iii) *μ is subadditive. If μ is σ-additive, then μ is also σ-subadditive.*

(iv) *If $\mathscr{E}$ is a ring, then $\sum_{j\in[\![n]\!]} \mu(A_j) = \mu(\biguplus_{j\in[\![n]\!]} A_j) \leqslant \mu(\biguplus_{j\in\mathbb{N}} A_j)$ for all $n \in \mathbb{N}$, and hence $\sum_{j\in\mathbb{N}} \mu(A_j) \leqslant \mu(\biguplus_{j\in\mathbb{N}} A_j)$, for any choice of countably many mutually disjoint sets $A_j \in \mathscr{E}$, $j \in \mathbb{N}$, with $\biguplus_{j\in\mathbb{N}} A_j \in \mathscr{E}$.*

(v) *If $\mathscr{E}$ is a ring, then for any $n \in \mathbb{N}$ and $(A_i)_{i\in[\![n]\!]}$ in $\mathscr{E}$ with $\mu(\bigcup_{i\in[\![n]\!]} A_i) \in \mathbb{R}_{\geqslant 0}$ the Inclusion-exclusion formulas (Poincaré and Sylvester) hold:*

$$\mu\Big(\bigcup_{i\in[\![n]\!]} A_i\Big) = \sum_{\mathcal{I}\in 2^{[\![n]\!]}_{\neq\emptyset}} (-1)^{|\mathcal{I}|-1}\mu\Big(\bigcap_{i\in\mathcal{I}} A_i\Big) \quad \text{and} \quad \mu\Big(\bigcap_{i\in[\![n]\!]} A_i\Big) = \sum_{\mathcal{I}\in 2^{[\![n]\!]}_{\neq\emptyset}} (-1)^{|\mathcal{I}|-1}\mu\Big(\bigcup_{i\in\mathcal{I}} A_i\Big).$$

§01.20 **Proof** of Lemma §01.19. (i), (ii) and (iv) are given in the lecture, (iii) and (v) are exercises. □

§01.21 **Notation**. We agree on the following conventions.

(a) A sequence $(x_n)_{n\in\mathbb{N}}$ in $\overline{\mathbb{R}}$ is called *increasing* (respectively *decreasing*), if $x_n \leqslant x_{n+1}$ (respectively $x_{n+1} \leqslant x_n$) for all $n \in \mathbb{N}$. If an increasing (respectively decreasing) sequence $(x_n)_{n\in\mathbb{N}}$ is convergent, say $x = \lim_{n\to\infty} x_n$, then we write $x_n \uparrow x$ (respectively $x_n \downarrow x$) for short.

(b) A sequence $(A_n)_{n\in\mathbb{N}}$ in 2^Ω is called *increasing* (respectively *decreasing*), if $A_n \subseteq A_{n+1}$ (respectively $A_{n+1} \subseteq A_n$) for all $n \in \mathbb{N}$. We call

$$A_\star := \liminf_{n\to\infty} A_n := \bigcup_{n\in\mathbb{N}} \bigcap_{m\in\mathbb{N}_{\geqslant n}} A_m := \bigcup\Big\{\bigcap\{A_m : m \in \mathbb{N}_{\geqslant n}\} : n \in \mathbb{N}\Big\} \quad \text{and}$$

$$A^\star := \limsup_{n\to\infty} A_n := \bigcap_{n\in\mathbb{N}} \bigcup_{m\in\mathbb{N}_{\geqslant n}} A_m$$

limes inferior, respectively, *limes superior* of the sequence $(A_n)_{n\in\mathbb{N}}$. The sequence $(A_n)_{n\in\mathbb{N}}$ is called *convergent*, if $A_\star = A^\star =: A$. In this case we write $\lim_{n\to\infty} A_n = A$ for short.

An increasing (respectively decreasing) sequence $(A_n)_{n\in\mathbb{N}}$ in 2^Ω is convergent with $A := \lim_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n$ (respectively $A := \lim_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} A_n$). In this case we write $A_n \uparrow A$ (respectively $A_n \downarrow A$).

(c) For functions $f, g : \Omega \to \overline{\mathbb{R}}$ we write *$f \leqslant g$* if $f(\omega) \leqslant g(\omega)$ for any $\omega \in \Omega$. Analogously, we write $f \geqslant 0$ and so on. A sequence $(f_n)_{n\in\mathbb{N}}$ of functions on Ω is called *(pointwise) increasing*, or briefly *isotone* (respectively, *(pointwise) decreasing*, or briefly *antitone*) if $f_n \leqslant f_{n+1}$ (respectively, $f_{n+1} \leqslant f_n$) for all $n \in \mathbb{N}$. We denote by

$$f_\star := \liminf_{n\to\infty} f_n := \sup\{\inf\{f_m : m \in \mathbb{N}_{\geqslant n}\} : n \in \mathbb{N}\} \text{ and}$$
$$f^\star := \limsup_{n\to\infty} f_n := \inf\{\sup\{f_m : m \in \mathbb{N}_{\geqslant n}\} : n \in \mathbb{N}\}$$

the *limes inferior*, respectively, *limes superior*. The sequence $(f_n)_{n\in\mathbb{N}}$ is *convergent* if $f_\star = f^\star =: f$, that is, the pointwise limit exists everywhere. In this case we write $\lim_{n\to\infty} f_n = f$.

An isotone (respectively, antitone) sequence $(f_n)_{n\in\mathbb{N}}$ is convergent with $f := \lim_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} f_n$ (respectively, $f := \lim_{n\to\infty} f_n = \inf_{n\in\mathbb{N}} f_n$). In this case we briefly write $f_n \uparrow f$ (respectively, $f_n \downarrow f$). □.

§01.22 **Definition**. A content μ on a ring $\mathscr{R} \subseteq 2^\Omega$ is called

lower semicontinuous if $\lim_{n\to\infty} \mu(A_n) = \mu(A)$ for any $A \in \mathscr{R}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{R}$ with $A_n \uparrow A$.

upper semicontinuous if $\lim_{n\to\infty} \mu(A_n) = \mu(A)$ for any $A \in \mathscr{R}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{R}$ with $\mu(A_n) \in \mathbb{R}_{\geqslant 0}$ for some (and then eventually all) $n \in \mathbb{N}$ and $A_n \downarrow A$.

$\emptyset$-continuous if $\lim_{n\to\infty} \mu(A_n) = 0 = \mu(\emptyset)$ for any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathscr{R}$ with $\mu(A_n) \in \mathbb{R}_{\geqslant 0}$ for some (and then eventually all) $n \in \mathbb{N}$ and $A_n \downarrow \emptyset$.

§01.23 **Remark**. In the definition of upper semicontinuity, we needed the assumption $\mu(A_n) \in \mathbb{R}_{\geqslant 0}$ since otherwise we would not even have $\emptyset$-continuity for an example as simple as the counting measure $\zeta_{\mathbb{N}}$ on $(\mathbb{N}, 2^{\mathbb{N}})$. Indeed, $A_n := \mathbb{N}_{\geqslant n} \downarrow \emptyset$ but $\zeta_{\mathbb{N}}(A_n) = \infty$ for all $n \in \mathbb{N}$. □

§01.24 **Lemma**. *Let μ be a content on the ring $\mathscr{R} \subseteq 2^\Omega$. Consider the following five properties.* (p1) *μ is σ-additive (and hence $\mu \in \mathfrak{M}(\mathscr{R})$ is a premeasure),* (p2) *μ is σ-subadditive,* (p3) *μ is lower semicontinuous,* (p4) *μ is $\emptyset$-continuous,* (p5) *μ is upper semicontinuous. Then the following implications hold:* (p1)$\Leftrightarrow$(p2)$\Leftrightarrow$(p3)$\Rightarrow$(p4)$\Leftrightarrow$(p5). *If μ is finite, then we also have* (p4)$\Rightarrow$(p3).

§01.25 **Proof** of Lemma §01.24. is given in the lecture. □

§01.26 **Example** *(§01.18 (b) continued)*. ν is a $\emptyset$-continuous content, but it is not a premeasure. □

§01.27 **Definition**.

(a) A pair $(\Omega, \mathscr{A})$ consisting of a nonempty set Ω and a σ-algebra $\mathscr{A} \subseteq 2^\Omega$ is called a *measurable space*. The sets $A \in \mathscr{A}$ are called *measurable sets*. If Ω is at most countably infinite and if $\mathscr{A} = 2^\Omega$, then the measurable space $(\Omega, 2^\Omega)$ is called *discrete*.

(b) A triple $(\Omega, \mathscr{A}, \mu)$ is called *measure space* if $(\Omega, \mathscr{A})$ is a measurable space and $\mu \in \mathfrak{M}(\mathscr{A})$ is a measure on $\mathscr{A}$.

(c) If in addition $\mu(\Omega) = 1$, then $(\Omega, \mathscr{A}, \mu)$ is called a *probability space* and $\mu \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ a *probability measure*. In this case, the sets $A \in \mathscr{A}$ are called *events*. □

End LE01
2026/04/14

§01|03 Measure extension

§01.28 **Lemma** *(Uniqueness)*. *Let $(\Omega, \mathscr{A})$ be a measurable space, let $\mathscr{E} \subseteq \mathscr{A}$ be a $\cap$-closed generator of $\mathscr{A}$ and let $\mu, \nu \in \mathfrak{M}_\sigma(\mathscr{A})$ be two σ-finite measures on $\mathscr{A}$, which agree on $\mathscr{E}$, that is,* (gC) $\mu(E) = \nu(E)$ *for all* $E \in \mathscr{E}$.

Assume in addition that

(uC) *there exist sets* $(\mathcal{E}_n)_{n\in\mathbb{N}}$ *in* $\mathscr{E}$ *with* $\bigcup_{n\in\mathbb{N}} \mathcal{E}_n = \Omega$ *and* $\mu(\mathcal{E}_n) \in \mathbb{R}_{\geqslant 0}$ *for all* $n \in \mathbb{N}$.

Then μ *and* ν *agree also on* $\mathscr{A}$.

If $\mu, \nu \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ *are two probability measures on* $\mathscr{A}$, *then* (uC) *is not needed.*

§01.29 **Proof** of Lemma §01.28. is given in the lecture. □

§01.30 **Remark**. In other words under the assumptions of Lemma §01.28 a σ-finite measure $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ is uniquely determined by its values $\mu(E)$, $E \in \mathscr{E}$. The uniqueness without (uC), the existence of the sequence $(\mathcal{E}_n)_{n\in\mathbb{N}}$, does generally not apply, even if $\mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ is a finite measure on $\mathscr{A}$. In this case the total mass $\mu(\Omega)$ is generally not uniquely determined. Let $\Omega = \{1, 2\}$ and $\mathscr{E} = \{\{1\}\}$. Then $\mathscr{E}$ is a $\cap$-closed generator of 2^Ω. A probability measure $\mu \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ is uniquely determined by the value $\mu(\{1\})$. However, a finite measure is not determined by its value on $\{\{1\}\}$, as $\mu \equiv 0$ and $\nu = \delta_2$ are different finite measures that (gC) agree on $\mathscr{E}$. □

§01.31 **Definition**. A set function $\mu^\star : 2^\Omega \to \overline{\mathbb{R}}_{\geqslant 0}$ is called an *outer measure* if (oM1) $\mu^\star(\emptyset) = 0$, (oM2) $\mu^\star$ is monotone, and (oM3) $\mu^\star$ is σ-subadditive. A set $A \in 2^\Omega$ is called *$\mu^\star$-measurable* if

$$\mu^\star(A \cap B) + \mu^\star(A^c \cap B) = \mu^\star(B) \quad \text{for any } B \in 2^\Omega.$$

We write $\sigma(\mu^\star) := \{A \in 2^\Omega : A \text{ is } \mu^\star\text{-measurable}\}$. □

§01.32 **Remark**. Since $\mu^\star(\emptyset) = 0$ we evidently have $\Omega \in \sigma(\mu^\star)$. As $\mu^\star$ is subadditive it follows that $A \in \sigma(\mu^\star)$ if and only if $\mu^\star(A \cap B) + \mu^\star(A^c \cap B) \leqslant \mu^\star(B)$ for any $B \in 2^\Omega$. □

§01.33 **Lemma**. *Let* $\mathscr{E} \subseteq 2^\Omega$ *be an arbitrary class of sets with* $\emptyset \in \mathscr{E}$ *and let* $\mu : \mathscr{E} \to \overline{\mathbb{R}}_{\geqslant 0}$ *be a set function with* $\mu(\emptyset) = 0$. *For* $A \in 2^\Omega$ *define the set of countable* $\mathscr{E}$*-coverings* $\mathscr{F}$ *of* A*:*

$$\mathcal{U}(A) = \{\mathscr{F} \subseteq \mathscr{E} : \mathscr{F} \text{ is countable and } A \subseteq \textstyle\bigcup_{F\in\mathscr{F}} F\}$$

Define the set function

$$\mu^\star : 2^\Omega \to \overline{\mathbb{R}}_{\geqslant 0} \text{ with } A \mapsto \mu^\star(A) := \begin{cases} \infty & , \mathcal{U}(A) = \emptyset; \\ \inf\{\sum_{F\in\mathscr{F}} \mu(F) : \mathscr{F} \in \mathcal{U}(A)\} & , \mathcal{U}(A) \neq \emptyset. \end{cases}$$

Then $\mu^\star$ *is an outer measure. If in addition* μ *is* σ*-subadditive, then* $\mu^\star$ *and* μ *agree on* $\mathscr{E}$*, i.e.* $\mu^\star(E) = \mu(E)$ *for all* $E \in \mathscr{E}$.

§01.34 **Proof** of Lemma §01.33. is given in the lecture. □

§01.35 **Lemma**. *If* $\mu^\star$ *is an outer measure, then* $\sigma(\mu^\star)$ *is a* σ*-algebra and the restriction of* $\mu^\star$ *on* $\sigma(\mu^\star)$ *is a measure, i.e.* $\mu^\star|_{\sigma(\mu^\star)} \in \mathfrak{M}(\sigma(\mu^\star))$.

§01.36 **Proof** of Lemma §01.35. is given in the lecture. □

§01.37 **Extension theorem for measures**. *Let* $\mathscr{E} \subseteq 2^\Omega$ *be a semiring and let* $\mu : \mathscr{E} \to \overline{\mathbb{R}}_{\geqslant 0}$ *be an additive with* $\mu(\emptyset) = 0$ *(content),* σ*-subadditive and* σ*-finite set function. Then there exists a unique* σ*-finite measure* $\widetilde{\mu} \in \mathfrak{M}_\sigma(\sigma(\mathscr{E}))$ *such that* $\widetilde{\mu}$ *and* μ *agree on* $\mathscr{E}$*, i.e.* $\widetilde{\mu}(E) = \mu(E)$ *for all* $E \in \mathscr{E}$.

§01.38 **Proof** of Theorem §01.37. is given in the lecture. □

§01.39 **Example.**

(a) There exists a uniquely determined measure λ^n on $(\mathbb{R}^n, \mathscr{B}^n)$ with the property that $\lambda^n((a,b]) = \prod_{i\in[\![n]\!]}(b_i - a_i)$ for all $a, b \in \mathbb{R}^n$ with $a < b$. $\boxed{\lambda^n}$ is called *Lebesgue measure* on $(\mathbb{R}^n, \mathscr{B}^n)$ (see lecture Analysis 3).

(b) Let $\mathbb{F} : \mathbb{R} \to \mathbb{R}$ be monotone increasing and right continuous. There is a uniquely determined measure $\mu_{\mathbb{F}}$ on $(\mathbb{R}, \mathscr{B})$ with the property that $\mu_{\mathbb{F}}((a,b]) = \mathbb{F}(b) - \mathbb{F}(a)$ for all $a, b \in \mathbb{R}$ with $a < b$. $\boxed{\mu_{\mathbb{F}}}$ is called *Lebesgue-Stieltjes measure* on $(\mathbb{R}, \mathscr{B})$ (Exercise). If in addition $\lim_{x\to\infty}(\mathbb{F}(x) - \mathbb{F}(-x)) = 1$, then $\mu_{\mathbb{F}}$ is a probability measure. □

§01.40 **Definition.** Let $(\Omega, \mathscr{A}, \mu)$ be a measure space.

(a) A set $N \in \mathscr{A}$ is called a *μ-null set*, or briefly null set, if $\mu(N) = 0$. By

$$\mathcal{N}_\mu := \bigcup \{2^N : N \in \mathscr{A} \text{ is } \mu\text{-null set}\}$$

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

(b) Let $\mathfrak{p}(\omega)$ be a property that a point $\omega \in \Omega$ can have or not have. We say that $\mathfrak{p}$ holds *μ-almost everywhere (μ-a.e.)* if there exists a μ-null set $N \in \mathcal{N}_\mu$ such that $\mathfrak{p}(\omega)$ holds for every $\omega \in \Omega \backslash N = N^c$. If $A \in \mathscr{A}$ and if there exists a μ-null set N such that $\mathfrak{p}(\omega)$ holds for every $\omega \in A\backslash N$, then we say that $\mathfrak{p}$ holds μ-almost everywhere on A. If $\mu = \mathbb{P} \in \mathfrak{M}_{\text{pr}}(\mathscr{A})$ is a probability measure then we say that $\mathfrak{p}$ holds *$\mathbb{P}$-almost surely ($\mathbb{P}$-a.s.)* respectively $\mathbb{P}$-almost surely on A.

(c) The measure space $(\Omega, \mathscr{A}, \mu)$ is called *complete*, if $\mathcal{N}_\mu \subseteq \mathscr{A}$. □

§01.41 **Remark.** Let $(\Omega, \mathscr{A}, \mu)$ be a σ-finite measure space. Due to Theorem §01.37 there exists a unique smallest σ-algebra $\tilde{\mathscr{A}} \supseteq \mathscr{A}$ and an extension $\tilde{\mu}$ of μ to $\tilde{\mathscr{A}}$ such that $(\Omega, \tilde{\mathscr{A}}, \tilde{\mu})$ is complete where $\tilde{\mathscr{A}} = \sigma(\mathscr{E})$ with $\mathscr{E} = \mathscr{A} \cup \mathcal{N}_\mu = \{A \cup N : A \in \mathscr{A}, N \in \mathcal{N}_\mu\}$ and $\tilde{\mu}(A \cup N) = \mu(A)$ for any $A \in \mathscr{A}$ and $N \in \mathcal{N}_\mu$. □

§01.42 **Definition.** Let $(\Omega, \mathscr{A}, \mu)$ be a measure space and $B \in \mathscr{A}$. On the trace σ-algebra $\mathscr{A}_B$ we define a measure by $\mu_B(A) := \mu(A)$ for $A \in \mathscr{A}$ with $A \subseteq B$. This measure is called the *restriction* of μ to B. □

§01.43 **Example.** The restriction $\lambda_{[0,1]}$ of the Lebesgue-Borel measure λ on $(\mathbb{R}, \mathscr{B})$ to $[0,1]$ is a probability measure on $([0,1], \mathscr{B}_{[0,1]})$, i.e. $\lambda_{[0,1]} \in \mathfrak{M}_{\text{pr}}(\mathscr{B}_{[0,1]})$. More generally, for a Borel set $B \in \mathscr{B}$ we call the restriction λ_B the Lebesgue measure on B, i.e. $\lambda_B \in \mathfrak{M}_\sigma(\mathscr{B}_B)$. □

End LE02
2026/04/17

§02 Integration theory

§02|01 The integral

§02.01 **Reminder.** Let $(\Omega, \mathscr{A}, \mu)$ be a measure space and let $(\mathcal{S}, \mathscr{S})$ be a measurable space.

(a) A function $f : \Omega \to \mathcal{S}$ is called *$\mathscr{A}$-$\mathscr{S}$-measurable* (or, briefly, *measurable*) if

$$\sigma(f) := f^{-1}(\mathscr{S}) := \{f^{-1}(S) : S \in \mathscr{S}\} \subseteq \mathscr{A}.$$

If f is measurable, we write $f : (\Omega, \mathscr{A}) \to (\mathcal{S}, \mathscr{S})$. We denote by $\mathcal{M}(\mathscr{A}, \mathscr{S})$ the set of all $\mathscr{A}$-$\mathscr{S}$-measurable functions. If $\mathscr{S} = \mathscr{B}_{\mathcal{S}}$ is the Borel σ-algebra on $\mathcal{S}$ then we write $\mathcal{M}_{\mathcal{S}}(\mathscr{A}) := \mathcal{M}(\mathscr{A}, \mathscr{B}_{\mathcal{S}})$ for short. If $\mu = \mathbb{P} \in \mathfrak{M}_{\text{pr}}(\mathscr{A})$ is a probability measure then $f \in \mathcal{M}(\mathscr{A}, \mathscr{S})$ is called ($(\mathcal{S}, \mathscr{S})$-valued) *random variable*. The σ-algebra $\sigma(f)$ is called the σ-algebra on Ω that is *generated* by f. This is the smallest σ-algebra with respect to which f is measurable.

(b) The identity map $\mathrm{id}_\Omega : \Omega \to \Omega$ is $\mathscr{A}$-$\mathscr{A}$-measurable. If $\mathscr{A} = 2^\Omega$ or $\mathscr{S} = \{\emptyset, \mathcal{S}\}$, then any map $f : \Omega \to \mathcal{S}$ belongs to $\mathcal{M}(\mathscr{A},\mathscr{S})$. The indicator function $\mathbb{1}_A$ for $A \in 2^\Omega$ belongs to $\mathcal{M}(\mathscr{A}, 2^{\{0,1\}})$ if and only if $A \in \mathscr{A}$.

(c) A measurable function $f : (\Omega, \mathscr{A}) \to (\mathcal{S}, \mathscr{S})$ is called

numerical if $(\mathcal{S}, \mathscr{S}) = (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, briefly $f \in \overline{\mathcal{M}}(\mathscr{A}) := \mathcal{M}_{\overline{\mathbb{R}}}(\mathscr{A}) = \mathcal{M}(\mathscr{A}, \overline{\mathscr{B}})$,

positive numerical if $(\mathcal{S}, \mathscr{S}) = (\overline{\mathbb{R}}_{\geqslant 0}, \overline{\mathscr{B}}_{\geqslant 0})$, briefly $f \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) := \mathcal{M}_{\overline{\mathbb{R}}_{\geqslant 0}}(\mathscr{A}) = \mathcal{M}(\mathscr{A}, \overline{\mathscr{B}}_{\geqslant 0})$,

real if $(\mathcal{S}, \mathscr{S}) = (\mathbb{R}, \mathscr{B})$, briefly $f \in \mathcal{M}(\mathscr{A}) := \mathcal{M}_{\mathbb{R}}(\mathscr{A}) = \mathcal{M}(\mathscr{A}, \mathscr{B})$,

positive real if $(\mathcal{S}, \mathscr{S}) = (\mathbb{R}_{\geqslant 0}, \mathscr{B}_{\geqslant 0})$, briefly $f \in \mathcal{M}_{\geqslant 0}(\mathscr{A}) := \mathcal{M}_{\mathbb{R}_{\geqslant 0}}(\mathscr{A}) = \mathcal{M}(\mathscr{A}, \mathscr{B}_{\geqslant 0})$.

For any $B_1, B_2 \in \overline{\mathscr{B}}$ with $B_1 \subseteq B_2$ we agree on $\mathcal{M}_{B_1}(\mathscr{A}) \subseteq \mathcal{M}_{B_2}(\mathscr{A})$, e.g. $\mathcal{M}_{(0,1]}(\mathscr{A}) \subseteq \mathcal{M}_{\geqslant 0}(\mathscr{A}) \subseteq \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ or $\mathcal{M}(\mathscr{A}) \subseteq \overline{\mathcal{M}}(\mathscr{A})$. If the preimage $(\Omega, \mathscr{A})$ is irrelevant we also write shortly $\overline{\mathcal{M}} := \overline{\mathcal{M}}(\mathscr{A})$, $\overline{\mathcal{M}}_{\geqslant 0} := \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, $\mathcal{M} := \mathcal{M}(\mathscr{A})$, and $\mathcal{M}_{\geqslant 0} := \mathcal{M}_{\geqslant 0}(\mathscr{A})$. If $(f_n)_{n\in\mathbb{N}}$ is a sequence in $\overline{\mathcal{M}}$, then $\sup_{n\in\mathbb{N}} f_n$, $\inf_{n\in\mathbb{N}} f_n$, $f_\star := \liminf_{n\to\infty} f_n$, and $f^\star := \limsup_{n\to\infty} f_n$ belong also to $\overline{\mathcal{M}}$ (see lecture EWS).

(d) A real map $f \in \mathcal{M}(\mathscr{A})$ assuming only finitely many values is called *simple* or *elementary*. If $f \in \mathcal{M}(\mathscr{A})$ is simple then there is an $n \in \mathbb{N}$ and mutually disjoint measurable sets $(A_j)_{i\in[\![n]\!]}$ in $\mathscr{A}$ as well as numbers $(a_j)_{i\in[\![n]\!]}$ in $\mathbb{R}$ such that $f = \sum_{i\in[\![n]\!]} a_j \mathbb{1}_{A_j}$. We denote by $\mathcal{M}^{\mathrm{sim}}(\mathscr{A})$ and $\mathcal{M}^{\mathrm{sim}}_{\geqslant 0}(\mathscr{A})$ the set of all simple, respectively, positive simple functions on $(\Omega, \mathscr{A})$. If $f = \sum_{i\in[\![n]\!]} a_j \mathbb{1}_{A_j}$ and $f = \sum_{i\in[\![m]\!]} b_j \mathbb{1}_{B_j}$ are two representations of $f \in \mathcal{M}^{\mathrm{sim}}_{\geqslant 0}(\mathscr{A})$, then $\sum_{i\in[\![n]\!]} a_j \mu(A_j) = \sum_{i\in[\![m]\!]} b_j \mu(B_j)$ (check it!).

(e) Let $f \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ be positive numerical. Then there exists an isotone sequence of simple functions $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{M}^{\mathrm{sim}}_{\geqslant 0}(\mathscr{A})$ such that $f_n \uparrow f$ (see lecture EWS).

(f) For functions $f, g \in \overline{\mathcal{M}}(\mathscr{A})$ we write $f \leqslant g$ if $f(\omega) \leqslant g(\omega)$ for any $\omega \in \Omega$ and $f \leqslant g$ μ-a.e. if there is $N \in \mathcal{N}_\mu$ such that $f(\omega) \leqslant g(\omega)$ for any $\omega \in N^c$ (see Definition §01.40). □

§02.02 **Theorem**. *For each measure μ on a measurable space $(\Omega, \mathscr{A})$ we call integral with respect to μ the uniquely determined functional $\mathbb{I}_\mu : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ satisfying the following properties:*

(I1) $\mathbb{I}_\mu(af + bg) = a\mathbb{I}_\mu(f) + b\mathbb{I}_\mu(g)$ *for all* $f, g \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ *and* $a, b \in \mathbb{R}_{\geqslant 0}$, (linearity)

(I2) $\mathbb{I}_\mu(f_n) \uparrow \mathbb{I}_\mu(f)$ *for all* $(f_n)_{n\in\mathbb{N}} \uparrow f$ *in* $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, (monotone convergence)

(I3) $\mathbb{I}_\mu(\mathbb{1}_A) = \mu(A)$ *for all* $A \in \mathscr{A}$. (normed)

For each $f \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ we call $\int f \,\mathrm{d}\mu := \mathbb{I}_\mu(f)$ *the integral of f with respect to μ. For $A \in \mathscr{A}$ we write shortly* $\int_A f \,\mathrm{d}\mu := \int (f\mathbb{1}_A)\,\mathrm{d}\mu$. *$f$ is called μ-integrable, if $\int f \,\mathrm{d}\mu \in \mathbb{R}_{\geqslant 0}$.* □

§02.03 **Proof** of Theorem §02.02. The theorem summarises the main result of this section; its proof takes place in several steps. We first show in Theorem §02.05 the uniqueness result and then explicitly state in Theorem §02.09 a functional $\mathbb{I}_\mu : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ for which we verify the required conditions (I1)-(I3). In summary, we then show therewith in Theorem §02.09 the existence result. □

§02.04 **Notation**. For $f \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ and $A \in \mathscr{A}$ we write shortly $\mu(f) := \int f \,\mathrm{d}\mu = \int_\Omega f(\omega)\mu(\mathrm{d}\omega)$ as well as $\mu(f\mathbb{1}_A) = \int_A f \,\mathrm{d}\mu = \int_A f(\omega)\mu(\mathrm{d}\omega)$. If $\mathbb{P} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ is a probability measure then we call the uniquely determined functional $\mathbb{E} := \mathbb{I}_{\mathbb{P}} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ *expectation* with respect to $\mathbb{P}$. □

§02.05 **Uniqueness theorem**. *The integral is uniquely determined.*

§02.06 **Proof** of Theorem §02.05. is given in the lecture. □

Reminder §02.01 (d) allows the following definition to be made since the defined value $\widetilde{\mathbb{I}}_\mu(f)$ does not depend on the chosen representation of f.

§02.07 **Lemma**. *The map $\widetilde{\mathbb{I}}_\mu : \mathcal{M}^{\text{sim}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ given by*

$$f = \sum_{i\in[\![n]\!]} a_j \mathbb{1}_{A_j} \mapsto \widetilde{\mathbb{I}}_\mu(f) := \sum_{i\in[\![n]\!]} a_j \mu(A_j).$$

is normed, positive, linear and monotone:

(i) $\widetilde{\mathbb{I}}_\mu(\mathbb{1}_A) = \mu(A)$ *for every* $A \in \mathscr{A}$, (normed)

(ii) $\widetilde{\mathbb{I}}_\mu(af + bg) = a\widetilde{\mathbb{I}}_\mu(f) + b\widetilde{\mathbb{I}}_\mu(g)$ *for all* $f, g \in \mathcal{M}^{\text{sim}}_{\geqslant 0}(\mathscr{A})$ *and* $a, b \in \mathbb{R}_{\geqslant 0}$, (linearity)

(iii) $\widetilde{\mathbb{I}}_\mu(f) \leqslant \widetilde{\mathbb{I}}_\mu(g)$ *for all* $f, g \in \mathcal{M}^{\text{sim}}_{\geqslant 0}(\mathscr{A})$ *with* $f \leqslant g$. (monotonicity).

§02.08 **Proof** of Lemma §02.07. Exercise. □

§02.09 **Existence theorem**. *The functional $\mathbb{I}_\mu : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ with*

$$f \mapsto \mathbb{I}_\mu(f) := \sup\{\widetilde{\mathbb{I}}_\mu(g) : g \in \mathcal{M}^{\text{sim}}_{\geqslant 0}(\mathscr{A}), g \leqslant f\}$$

is an integral with respect to μ, that is, it shares the properties (I1)-(I3) *in Theorem §02.02:*

(i) $\mathbb{I}_\mu(\mathbb{1}_A) = \mu(A)$ *for every* $A \in \mathscr{A}$, (normed)

(ii) $\mathbb{I}_\mu(f) \leqslant \mathbb{I}_\mu(g)$ *for all* $f, g \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ *with* $f \leqslant g$, (monotonicity)

(iii) $\mathbb{I}_\mu(f_n) \uparrow \mathbb{I}_\mu(f)$ *for all* $(f_n)_{n\in\mathbb{N}} \uparrow f$ *in* $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$. (monotone convergence)

(iv) $\mathbb{I}_\mu(af + bg) = a\mathbb{I}_\mu(f) + b\mathbb{I}_\mu(g)$ *for all* $f, g \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ *and* $a, b \in \overline{\mathbb{R}}_{\geqslant 0}$ (linearity)
(with convention $\infty \cdot 0 = 0$).

§02.10 **Proof** of Theorem §02.09. is given in the lecture. □

§02.11 **Remark**. By Lemma §02.07 (iii) we have the identity $\mathbb{I}_\mu(f) = \widetilde{\mathbb{I}}_\mu(f)$ for any $f \in \mathcal{M}^{\text{sim}}_{\geqslant 0}(\mathscr{A})$. Hence $\mathbb{I}_\mu$ is an extension of the map $\widetilde{\mathbb{I}}_\mu$ from $\mathcal{M}^{\text{sim}}_{\geqslant 0}(\mathscr{A})$ to the set of positive numerical functions $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$. □

§02.12 **Comment**. A measurable partition $\mathcal{P} := \{A_i : i \in \mathcal{I}\} \subseteq \mathscr{A}_{\neq\varnothing}$ of Ω is finite, if $|\mathcal{I}| \in \mathbb{N}$, and hence $\varnothing \neq A \in \mathscr{A}$ for each $A \in \mathcal{P}$. If we set $\mathscr{P} := \{\mathcal{P} \subseteq \mathscr{A}_{\neq\varnothing} : \mathcal{P}$ finite, measurable partition of $\Omega\}$, then the functional $\mathbb{I}_\mu : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ given by (with convention $\infty \cdot 0 = 0$)

$$f \mapsto \mathbb{I}_\mu(f) := \sup\Big\{\sum_{A\in\mathcal{P}} \big(\inf_{\omega\in A} f(w)\big)\mu(A) : \mathcal{P} \in \mathscr{P}\Big\}$$

shares also the properties (I1)-(I3) in Theorem §02.02, and hence it is an alternative but equivalent representation of the uniquely determined integral with respect to μ. □

§02.13 **Notation**. For arbitrary measures $\mu, \nu \in \mathfrak{M}(\mathscr{A})$ we write $\nu \leqslant \mu$ if $\nu(A) \leqslant \mu(A)$ for all $A \in \mathscr{A}$. Evidently, $\nu \leqslant \mu$ and $\mu \leqslant \nu$ imply together $\mu = \nu$. □

§02.14 **Lemma** *(Properties)*. *Let $(\Omega, \mathscr{A}, \mu)$ be an arbitrary measure space and let $(f_n)_{n\in\mathbb{N}}$ be a sequence in $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$.*

(i) *(Fatou's lemma)* $\mu(\liminf_{n\to\infty} f_n) = \int \big(\liminf_{n\to\infty} f_n\big)\,\mathrm{d}\mu \leqslant \liminf_{n\to\infty} \int f_n\,\mathrm{d}\mu = \liminf_{n\to\infty} \mu(f_n)$ *and in particular* $\mu\big(\liminf_{n\to\infty} A_n\big) \leqslant \liminf_{n\to\infty} \mu(A_n)$ *for every sequence* $(A_n)_{n\in\mathbb{N}}$ *of sets in* $\mathscr{A}$. *If* $\mu \in \mathfrak{M}_{\mathrm{f}}(\mathscr{A})$ *is finite, then also* $\limsup_{n\to\infty} \mu(A_n) \leqslant \mu\big(\limsup_{n\to\infty} A_n\big)$.

(ii) $\sum_{n\in\mathbb{N}} f_n \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ *and* $\mu(\sum_{n\in\mathbb{N}} f_n) = \int\big(\sum_{n\in\mathbb{N}} f_n\big)\,\mathrm{d}\mu = \sum_{n\in\mathbb{N}}\int f_n\,\mathrm{d}\mu = \sum_{n\in\mathbb{N}}\mu(f_n)$.

Let in addition $f, g \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$.

(iii) $f = 0$ μ-a.e. *if and only if* $\mu(f) = \int f\,\mathrm{d}\mu = 0$. *If* $\mu(f) \in \mathbb{R}_{\geqslant 0}$ *then* $f \in \mathbb{R}_{\geqslant 0}$ μ-a.e. *and the restriction of* μ *on* $\{f \neq 0\}$ *is a* σ*-finite measure.*

(iv) *The set function*

$$\boxed{f\mu} : \mathscr{A} \ni A \mapsto f\mu(A) := \mu(\mathbb{1}_A f) = \int(\mathbb{1}_A f)\,\mathrm{d}\mu \in \overline{\mathbb{R}}_{\geqslant 0}$$

is a measure on $(\Omega, \mathscr{A})$. *For all* $A \in \mathscr{A}$ *with* $\mu(A) = 0$ *we have* $f\mu(A) = 0$.

(v) *If* $f \leqslant g$ *(respectively* $f = g$*)* μ-a.e. *then* $f\mu \leqslant g\mu$ *(respectively* $f\mu = g\mu$*). The converse holds, if* (c1) f *is* μ*-integrable, or* (c2) $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$, *or* (c3) $g\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ σ*-finite.* *In particular,* $\mu(f) = \int f\,\mathrm{d}\mu \leqslant \int g\,\mathrm{d}\mu = \mu(g)$ *(respectively,* $\mu(f) = \mu(g)$*).*

(vi) $\mu \in \mathfrak{M}(\mathscr{A})$ *is* σ*-finite if and only if there is* $h \in \mathcal{M}_{(0,1]}(\mathscr{A})$ *with* $\mu(h) \in \mathbb{R}_{\geqslant 0}$ *(*μ*-integrable). In particular, for each* σ*-finite* $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ *there exists* $h \in \mathcal{M}_{(0,1]}(\mathscr{A})$ *such that* $h\mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ *is finite and* $h\mu$ *shares the same null-sets as* μ. End LE03 2026/04/22

(vii) $\sum_{n\in\mathbb{N}}\mu\big(\{f \geqslant n\}\big) \leqslant \mu(f) \leqslant \sum_{n\in\mathbb{N}_0}\mu\big(\{f > n\}\big)$ *and* $\mu(f) = \int_0^\infty \mu\big(\{f \geqslant t\}\big)\mathrm{d}t$ *for every* $f \in \mathcal{M}(\mathscr{A})$ *with* $f \in \mathbb{R}_{\geqslant 0}$ μ-a.e.

§02.15 **Proof** of Lemma §02.14. is given in the lecture. □

§02.16 **Definition**. Let $\mu \in \mathfrak{M}(\mathscr{A})$ be a measure on $(\Omega, \mathscr{A})$ and let $\mathfrak{f} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$. Define the measure $\nu \in \mathfrak{M}(\mathscr{A})$ by $\nu(A) := \mu(\mathbb{1}_A\mathfrak{f})$ for $A \in \mathscr{A}$. We say that $\boxed{\mathfrak{f}\mu} := \nu$ has the *density* $\mathrm{d}\nu/\mathrm{d}\mu := \mathfrak{f}$ with respect to μ, or briefly *μ-density*. □

§02.17 **Lemma** *(Properties)*. *Let* $(\Omega, \mathscr{A}, \mu)$ *be a measure space and let* $\nu := \mathfrak{f}\mu \in \mathfrak{M}(\mathscr{A})$ *admit the density* $\mathrm{d}\nu/\mathrm{d}\mu = \mathfrak{f} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$.

(i) $\nu(g) = \int g\,\mathrm{d}\nu = \int(g\mathfrak{f})\,\mathrm{d}\mu = \mu(g\mathfrak{f}) = \mathfrak{f}\mu(g)$ *for every* $g \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$.

(ii) $\rho = \mathfrak{q}\nu = \mathfrak{q}(\mathfrak{f}\mu) = (\mathfrak{q}\mathfrak{f})\mu$ *for every* $\rho := \mathfrak{q}\nu \in \mathfrak{M}(\mathscr{A})$ *with* $\mathfrak{q} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$.

(iii) *If* $\nu \in \mathfrak{M}_\sigma(\mathscr{A})$ *or* $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ *is* σ*-finite then the* μ*-density* $\mathrm{d}\nu/\mathrm{d}\mu = \mathfrak{f}$ *of* ν *is unique up to equality* μ*-almost everywhere.*

(iv) *If* $\nu \in \mathfrak{M}_\sigma(\mathscr{A})$ *is* σ*-finite, then* $\mathrm{d}\nu/\mathrm{d}\mu = \mathfrak{f} \in \mathbb{R}_{\geqslant 0}$ μ-a.e.. *The converse holds, if* $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$.

§02.18 **Proof** of Lemma §02.17. is given in the lecture. □

§02.19 **Notation**. If $f \in \overline{\mathcal{M}}(\mathscr{A})$ is numerical then $f^+ := f \vee 0$, $f^- := (-f)^+$, $|f| = f^+ + f^- \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ are positive numerical. □

§02.20 **Definition**. Let $(\Omega, \mathscr{A}, \mu)$ be a measure space and let $f \in \overline{\mathcal{M}}(\mathscr{A})$ be numerical.

(a) If f^+ or f^- is μ-integrable, that is, $\mu(f^+) \wedge \mu(f^-) \in \mathbb{R}_{\geqslant 0}$, then we define the *integral*

$$\mu(f) := \boxed{\int f\,\mathrm{d}\mu} := \int f^+\,\mathrm{d}\mu - \int f^-\,\mathrm{d}\mu = \mu(f^+) - \mu(f^-)$$

of f with respect to μ where we use the usual conventions $\infty + x = \infty$ and $-\infty + x = -\infty$ for all $x \in \mathbb{R}$. In this case f is called *μ-quasiintegrable*. The integral of f is not defined, if $\mu(f^+) = \infty = \mu(f^-)$.

(b) If $\mu(|f|) \in \mathbb{R}_{\geqslant 0}$, that is, $\mu(f^+) \vee \mu(f^-) \in \mathbb{R}_{\geqslant 0}$, then f is called *μ-integrable*. The vector space of all μ-integrable *real* functions is denoted by

$$\mathcal{L}^1 := \mathcal{L}^1(\mu) := \mathcal{L}^1(\Omega, \mathscr{A}, \mu) := \{f \in \mathcal{M}(\mathscr{A}) \colon \mu(|f|) \in \mathbb{R}_{\geqslant 0}\}.$$

The set of all μ-integrable *numerical* functions is denoted by

$$\bar{\mathcal{L}}^1 := \bar{\mathcal{L}}^1(\mu) := \bar{\mathcal{L}}^1(\Omega, \mathscr{A}, \mu) := \{f \in \overline{\mathcal{M}}(\mathscr{A}) \colon \mu(|f|) \in \mathbb{R}_{\geqslant 0}\}.$$

(c) For $s \in \mathbb{R}_{>0}$ define

$$\|f\|_{\mathcal{L}^s} := \big(\mu(|f|^s)\big)^{1/s} \quad \text{and} \quad \|f\|_{\mathcal{L}^\infty} := \inf\{x \in \mathbb{R}_{\geqslant 0} \colon \mu(\{|f| > x\}) = 0\}.$$

For $s \in \overline{\mathbb{R}}_{>0}$, f is called *$\mathcal{L}^s$-integrable* if $\|f\|_{\mathcal{L}^s} \in \mathbb{R}_{\geqslant 0}$. The vector space of all $\mathcal{L}^s$-integrable *real* functions we denote by

$$\mathcal{L}^s := \mathcal{L}^s(\mu) := \mathcal{L}^s(\Omega, \mathscr{A}, \mu) := \{f \in \mathcal{M}(\mathscr{A}) \colon \|f\|_{\mathcal{L}^s} \in \mathbb{R}_{\geqslant 0}\}.$$

The set of all $\mathcal{L}^s$-integrable *numerical* functions we denote by

$$\bar{\mathcal{L}}^s := \bar{\mathcal{L}}^s(\mu) := \bar{\mathcal{L}}^s(\Omega, \mathscr{A}, \mu) := \{f \in \overline{\mathcal{M}}(\mathscr{A}) \colon \|f\|_{\mathcal{L}^s} \in \mathbb{R}_{\geqslant 0}\}.$$

For $s \in \overline{\mathbb{R}}_{\geqslant 1}$, the map $\|\cdot\|_{\mathcal{L}^s}$ is a seminorm on $\mathcal{L}^s(\mu)$ (see Subsection §02|03 below), that is, for all $f, g \in \mathcal{L}^s(\mu)$ and $a \in \mathbb{R}$, (s1) $\|af\|_{\mathcal{L}^s} = |a|\|f\|_{\mathcal{L}^s}$, (s2) $\|f+g\|_{\mathcal{L}^s} \leqslant \|f\|_{\mathcal{L}^s} + \|g\|_{\mathcal{L}^s}$, (s3) $\|f\|_{\mathcal{L}^s} \in \mathbb{R}_{\geqslant 0}$ and $\|f\|_{\mathcal{L}^s} = 0$ if $f = 0$ μ-a.e.

(d) The map $\langle \cdot, \cdot \rangle_{\mathcal{L}^2} : \mathcal{L}^2(\mu) \times \mathcal{L}^2(\mu) \to \mathbb{R}$ with $(f,g) \mapsto \langle f, g \rangle_{\mathcal{L}^2} := \mu(fg)$ is a positive semidefinite symmetric bilinearform. □

§02.21 **Lemma** *(Properties)*. *Let $f, g \in \bar{\mathcal{L}}^1(\Omega, \mathscr{A}, \mu)$.*

(i) *If $af + bg \in \overline{\mathcal{M}}(\mathscr{A})$ for $a, b \in \mathbb{R}$, then $af + bg \in \bar{\mathcal{L}}^1(\Omega, \mathscr{A}, \mu)$ and $\mu(af+bg) = a\mu(f) + b\mu(g)$.* *(linear)*

(ii) *Let $h \in \overline{\mathcal{M}}(\mathscr{A})$. If $h = f$ μ-a.e., then h is μ-integrable and $\mu(h) = \mu(f)$.*
If $|h| \leqslant |g|$ μ-a.e. then h is μ-integrable.

(iii) *If $f \leqslant g$ μ-a.e., then $\mu(f) \leqslant \mu(g)$.* *(monotone)*
In particular, if $g \in \overline{\mathbb{R}}_{\geqslant 0}$ μ-a.e., then $\mu(g) \in \mathbb{R}_{\geqslant 0}$ and *(positive)*
$\mu(\{g \geqslant \varepsilon\}) \leqslant \frac{1}{\varepsilon}\mu(g)$ for all $\varepsilon \in \mathbb{R}_{>0}$. *(Markov inequality)*

(iv) *$|\mu(f)| \leqslant \mu(|f|)$.* *(triangle inequality)*

(v) *$f = 0$ μ-a.e. if and only if $\mu(f\mathbb{1}_A) = 0$ for all $A \in \mathscr{A}$.*

(vi) *If $\mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ is finite and $h \in \mathcal{M}(\mathscr{A})$ is bounded, hence $\|h\|_\infty := \sup_{\omega\in\Omega}|h(\omega)| \in \mathbb{R}_{\geqslant 0}$, then $h \in \mathcal{L}^1(\mu)$.*

(vii) *If $\mu, \nu \in \mathfrak{M}(\mathscr{A})$ then $h \in \bar{\mathcal{L}}^1(\mu+\nu)$ if and only if $h \in \bar{\mathcal{L}}^1(\mu) \cap \bar{\mathcal{L}}^1(\nu)$. In this case, $(\mu+\nu)(h) = \mu(h) + \nu(h)$.*

(viii) *If $\nu = \mathfrak{f}\mu$ with $\mathrm{d}\nu/\mathrm{d}\mu = \mathfrak{f} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ then $g \in \overline{\mathcal{M}}(\mathscr{A})$ is ν-(quasi)integrable if and only if $g\mathfrak{f} \in \overline{\mathcal{M}}(\mathscr{A})$ is μ-(quasi)integrable. In this case $\nu(g) = \mu(g\mathfrak{f}) = \int (g\mathfrak{f})\,\mathrm{d}\mu = \int g\,\mathrm{d}(\mathfrak{f}\mu) = \int g\,\mathrm{d}\nu$.*

§02.22 **Proof** of Lemma §02.21. is given in the lecture. □

§02.23 **Corollary** *(Properties)*. *Let $f, g \in \overline{\mathcal{M}}(\mathscr{A})$ and $\mu \in \mathfrak{M}(\mathscr{A})$.*

(i) *Let $s \in \mathbb{R}_{>0}$. f is $\mathcal{L}^s(\mu)$-integrable if and only if $|f|^s$ is μ-integrable. Moreover, if f is $\mathcal{L}^\infty(\mu)$-integrable then $\mu\big(\{|f| > \|f\|_{\mathcal{L}^\infty}\}\big) = 0$.*

(ii) *Let $s \in \overline{\mathbb{R}}_{>0}$. $\|f\|_{\mathcal{L}^s} = 0$ if and only if $f = 0$ μ-a.e.. If $a \in \mathbb{R}$ then $\|af\|_{\mathcal{L}^s} = |a|\|f\|_{\mathcal{L}^s}$. If f is $\mathcal{L}^s$-integrable and $f = g$ μ-a.e., then $|f| \in \mathbb{R}_{\geqslant 0}$ μ-a.e. and $\|f\|_{\mathcal{L}^s} = \|g\|_{\mathcal{L}^s}$.*

§02.24 **Proof** of Corollary §02.23. Exercise. □

§02.25 **Lemma**. *Let $(\Omega, \mathscr{A}, \mu)$ be an arbitrary measure space and $f, g \in \overline{\mathcal{M}}(\mathscr{A})$.*

(i) *(Hölder's inequality) Let $s, r \in \overline{\mathbb{R}}_{\geqslant 1}$ with $\frac{1}{s} + \frac{1}{r} = 1$. Then $\mu(|fg|) \leqslant \|f\|_{\mathcal{L}^s}\|g\|_{\mathcal{L}^r}$.*
(Cauchy-Schwarz inequality) If $f, g \in \mathcal{L}^2$ then $|\langle f, g\rangle_{\mathcal{L}^2}| \leqslant \|f\|_{\mathcal{L}^2}\|g\|_{\mathcal{L}^2}$.

(ii) *If $\mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ is finite, $s \in \overline{\mathbb{R}}_{>0}$ and $r \in (0, s)$. Then $\mu(\Omega)^{1/s}\|f\|_{\mathcal{L}^r(\mu)} \leqslant \mu(\Omega)^{1/r}\|f\|_{\mathcal{L}^s(\mu)}$ and hence $\mathcal{L}^s(\mu) \subseteq \mathcal{L}^r(\mu)$ (and $\bar{\mathcal{L}}^s(\mu) \subseteq \bar{\mathcal{L}}^r(\mu)$).*

(iii) *(Minkowski's inequality) For $f + g \in \overline{\mathcal{M}}(\mathscr{A})$ and any $s \in \overline{\mathbb{R}}_{\geqslant 1}$, $\|f + g\|_{\mathcal{L}^s} \leqslant \|f\|_{\mathcal{L}^s} + \|g\|_{\mathcal{L}^s}$.*

§02.26 **Proof** of Lemma §02.25. is given in the lecture. □ End LE04 2026/04/24

§02.27 **Lemma** *(Image measure)*. *Let $(\mathcal{X}, \mathscr{X})$ and $(\Omega, \mathscr{A})$ ... be measurable spaces* — *Let $(\Omega, \mathscr{A})$ and $(\mathcal{X}, \mathscr{X})$ be measurable spaces, let $\mu \in \mathfrak{M}(\mathscr{A})$ be a measure and let $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$ be measurable. Let $\mu^X := \mu \circ X^{-1} \in \mathfrak{M}(\mathscr{X})$ be the image measure on $(\mathcal{X}, \mathscr{X})$ of μ under the map X. If $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X})$ then $\mu(h(X)) = \mu^X(h)$. Consequently, $h \in \overline{\mathcal{M}}(\mathscr{X})$ is μ^X-(quasi)integrable if and only if $h(X) \in \overline{\mathcal{M}}(\mathscr{A})$ is μ-(quasi)integrable. In this case, $\mu(h(X)) = \mu^X(h)$.*

Fig. 01 [§02] Measurability implications.

$$\begin{tikzcd}
(\Omega, \mathscr{A}) \arrow[r, "X"] \arrow[dr, swap, "{h(X) \in \overline{\mathcal{M}}(\mathscr{A})\ \mu\text{-(quasi)integrable}}"] & (\mathcal{X}, \mathscr{X}) \arrow[d, "{h \in \overline{\mathcal{M}}(\mathscr{X})\ \mu^X\text{-(quasi)integrable}}"] \\
 & (\overline{\mathbb{R}}, \overline{\mathscr{B}})
\end{tikzcd}$$

In particular, if X is a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$, then

$$\int_{\mathcal{X}} h(x)\mathbb{P}^X(\mathrm{d}x) = \int h \,\mathrm{d}\mathbb{P}^X = \mathbb{P}^X(h) = \mathbb{P}(h(X)) = \int h(X)\,\mathrm{d}\mathbb{P} = \int_\Omega h(X(\omega))\mathbb{P}(\mathrm{d}\omega).$$

§02.28 **Proof** of Lemma §02.27. is given in the lecture. □

§02|02 Convergence criteria

§02.29 **Definition**. Let $(\Omega, \mathscr{A}, \mu)$ be a measure space. We say that a sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{M}(\mathscr{A})$ converges to $f \in \overline{\mathcal{M}}(\mathscr{A})$

μ-almost everywhere (μ-a.e.), symbolically $f_n \xrightarrow{\mu\text{-a.e.}} f$, if $\limsup_{n\to\infty}|f_n - f| = 0$ μ-a.e., that is, there exists a μ-null set $N \in \mathscr{A}$ such that $\lim_{n\to\infty}|f_n(\omega) - f(\omega)| = 0$ for any $\omega \in N^c := \Omega \backslash N$.

μ-almost complete (μ-a.c.), symbolically $f_n \xrightarrow{\mu\text{-a.c.}} f$, if $\sum_{n\in\mathbb{N}} \mu\big(\{|f_n - f| > \varepsilon\} \cap A\big) \in \mathbb{R}_{\geqslant 0}$ for every $A \in \mathscr{A}$ with $\mu(A) \in \mathbb{R}_{\geqslant 0}$ and for every $\varepsilon \in \mathbb{R}_{>0}$.

in μ-measure (or, briefly, in measure), symbolically $f_n \xrightarrow{\mu} f$, if $\lim_{n\to\infty} \mu\big(\{|f_n - f| > \varepsilon\} \cap A\big) = 0$ for every $A \in \mathscr{A}$ with $\mu(A) \in \mathbb{R}_{\geqslant 0}$ and for every $\varepsilon \in \mathbb{R}_{>0}$.

in $\mathcal{L}^{s}(\mu)$ (or in s-th μ-mean) for $s \in \overline{\mathbb{R}}_{>0}$, symbolically $f_n \xrightarrow{\mathcal{L}^{s}(\mu)} f$, if $(f_n)_{n\in\mathbb{N}}$ and f in $\bar{\mathcal{L}}^{s}(\mu)$ such that $\lim_{n\to\infty}\|f_n - f\|_{\mathcal{L}^s} = 0$.

If μ is a probability measure, then convergence in μ-measure is also called convergence *in probability*. Sometimes we write briefly $f_n \xrightarrow{\text{a.e.}} f$, $f_n \xrightarrow{\text{a.c.}} f$ or $f_n \xrightarrow{\mathcal{L}^s} f$ if the underlying measure emerges from the context. □

§02.30 **Remark**. Convergence in $\mathcal{L}^{s}(\mu)$ and convergence μ-almost everywhere evidently determine the limit up to equality μ-almost everywhere. This also applies to convergence in μ-measure, if $\mu \in \mathfrak{M}_{\sigma}(\mathscr{A})$ is σ-finite. Indeed, if $f_n \xrightarrow{\mu} f$ and $f_n \xrightarrow{\mu} g$ then for every $\varepsilon \in \mathbb{R}_{>0}$ and $A \in \mathscr{A}$ with $\mu(A) \in \mathbb{R}_{\geqslant 0}$ (since $|f-g| \leqslant |f - f_n| + |g - f_n|$)

$$\mu(\{|f-g| > \varepsilon\} \cap A) \leqslant \mu(\{|f - f_n| > \varepsilon/2\} \cap A) + \mu(\{|g - f_n| > \varepsilon/2\} \cap A) \xrightarrow{n\to\infty} 0.$$

and hence $\mu(\{|f-g| > \varepsilon\} \cap A) = 0$. Therefore, we have $\mu(\{f \neq g\} \cap A) = 0$ making use of $\{f \neq g\} \cap A = \bigcup_{k\in\mathbb{N}}\{|f-g| > 1/k\} \cap A$. Selecting $A_n \uparrow \Omega$ with $\mu(A_n) \in \mathbb{R}_{\geqslant 0}$ (since $\mu \in \mathfrak{M}_{\sigma}(\mathscr{A})$) implies $f = g$ μ-a.e.. If $\mu \in \mathfrak{M}_{\text{fi}}(\mathscr{A})$ is finite, then $\lim_{n\to\infty}\mu(\{|f_n - f| > \varepsilon\}) = 0$ for every $\varepsilon \in \mathbb{R}_{>0}$ and $f_n \xrightarrow{\mu} f$ are equivalent. The last statement does not apply, if $\mu \in \mathfrak{M}_{\sigma}(\mathscr{A})$ is σ-finite. For instance, on $(\mathbb{N}, 2^{\mathbb{N}}, \zeta_{\mathbb{N}})$ (see Example §01.18 (c) for the counting measure $\zeta_{\mathbb{N}}$) for $A_n := \mathbb{N}_{\geqslant n}$, $n \in \mathbb{N}$, we have $\{\mathbb{1}_{A_n} > \varepsilon\} = A_n$ for every $\varepsilon \in (0,1)$ and $\{\mathbb{1}_{A_n} > \varepsilon\} = \varnothing$ for every $\varepsilon \in \mathbb{R}_{\geqslant 1}$. Since $A_n \downarrow \varnothing$, and hence $\zeta_{\mathbb{N}}(A_n \cap A) \downarrow 0$ for each $A \in \mathscr{A}$ with $\zeta_{\mathbb{N}}(A) \in \mathbb{R}_{\geqslant 0}$ (upper semicontinuous), we evidently have $\mathbb{1}_{A_n} \xrightarrow{\zeta_{\mathbb{N}}} 0$. On the other hand side, for each $\varepsilon \in (0,1)$ we have $\zeta_{\mathbb{N}}(\{\mathbb{1}_{A_n} > \varepsilon\}) = \zeta_{\mathbb{N}}(A_n) = \infty$ for all $n \in \mathbb{N}$. □

§02.31 **Lemma**. *Let $(\Omega, \mathscr{A}, \mu)$ be an arbitrary measure space.*

(i) *(Monotone convergence) Let $f \in \overline{\mathcal{M}}(\mathscr{A})$ and let $f_n \in \bar{\mathcal{L}}^{1}(\mu)$, $n \in \mathbb{N}$. Assume $f_n \uparrow f$ μ-a.e. Then $\mu(f_n) \uparrow \mu(f)$ where both sides can equal $+\infty$.*

(ii) *(Dominated convergence) Let $(f_n)_{n\in\mathbb{N}}$ in $\overline{\mathcal{M}}(\mathscr{A})$ be μ-a.e. convergent. Assume $\sup_{n\in\mathbb{N}}|f_n| \leqslant g$ μ-a.e. with $g \in \bar{\mathcal{L}}^{1}(\mu)$. Then there exists $f \in \mathcal{M}(\mathscr{A})$ with $f_n \xrightarrow{\mu\text{-a.e.}} f$, $(f_n)_{n\in\mathbb{N}}$ and f belong to $\bar{\mathcal{L}}^{1}(\mu)$ and $\lim_{n\to\infty}\mu((|f - f_n|)) = 0$ as well as $\lim_{n\to\infty}\mu(f_n) = \mu(f)$. If $g \in \bar{\mathcal{L}}^{s}(\mu)$ for $s \in \mathbb{R}_{\geqslant 1}$, then $(f_n)_{n\in\mathbb{N}}$ and f belong to $\bar{\mathcal{L}}^{s}(\mu)$ and $\lim_{n\to\infty}\|f_n - f\|_{\mathcal{L}^s} = 0$.*

(iii) *(Scheffé's theorem) Let $f, f_n \in \mathcal{M}_{\geqslant 0}(\mathscr{A})$, $n \in \mathbb{N}$, be μ-integrable. Assume $f_n \xrightarrow{\mu\text{-a.e.}} f$ and $\mu(f_n) \xrightarrow{n\to\infty} \mu(f)$, then $f_n \xrightarrow{\mathcal{L}^{1}(\mu)} f$.*

(iv) *(Theorem of Riesz) Let $f, f_n \in \mathcal{L}^{s}(\mu)$, $n \in \mathbb{N}$, $s \in \mathbb{R}_{\geqslant 1}$. Assume $f_n \xrightarrow{\mu\text{-a.e.}} f$. $\mu(|f_n|^s) \xrightarrow{n\to\infty} \mu(|f|^s)$ if and only if $f_n \xrightarrow{\mathcal{L}^{s}(\mu)} f$.*

(v) *Let $f, f_n \in \mathcal{M}(\mathscr{A})$, $n \in \mathbb{N}$. Then the following implications hold:*

$$f_n \xrightarrow{\mu\text{-a.e.}} f \Longrightarrow f_n \xrightarrow{\mu} f \Longleftarrow f_n \xrightarrow{\mathcal{L}^{s}(\mu)} f.$$

If $\mu \in \mathfrak{M}_{\sigma}(\mathscr{A})$ is σ-finite, then we also have $f_n \xrightarrow{\mu\text{-a.c.}} f \Longrightarrow f_n \xrightarrow{\mu\text{-a.e.}} f$. Moreover, $f_n \xrightarrow{\mu} f$ if and only if for any subsequent of $(f_n)_{n\in\mathbb{N}}$ there exists a sub-subsequence that converges to f μ-almost everywhere.

§02.32 **Proof** of Lemma §02.31. is given in the lecture. □

§02.33 **Reminder**. A sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{L}^{s}(\mu)$ is called ($\mathcal{L}^{s}(\mu)$-)Cauchy sequence, if for every $\varepsilon \in \mathbb{R}_{>0}$ there exists $n_\circ \in \mathbb{N}$ such that $\|f_n - f_m\|_{\mathcal{L}^s(\mu)} \leqslant \varepsilon$ for all $m, n \in \mathbb{N}_{\geqslant n_\circ}$, symbolically $\lim_{n,m\to\infty}\|f_n - f_m\|_{\mathcal{L}^s(\mu)} = 0$. For $s \in \overline{\mathbb{R}}_{\geqslant 1}$ keep in mind that every $\mathcal{L}^{s}(\mu)$ convergent sequence by applying Minkowski's inequality (see Lemma §02.25 (iii)) is also a $\mathcal{L}^{s}(\mu)$-Cauchy sequence. □

§02.34 **Lemma**. *Let* $\mathrm{s} \in \overline{\mathbb{R}}_{\geqslant 1}$ *and let* $(f_n)_{n\in\mathbb{N}}$ *be a* $\mathcal{L}^{\mathrm{s}}(\mu)$*-Cauchy sequence. Then there exists* $f \in \mathcal{L}^{\mathrm{s}}(\mu)$ *such that* $f_n \xrightarrow{\mathcal{L}^{\mathrm{s}}(\mu)} f$ *and there exists a subsequence of* $(f_n)_{n\in\mathbb{N}}$ *that converges* μ-a.e. *to* f.

§02.35 **Proof** of Lemma §02.34. is given in the lecture. □

§02.36 **Corollary**. *Let* $\mathrm{s} \in \overline{\mathbb{R}}_{\geqslant 1}$ *and let* $(f_n)_{n\in\mathbb{N}}$ *be a* $\mathcal{L}^{\mathrm{s}}(\mu)$*-Cauchy sequence that converges* μ-a.e. *to* $f \in \overline{\mathcal{M}}(\mathscr{A})$. *Then* f *belongs to* $\mathcal{L}^{\mathrm{s}}(\mu)$ *and* $f_n \xrightarrow{\mathcal{L}^{\mathrm{s}}(\mu)} f$.

§02.37 **Proof** of Corollary §02.36. is given in the lecture. □

§02.38 **Preliminaries**. Let $(\Omega, \mathscr{A}, \mu)$ be an arbitrary measure space and let $f \in \overline{\mathcal{M}}(\mathscr{A})$. f is μ-integrable if and only if for every $\varepsilon \in \mathbb{R}_{>0}$ there exists $g \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mu) := \bar{\mathcal{L}}^1(\mu) \cap \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ such that $\mu(|f|\mathbb{1}_{\{|f|\geqslant g\}}) \leqslant \varepsilon$ or in equal $\inf\{\mu(|f|\mathbb{1}_{\{|f|\geqslant g\}}) : g \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mu)\} = 0$. Assume $\mu(|f|) \in \mathbb{R}_{\geqslant 0}$. Setting $g := 2|f| \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mu)$ we evidently have $\{|f| \geqslant g\} = \{f = 0\} \cup \{|f| = \infty\}$ and hence applying Corollary §02.23 (ii) also $\mu(|f|\mathbb{1}_{\{|f|\geqslant g\}}) = 0$. We obtain the converse by exploiting $\mu(|f|) = \mu(|f|\mathbb{1}_{\{|f|\geqslant g\}}) + \mu(|f|\mathbb{1}_{\{|f|<g\}}) \leqslant \varepsilon + \mu(g) \in \mathbb{R}_{>0}$, which in turn implies $\mu(|f|) \in \mathbb{R}_{\geqslant 0}$. □

§02.39 **Definition**. A class of functions $\mathcal{F} \subseteq \bar{\mathcal{L}}^1(\mu)$ is called *uniformly μ-integrable* if

$$\inf\Big\{\sup_{f\in\mathcal{F}} \mu(|f|\mathbb{1}_{\{|f|\geqslant g\}}) : g \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mu)\Big\} = 0.$$

If $\mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ is finite, then uniform μ-integrability is equivalent to the condition:

$$\inf\Big\{\sup_{f\in\mathcal{F}} \mu(|f|\mathbb{1}_{\{|f|\geqslant a\}}) : a \in \mathbb{R}_{\geqslant 0}\Big\} = 0.$$ □

§02.40 **Remark**.

(a) Let $\mathcal{F}$ be uniformly μ-integrable and let $\varepsilon \in \mathbb{R}_{>0}$. A function $g \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mu)$ is called *ε-majorant* if $\sup_{f\in\mathcal{F}} \mu(|f|\mathbb{1}_{\{|f|\geqslant g\}}) \leqslant \varepsilon$. Evidently, there exists a ε-majorant g for $\mathcal{F}$ and every $h \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mu)$ with $h \geqslant g$ is also a ε-majorant for $\mathcal{F}$.

(b) A family $(f_i)_{i\in\mathcal{I}}$ in $\overline{\mathcal{M}}(\mathscr{A})$ is called *uniformly μ-integrable* if the class $\{f_i : i \in \mathcal{I}\}$ is.

(c) Let $\mathcal{F}_i$, $i \in [\![n]\!]$, be finitely many uniformly μ-integrable classes in $\overline{\mathcal{M}}(\mathscr{A})$. Then their union $\mathcal{F} := \cup_{i\in[\![n]\!]}\mathcal{F}_i$ is also uniformly μ-integrable. Indeed, for every $\varepsilon \in \mathbb{R}_{>0}$ and ε-majorant g_i for $\mathcal{F}_i$, $i \in [\![n]\!]$, the function $g_1 \vee \cdots \vee g_n$ is a ε-majorant for $\mathcal{F}$.

(d) Let $\mathcal{F} \subseteq \overline{\mathcal{M}}(\mathscr{A})$ and let $g \in \bar{\mathcal{L}}^{\mathrm{s}}_{\geqslant 0}(\mu) := \bar{\mathcal{L}}^{\mathrm{s}}(\mu) \cap \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, $\mathrm{s} \in \mathbb{R}_{\geqslant 1}$, satisfy $|f| \leqslant g$ μ-a.e. for all $f \in \mathcal{F}$. Then $\mathcal{F}^{\mathrm{s}} := \{|f|^{\mathrm{s}} : f \in \mathcal{F}\}$ is uniformly μ-integrable. For $\varepsilon \in \mathbb{R}_{>0}$ every ε-majorant h for $\{g^{\mathrm{s}}\}$ is a ε-majorant for $\mathcal{F}^{\mathrm{s}}$, since $\mu(|f|^{\mathrm{s}}\mathbb{1}_{\{|f|^{\mathrm{s}}>h\}}) \leqslant \mu(g^{\mathrm{s}}\mathbb{1}_{\{g^{\mathrm{s}}>h\}}) \leqslant \varepsilon$ for all $f \in \mathcal{F}$. □

End LE05 2026/04/29

§02.41 **Lemma**. *Let* $(\Omega, \mathscr{A}, \mu)$ *be an arbitrary measure space.*

(i) *If* $\mathcal{F} \subseteq \bar{\mathcal{L}}^1(\mu)$ *is a finite set then* $\mathcal{F}$ *is uniformly* μ*-integrable.*

(ii) *If* $\mathcal{F}, \mathcal{G} \subseteq \mathcal{L}^1(\mu)$ *are uniformly* μ*-integrable, then* $\{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$, $\{f - g : f \in \mathcal{F}, g \in \mathcal{G}\}$ *and* $\{|f| : f \in \mathcal{F}\}$ *are uniformly* μ*-integrable.*

(iii) *If* $\mathcal{F} \subseteq \bar{\mathcal{L}}^1(\mu)$ *is uniformly* μ*-integrable, and if, for any* $g \in \mathcal{G} \subseteq \overline{\mathcal{M}}(\mathscr{A})$, *there exists an* $f \in \mathcal{F}$ *with* $|g| \leqslant |f|$, *then* $\mathcal{G} \subseteq \bar{\mathcal{L}}^1(\mu)$ *is also uniformly* μ*-integrable.*

(iv) *Let* $\mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ *be finite, let* $\mathrm{s} \in \mathbb{R}_{>1}$ *and let* $\mathcal{F}$ *be bounded in* $\mathcal{L}^{\mathrm{s}}(\mu)$, *that is,* $\sup\{\|f\|_{\mathcal{L}^{\mathrm{s}}} : f \in \mathcal{F}\} \in \mathbb{R}_{\geqslant 0}$. *Then* $\mathcal{F}$ *is uniformly* μ*-integrable.*

§02.42 **Proof** of Lemma §02.41. We refer to the lecture EWS / Exercise. □

§02.43 **Theorem**. *Let $(\Omega, \mathscr{A}, \mu)$ be an arbitrary measure space. $\mathcal{F} \subseteq \overline{\mathcal{M}}(\mathscr{A})$ is uniformly μ-integrable if and only if the following two conditions hold:*

(uI1) *$\mathcal{F}$ is bounded in $\bar{\mathcal{L}}^1(\mu)$, i.e.* $\sup\{\mu(|f|) \colon f \in \mathcal{F}\} \in \mathbb{R}_{\geqslant 0}$.

(uI2) *For any $\varepsilon \in \mathbb{R}_{>0}$ there are $h \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mu)$ and $\delta \in \mathbb{R}_{>0}$ such that for all $A \in \mathscr{A}$ holds the implication: $\mu(h\mathbb{1}_A) \leqslant \delta \Rightarrow \sup_{f \in \mathcal{F}} \mu(|f|\mathbb{1}_A) \leqslant \varepsilon$.*

§02.44 **Proof** of Theorem §02.43. is given in the lecture. □

§02.45 **Theorem**. *Let $(\Omega, \mathscr{A}, \mu)$ be a measure space, let $s \in \mathbb{R}_{\geqslant 1}$ and let $(f_n)_{n\in\mathbb{N}}$ belong to $\bar{\mathcal{L}}^s(\mu)$. $(f_n)_{n\in\mathbb{N}}$ converges in $\mathcal{L}^s(\mu)$ if and only if the following two conditions hold:*

(C1) *$(|f_n|^s)_{n\in\mathbb{N}}$ is uniformly μ-integrable.*

(C2) *$(f_n)_{n\in\mathbb{N}}$ converges in μ-measure.*

§02.46 **Proof** of Theorem §02.45. „$\Rightarrow$“ in the lecture, for the converse we refer to Bauer (1992, Theorem 21.4, p.142) □

§02.47 **Remark**. The Theorem §02.45 guarantees the existence of a $\mathcal{L}^s(\mu)$-integrable function under the possible limits in μ-measure of the sequence $(f_n)_{n\in\mathbb{N}}$. □

§02.48 **Corollary**. *Let $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ be σ-finite, let $s \in \mathbb{R}_{\geqslant 1}$ and let $(f_n)_{n\in\mathbb{N}}$ belong to $\mathcal{L}^s(\mu)$. Assume $f_n \xrightarrow{\mu} f \in \mathcal{M}(\mathscr{A})$ and $(|f_n|^s)_{n\in\mathbb{N}}$ is uniformly μ-integrable. Then $f \in \mathcal{L}^s(\mu)$ and $f_n \xrightarrow{\mathcal{L}^s(\mu)} f$.*

§02.49 **Proof** of Corollary §02.48. is given in the lecture. □

§02.50 **Summary**. Let $(\Omega, \mathscr{A}, \mu)$ be an arbitrary measure space, let $s \in \overline{\mathbb{R}}_{\geqslant 1}$, and let $(f_n)_{n\in\mathbb{N}}$ belong to $\mathcal{L}^s(\mu)$. The following statements are equivalent:

(C1) There is $f \in \mathcal{L}^s(\mu)$ such that $f_n \xrightarrow{\mathcal{L}^s(\mu)} f$.

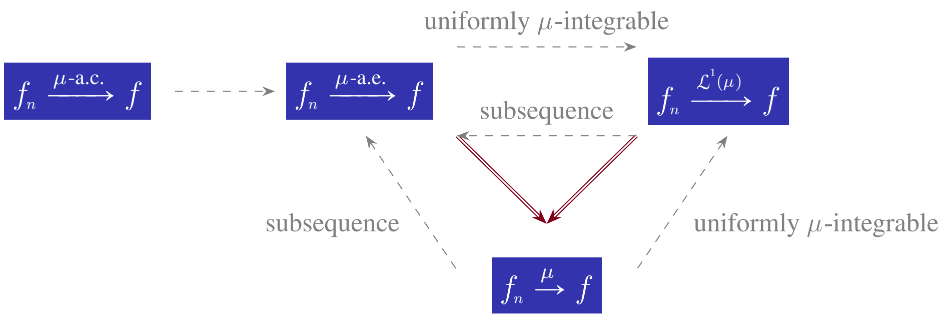
(C2) $(f_n)_{n\in\mathbb{N}}$ is a $\mathcal{L}^s(\mu)$-Cauchy sequence, i.e. $\lim_{n,m\to\infty} \|f_n - f_m\|_{\mathcal{L}^s} = 0$.

Assume in addition $s \in \mathbb{R}_{\geqslant 1}$ and $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ is σ-finite. Then (C1) and (C2) are equivalent to

(C3) $(|f_n|^s)_{n\in\mathbb{N}}$ is uniformly μ-integrable, and there is $f \in \mathcal{M}(\mathscr{A})$ such that $f_n \xrightarrow{\mu} f$.

The limes in (C1) and in (C3) coincide.

Fig. 02 [§02] Implications of convergence criteria for a σ-finite measure $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$.



The Figure 02 [§02] was created based on Klenke (2020, Fig. 6.1, p.159). □

§02|03 $\mathbb{L}^s$-Spaces

§02.51 **Reminder**. For $s \in \overline{\mathbb{R}}_{>0}$ and $f, g \in \mathcal{M}(\mathscr{A})$ we have shown that $\|f - g\|_{\mathcal{L}^s(\mu)} = 0$ if and only if $f = g$ μ-a.e.. In this case we now consider f and g as equivalent. More precisely, for each $f \in \mathcal{M}(\mathscr{A})$ we introduce the μ-equivalence class $\{f\}_\mu := \{g \in \mathcal{M}(\mathscr{A}) \colon g = f\ \mu\text{-a.e.}\}$ and hence

$\{0\}_\mu = \{g \in \mathcal{M}(\mathscr{A})\colon g = 0\ \mu\text{-a.e.}\}$. For any $s \in \overline{\mathbb{R}}_{\geqslant 1}$, $\{0\}_\mu$ is a subvector space of $\mathcal{L}^s(\mu)$. Thus formally we can build the factor space

$$\boxed{\mathbb{L}^s} := \mathbb{L}^s(\mu) := \mathbb{L}^s(\Omega, \mathscr{A}, \mu) := \{\{f\}_\mu = f + \{0\}_\mu\colon f \in \mathcal{L}^s(\mu)\}.$$

For $\{f\}_\mu \in \mathbb{L}^s(\mu)$, define $\|\{f\}_\mu\|_{\mathbb{L}^s(\mu)} := \|f\|_{\mathcal{L}^s}$ for any $f \in \{f\}_\mu$. Also let $\mu(\{f\}_\mu) := \mu(f)$ if this expression is defined for f. Note that $\|\{f\}_\mu\|_{\mathbb{L}^s(\mu)}$ and $\mu(\{f\}_\mu)$ do not depend on the choice of the representative $f \in \{f\}_\mu$. Similarly, for $\{f\}_\mu, \{g\}_\mu \in \mathbb{L}^2(\mu)$ define

$$\langle \{f\}_\mu, \{g\}_\mu \rangle_{\mathbb{L}^2(\mu)} := \langle f, g \rangle_{\mathcal{L}^2(\mu)} = \mu(fg)$$

with $f \in \{f\}_\mu$ and $g \in \{g\}_\mu$. □

§02.52 **Lemma** *(Fischer-Riesz). Let $(\Omega, \mathscr{A}, \mu)$ be an arbitrary measure space. $(\mathbb{L}^s(\mu), \|\cdot\|_{\mathbb{L}^s(\mu)})$ is a Banach space for any $s \in \overline{\mathbb{R}}_{\geqslant 1}$. $(\mathbb{L}^2(\mu), \langle\cdot,\cdot\rangle_{\mathbb{L}^2(\mu)})$ is a real Hilbert space.*

§02.53 **Proof** of Lemma §02.52. we refer, for example, to Klenke (2020, Theorem 7.18, p.171) □

§02.54 **Remark**. Let $(V, \langle\cdot,\cdot\rangle)$ be a Hilbert space. Then the *Riesz-Fréchet representation theorem* states, that a map $F : V \to \mathbb{R}$ is continuous and linear if and only if there is an $f \in V$ with $F(x) = \langle f, x\rangle$ for all $x \in V$. The uniquely determined element $f \in V$ is called *representative* of F. In the next section we will need the representation theorem for the space $\mathcal{L}^2$, which unlike $\mathbb{L}^2$ is not a Hilbert space. The representation theorem still holds if V is a linear vector space and $\langle\cdot,\cdot\rangle$ is a complete positive semidefinite symmetric bilinear form (complete semi-inner product) (c.f. Klenke (2020) section 7.3). □

§02.55 **Lemma**. *The map $F : \mathcal{L}^2(\mu) \to \mathbb{R}$ is continuous and linear if and only if there is an $f \in \mathcal{L}^2(\mu)$ with $F(g) = \mu(gf)$ for all $g \in \mathcal{L}^2(\mu)$.*

§02.56 **Proof** of Lemma §02.55. we refer to Klenke (2020, Corollary 7.28, p.174) □

§03 Measures with density - Theorem of Radon-Nikodym

§03.01 **Definition**. Let $\nu, \mu \in \mathfrak{M}(\mathscr{A})$ be arbitrary measures on $(\Omega, \mathscr{A})$.

$\boxed{\nu \ll \mu}$: ν is called *absolutely continuous* with respect to μ, *μ-continuous*, or *dominated* by μ, if any μ-null set is also a ν-null set, that is, $\nu(A) = 0$ for all $A \in \mathscr{A}$ with $\mu(A) = 0$. The measures Maße μ and ν are called *equivalent* (symbolically $\mu \ll\gg \nu$), if $\nu \ll \mu$ and $\mu \ll \nu$.

$\boxed{\mu \perp \nu}$: μ is called *singular* to ν or *ν-singular*, if there exists a μ-null set $N \in \mathscr{A}$ such that $\nu(\Omega\backslash N) = 0$, or in equal $\nu = \mathbb{1}_N \nu$, that is, $\nu(A) = \nu(A \cap N)$ for all $A \in \mathscr{A}$.

: ν is called *totally continuous* with respect to μ if, for any $\varepsilon \in \mathbb{R}_{>0}$ there exists a $\delta \in \mathbb{R}_{>0}$ such that $\nu(A) \leqslant \varepsilon$ for all $A \in \mathscr{A}$ with $\mu(A) \leqslant \delta$. □

§03.02 **Remark**. Evidently, $\mu \perp \nu$ if and only if there are $\Omega_\mu, \Omega_\nu \in \mathscr{A}$ with $\Omega = \Omega_\mu \biguplus \Omega_\nu$ and $\mu(\Omega_\nu) = 0 = \nu(\Omega_\mu)$, and hence if and only if $\nu \perp \mu$. Consequently measures $\mu, \nu \in \mathfrak{M}(\mathscr{A})$ with $\mu \perp \nu$ are also called mutually singular. The condition $\nu = \mathbb{1}_N \nu$ means the support of the meassure ν is contained in $N \in \mathscr{A}$. Note that $\nu \ll \mu$ and $\nu \perp \mu$ imply together $\nu(\Omega_\nu) = 0$, and hence $\nu(\Omega) = 0$. □

§03.03 **Lemma**. *Let $\nu, \mu \in \mathfrak{M}(\mathscr{A})$ be measures on $(\Omega, \mathscr{A})$. If ν is totally continuous with respect to μ, then $\nu \ll \mu$. If $\nu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ is finite, then the converse also holds.*

§03.04 **Proof** of Lemma §03.03. is given in the lecture. □

Reminder. For measures $\mu, \nu \in \mathfrak{M}(\mathscr{A})$ we write $\nu \leqslant \mu$ if $\nu(A) \leqslant \mu(A)$ for all $A \in \mathscr{A}$. □

§03.05 **Lemma**. *Let $\nu, \mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ be finite measures with $\nu \leqslant \mu$, then there exists $h \in \mathcal{M}_{[0,1]}(\mathscr{A})$ such that $\nu = h\mu$.*

§03.06 **Proof** of Lemma §03.05. is given in the lecture. □

§03.07 **Theorem of Radon-Nikodym**. *Let $\mu \in \mathfrak{M}_{\sigma}(\mathscr{A})$ be a σ-finite measure and let $\nu \in \mathfrak{M}(\mathscr{A})$ be a μ-continuous measure, i.e. $\nu \ll \mu$. Then ν has a density $\mathbb{f} = \mathrm{d}\nu/\mathrm{d}\mu \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ with respect to μ, that is, $\nu = \mathbb{f}\,\mu$.*

§03.08 **Proof** of Theorem §03.07. is given in the lecture. □

§03.09 **Remark**. Let $\mu, \nu \in \mathfrak{M}_{\sigma}(\mathscr{A})$ be σ-finite measures with $\nu \ll \mu$ and let $\mathbb{f} = \mathrm{d}\nu/\mathrm{d}\mu \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ be a μ-density of ν. Then Theorem §03.07 implies directly the usual *chain rules*:

(a) If $g \in \overline{\mathcal{M}}(\mathscr{A})$ is ν-quasiintegrable, then $\nu(g\mathbb{1}_A) = \mu(g\mathbb{f}\mathbb{1}_A)$ for all $A \in \mathscr{A}$.

(b) If $\rho \in \mathfrak{M}_{\sigma}(\mathscr{A})$ is a σ-finite measure with $\rho \ll \nu \ll \mu$ then $\frac{\mathrm{d}\rho}{\mathrm{d}\mu} = \frac{\mathrm{d}\rho}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ μ-a.e..

(c) If $h \in \mathcal{M}_{[0,1]}(\mathscr{A})$ with $h = \frac{\mathrm{d}\nu}{\mathrm{d}(\nu+\mu)}$ μ-a.e. then $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = \frac{h}{1-h}$ μ-a.e.. □

§03.10 **Example**.

(a) *Continuous probability measures* on $(\mathbb{R}^k, \mathscr{B}^k)$ as studied in the lecture EWS are probability measures dominated by the Lebesgue measure λ^k with corresponding (Radon-Nikodym-) density.

(b) *Discret probability measures* on a countable set Ω introduced in the lecture EWS are probability measures dominated by the counting measure ζ_Ω and the mass function corresponds to the (Radon-Nikodym-) density. Similarly, if $\Omega \subseteq \mathbb{R}$ then the discrete measure $\mu \in \mathfrak{M}_{\sigma}(\mathscr{B})$ with mass function $\mathbb{p}$ as in Example §01.18 (c) is absolutely continuous with respect to the counting measure $\zeta_\Omega \in \mathfrak{M}_{\sigma}(\mathscr{B})$ with (Radon-Nikodym-) density $\mathbb{p}$. □

§03.11 **Lebesgue's decomposition theorem**. *Let $\mu, \nu \in \mathfrak{M}_{\sigma}(\mathscr{A})$ be σ-finite measures on $(\Omega, \mathscr{A})$. Then there exists a unique decomposition $\nu = \nu_a + \nu_s$ of ν into two measures $\nu_a, \nu_s \in \mathfrak{M}(\mathscr{A})$ such that $\nu_a \ll \mu$ and $\nu_s \perp \mu$ is the μ-continuous, respectively the μ-singular part of ν. Moreover, $\nu_a, \nu_s \in \mathfrak{M}_{\sigma}(\mathscr{A})$ are σ-finite, and $\nu_a, \nu_s \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ are finite if and only if $\nu \in \mathfrak{M}_{e}(\mathscr{A})$ is finite. ν_a has a μ-density $\mathrm{d}\nu_a/\mathrm{d}\mu \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ with $\mathrm{d}\nu_a/\mathrm{d}\mu \in \mathbb{R}_{\geqslant 0}$ μ-a.e.*

§03.12 **Proof** of Theorem §03.11. is given in the lecture. □

End LE06
2026/05/06

§03.13 **Remark**. If $\mathbb{f} = \mathrm{d}\nu_a/\mathrm{d}\mu \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ is a μ-density of ν_a as in Theorem §03.11 then the positive real function $\tilde{\mathbb{f}} := \mathbb{f}\mathbb{1}_{\{\mathbb{f}\in\mathbb{R}_{\geqslant 0}\}} \in \mathcal{M}_{\geqslant 0}(\mathscr{A})$ is also a μ-density of ν_a, since $\mathbb{f} = \tilde{\mathbb{f}}$ μ-a.e. In other words $\tilde{\mathbb{f}} \in \mathcal{M}_{\geqslant 0}(\mathscr{A})$ is also a version of the Radon-Nikodym density of ν_a with respect to μ. Consequently, without loss of generality we chose here and subsequently a positive real version of the Radon-Nikodym density. Furthermore, given $\mathbb{f} = \mathrm{d}\nu_a/\mathrm{d}\mu \in \mathcal{M}_{\geqslant 0}(\mathscr{A})$ let us define a numerical function $\mathrm{L} := \mathbb{f}\mathbb{1}_{N^c} + \infty\mathbb{1}_N \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ with $\mu(N) = 0 = \nu_s(N^c)$ where $\{\mathrm{L} = \infty\} = N$ and the Lebesgue decomposition writes $\nu = \mathrm{L}\mu + \mathbb{1}_{\{\mathrm{L}=\infty\}}\nu$, i.e. for all $A \in \mathscr{A}$ we have $\nu(A) = \mu(\mathbb{1}_A\mathrm{L}) + \nu(A \cap \{\mathrm{L} = \infty\})$. □

§03.14 **Definition**. Let $\nu, \mu \in \mathfrak{M}_{\sigma}(\mathscr{A})$ be two σ-finite measures on $(\Omega, \mathscr{A})$, where $\nu \ll \mu$ does not necessarily hold. Any positive numerical function $\mathrm{L} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ satisfying

$$\mu(\mathrm{L} = \infty) = 0 \text{ and } \nu = \mathrm{L}\mu + \mathbb{1}_{\{\mathrm{L}=\infty\}}\nu \tag{03.01}$$

is called *density ratio (DR)* of ν with respect to μ, or *μ-density ratio* of ν. For probability measures a density ratio is also called *likelihood ratio (LR)*. □

§03.15 **Lemma**. *Let $\nu, \mu \in \mathfrak{M}_\sigma(\mathscr{A})$ be two σ-finite measures. Then the μ-density ratio $\mathrm{L} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ of ν is unique up to $(\nu + \mu)$-a.e. equivalence.*

§03.16 **Proof** of Lemma §03.15. is given in the lecture. □

Alternative formulation of the theorem of Radon-Nikodym

§03.17 **Definition**. Let $(\Omega, \mathscr{A}, \mu)$ be a measure space and let $\mathcal{F} \subseteq \overline{\mathcal{M}}(\mathscr{A})$ be a class of numerical functions. Any function $g \in \overline{\mathcal{M}}(\mathscr{A})$ satisfying the following two conditions

(eS1) $f \leqslant g$ μ-a.e. for all $f \in \mathcal{F}$;

(eS2) if $h \in \overline{\mathcal{M}}(\mathscr{A})$ satisfies $f \leqslant h$ μ-a.e. for all $f \in \mathcal{F}$, then $g \leqslant h$ μ-a.e.

is called a *μ-essential supremum* over $\mathcal{F}$, symbolically $g = \mu\text{-ess sup}_{f\in\mathcal{F}} f$. □

§03.18 **Remark**. The μ-essential supremum can be seen as an extension of the usual concept of the supremum. If $\mathcal{F}$ is countable and $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ is σ-finite, then $g := \sup_{f\in\mathcal{F}} f$ satisfies the conditions §03.17 (eS1) and (eS2), and hence $\sup_{f\in\mathcal{F}} f = \mu\text{-ess sup}_{f\in\mathcal{F}} f$ μ-a.e. In contrast, if for example $\mathcal{F} = \{\mathbb{1}_{\{x\}}, x \in B\}$ with uncountable $B \in \mathscr{B}$ such that $\lambda(B) \in \mathbb{R}_{>0}$, then the λ-essential supremum and the usual supremum differ. Precisely, $\sup_{f\in\mathcal{F}} f = \mathbb{1}_B \neq 0 = \lambda\text{-ess sup}_{f\in\mathcal{F}} f$. □

§03.19 **Lemma**. *Let $\mu \in \mathfrak{M}_\sigma(\mathscr{A})$ and $\mathcal{F} \subseteq \overline{\mathcal{M}}(\mathscr{A})$. Then:*

(i) *$g := \mu\text{-ess sup}_{f\in\mathcal{F}} f$ exists and it is μ-a.e. uniquely determined, that is, if $g \in \overline{\mathcal{M}}(\mathscr{A})$ is a solution of Definition §03.17 (eS1) and (eS2) then also $\widetilde{g} \in \overline{\mathcal{M}}(\mathscr{A})$ with $\mu(\{g \neq \widetilde{g}\}) = 0$.*

(ii) *There exists a sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{F}$ with $g = \sup_{n\in\mathbb{N}} f_n$ μ-a.e.*

(iii) *If $\mathcal{F}$ is increasing filtered (for all $h, k \in \mathcal{F}$ exists $f \in \mathcal{F}$ with $f \geqslant h \vee k$), then there exists an isotone sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathcal{F}$ with $f_n \uparrow g$ μ-a.e..*

§03.20 **Proof** of Lemma §03.19. We refer to Witting (1985, Satz 1.102, S.105). □

§03.21 **Lemma**. *Let $\mu, \nu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ be finite and mutually not singular measures on $(\Omega, \mathscr{A})$. Then there is $\Omega_\circ \in \mathscr{A}$ with $\mu(\Omega_\circ) \in \mathbb{R}_{>0}$ and $\varepsilon \in \mathbb{R}_{>0}$ with $\varepsilon \mathbb{1}_{\Omega_\circ}\mu \leqslant \mathbb{1}_{\Omega_\circ}\nu$.*

§03.22 **Proof** of Lemma §03.21. The claim is shown in Klenke (2020, Lemma 7.46, p.184) with help of the Hahn-decomposition for signed measures. An alternative proof of the claim is given in the proof of Bauer (1992, Theorem 17.10, p.117) exploiting Bauer (1992, Lemma 17.9, p.114). □

§03.23 **Lemma**. *Let $\nu, \mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{A})$ be finite with $\nu \leqslant \mu$. Set $\mathcal{F} := \{f \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) : f\mu \leqslant \nu\}$. There exists $g \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ such that $g = \mu\text{-ess sup}_{f\in\mathcal{F}} f$ μ-a.e. and $\nu = g\mu$, that is, g is a version of the μ-density of ν.*

§03.24 **Proof** of Lemma §03.23. is given in the lecture. □

§04 Measures on product spaces

§04|01 Finite product measures

§04.01 **Reminder**. Let $\mathcal{I}$ be an arbitrary nonempty index set and let $(\mathcal{S}_i, \mathscr{S}_i)$, $i \in \mathcal{I}$, be measurable spaces. The set $\mathcal{S}_\mathcal{I} := \bigtimes_{i\in\mathcal{I}} \mathcal{S}_i$ of all maps $(s_i)_{i\in\mathcal{I}} : \mathcal{I} \to \cup_{i\in\mathcal{I}}\mathcal{S}_i$ such that $s_i \in \mathcal{S}_i$ for all $i \in \mathcal{I}$ is called *product space* or *Cartesian product*. We identify the map $i \mapsto s_i$ and the family $(s_i)_{i\in\mathcal{I}}$. If $\mathcal{S}_i = \mathcal{S}$ for all $i \in \mathcal{I}$ then we write $\mathcal{S}^{\mathcal{I}} := \mathcal{S}_{\mathcal{I}}$, and in case $n := |\mathcal{I}| \in \mathbb{N}$ also $\mathcal{S}^n := \mathcal{S}^{\mathcal{I}}$ for short. For every $\mathcal{J} \subseteq \mathcal{I}$ the map

$$\Pi_{\mathcal{J}} : \mathcal{S}_i \ni s_{\mathcal{I}} := (s_i)_{i\in\mathcal{I}} \mapsto (s_j)_{j\in\mathcal{J}} =: s_{\mathcal{J}} \in \mathcal{S}_{\mathcal{J}}$$

is called *canonical projection* and in particular for $j \in \mathcal{I}$ the map

$$\Pi_j := \Pi_{\{j\}} : \mathcal{S}_{\mathcal{I}} \ni s_{\mathcal{I}} := (s_i)_{i\in\mathcal{I}} \mapsto s_j \in \mathcal{S}_j$$

is called *coordinate map* such that $\times_{i\in\mathcal{I}} \mathcal{E}_i = \bigcap_{i\in\mathcal{I}} \Pi_i^{-1}(\mathcal{E}_i)$ for all $\mathcal{E}_i \subseteq \mathcal{S}_i$ and $i \in \mathcal{I}$. □

§04.02 **Definition**. Let $\mathcal{I}$ be an arbitrary nonempty index set.

(a) Let $(\Omega, \mathscr{A})$ be a measurable space and for each $i \in \mathcal{I}$ let $\mathscr{A}_i \subseteq \mathscr{A}$ be a σ-algebra. The σ-algebra

$$\bigwedge_{i\in\mathcal{I}} \mathscr{A}_i := \bigcap_{i\in\mathcal{I}} \mathscr{A}_i \quad \text{and} \quad \bigvee_{i\in\mathcal{I}} \mathscr{A}_i := \sigma\big(\bigcup_{i\in\mathcal{I}} \mathscr{A}_i\big)$$

is respectively the *largest σ-algebra* on Ω, that belongs to all $\mathscr{A}_i$, $i \in \mathcal{I}$, and the *smallest σ-algebra* on Ω, that contains all $\mathscr{A}_i$, $i \in \mathcal{I}$.

(b) For each $i \in \mathcal{I}$ let $(\mathcal{S}_i, \mathscr{S}_i)$ be a measurable space. The *product-σ-algebra*

$$\mathscr{S}_{\mathcal{I}} := \bigotimes_{i\in\mathcal{I}} \mathscr{S}_i$$

is the smallest σ-algebra on the product space $\mathcal{S}_{\mathcal{I}} = \times_{i\in\mathcal{I}} \mathcal{S}_i$ such that for every $i \in \mathcal{I}$ the coordinate map $\Pi_i : \mathcal{S}_{\mathcal{I}} \to \mathcal{S}_i$ is measurable with respect to $\mathscr{S}_{\mathcal{I}}$-$\mathscr{S}_i$, i.e. $\Pi_i \in \mathcal{M}(\mathcal{S}_{\mathcal{I}}, \mathcal{S}_i)$; that is,

$$\mathscr{S}_{\mathcal{I}} = \bigotimes_{i\in\mathcal{I}} \mathscr{S}_i := \bigvee_{i\in\mathcal{I}} \sigma(\Pi_i) = \bigvee_{i\in\mathcal{I}} \Pi_i^{-1}(\mathscr{S}_i).$$

If $(\mathcal{S}_i, \mathscr{S}_i) = (\mathcal{S}, \mathscr{S})$ for all $i \in \mathcal{I}$, then we also write $\mathscr{S}^{\mathcal{I}} := \mathscr{S}_{\mathcal{I}}$, and $\mathscr{S}^n := \mathscr{S}^{\mathcal{I}}$ in case $n := |\mathcal{I}| \in \mathbb{N}$. The family $(\Pi_i)_{i\in\mathcal{I}}$ is called the *canonical process* on $(\mathcal{S}_{\mathcal{I}}, \mathscr{S}_{\mathcal{I}})$. □

Consider now the situation of finitely many measure spaces $(\mathcal{S}_i, \mathscr{S}_i, \mu_i)$, $i \in [\![n]\!]$, where $n \in \mathbb{N}$.

§04.03 **Lemma**. *For every $i \in [\![n]\!]$ let $\mathscr{E}_i$ be a generator of the σ-algebra $\mathscr{S}_i$ on $\mathcal{S}_i$ and let $(\mathcal{E}_{ik})_{k\in\mathbb{N}}$ be a sequence in $\mathscr{E}_i$ such that $\mathcal{E}_{ik} \uparrow \mathcal{S}_i$. Then the product-$\sigma$-algebra $\mathscr{S}_{[\![n]\!]} = \bigotimes_{i\in[\![n]\!]} \mathscr{S}_i$ is generated by the class of sets $\{\times_{i\in[\![n]\!]} \mathcal{E}_i \colon \mathcal{E}_i \in \mathscr{E}_i, i \in [\![n]\!]\}$.*

§04.04 **Proof** of Lemma §04.03. is given in the lecture. □

§04.05 **Remark**. Let $\mathscr{S}_1 = \{\emptyset, \mathcal{S}_1\}$ and $\mathscr{E}_1 = \{\emptyset\}$. Let $\mathscr{E}_2 = \mathscr{S}_2$ be a σ-algebra on $\mathcal{S}_2$ containing at least 4 elements. Then the class of sets $\{\emptyset \times E \colon E \in \mathscr{E}_2\}$ does not generate the product-σ-algebra $\mathscr{S}_1 \otimes \mathscr{S}_2$. Consequently, the restrictive assumption on the generator in Lemma §04.03 cannot simply be dispensed with. On the other hand side by applying Lemma §04.03 the product-σ-Algebra $\mathscr{S}_{[\![n]\!]} = \bigotimes_{i\in[\![n]\!]} \mathscr{S}_i$ is generated by the class of sets $\{\times_{i\in[\![n]\!]} \mathcal{E}_i \colon \mathcal{E}_i \in \mathscr{S}_i, i \in [\![n]\!]\}$ □

§04.06 **Definition**. A measure $\mu_{[\![n]\!]} \in \mathfrak{M}(\mathscr{S}_{[\![n]\!]})$ on $(\mathcal{S}_{[\![n]\!]}, \mathscr{S}_{[\![n]\!]})$ is called *product measure* if

$$\mu_{[\![n]\!]}\big(\times_{i\in[\![n]\!]} \mathcal{E}_i\big) = \mu_{[\![n]\!]}\Big(\bigcap_{i\in[\![n]\!]} \Pi_i^{-1}(\mathcal{E}_i)\Big) = \prod_{i\in[\![n]\!]} \mu_i(\mathcal{E}_i) \quad \text{for} \quad \mathcal{E}_i \in \mathscr{S}_i,\ i \in [\![n]\!].$$

In this case we write $\bigotimes_{i\in[\![n]\!]} \mu_i := \mu_{[\![n]\!]}$. If $\mu_i = \mu$ for all $i \in [\![n]\!]$, then we write $\mu^n := \mu_{[\![n]\!]}$. □

§04.07 **Lemma** *(Uniqueness of finite product measures)*. *For every $i \in [\![n]\!]$ let $\mathscr{E}_i$ be a $\cap$-closed generator of the σ-algebra $\mathscr{S}_i$ on $\mathcal{S}_i$ and let* (uC) $(\mathcal{E}_{ik})_{k\in\mathbb{N}}$ *be a sequence in $\mathscr{E}_i$ such that $\mu_i(\mathcal{E}_{ik}) \in \mathbb{R}_{\geqslant 0}$ for every $k \in \mathbb{N}$ and $\mathcal{E}_{ik} \uparrow \mathcal{S}_i$. Then there is at most one measure $\mu_{[\![n]\!]} \in \mathfrak{M}(\mathscr{S}_{[\![n]\!]})$ on $(\mathcal{S}_{[\![n]\!]}, \mathscr{S}_{[\![n]\!]})$ with*

$$\mu_{[\![n]\!]}\big(\times_{i\in[\![n]\!]} \mathcal{E}_i\big) = \prod_{i\in[\![n]\!]} \mu_i(\mathcal{E}_i) \quad \text{for} \quad \mathcal{E}_i \in \mathscr{E}_i,\ i \in [\![n]\!].$$

§04.08 **Proof** of Lemma §04.07. is given in the lecture. □

§04.09 **Remark**. Under the assumptions of Lemma §04.07 follows immediately that for every $i \in [\![n]\!]$ the measure $\mu_i \in \mathfrak{M}_\sigma(\mathscr{S}_i)$ is σ-finite. □

§04.10 **Notation**. For $i \in [\![2]\!]$ let $(\mathcal{S}_i, \mathscr{S}_i)$ be a measurable space. For all $\mathcal{E} \subseteq \mathcal{S}_1 \times \mathcal{S}_2$, $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$ we denote with $\mathcal{E}_{s_1} := \{s \in \mathcal{S}_2 \colon (s_1, s) \in \mathcal{E}\} \subseteq \mathcal{S}_2$ or $\mathcal{E}^{s_2} := \{s \in \mathcal{S}_1 \colon (s, s_2) \in \mathcal{E}\} \subseteq \mathcal{S}_1$, respectively, the *$s_1$-slice* or *$s_2$-slice* of $\mathcal{E}$. □

§04.11 **Lemma**. *For all $\mathcal{E} \in \mathscr{S}_1 \otimes \mathscr{S}_2$, $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$ we have $\mathcal{E}_{s_1} \in \mathscr{S}_2$ und $\mathcal{E}^{s_2} \in \mathscr{S}_1$.*

§04.12 **Proof** of Lemma §04.11. is given in the lecture. □

§04.13 **Remark**. Due to Lemma §04.11 $\mu_2(\mathcal{E}_{s_1})$ and $\mu_1(\mathcal{E}^{s_2})$ are well-defined for all $\mathcal{E} \in \mathscr{S}_1 \otimes \mathscr{S}_2$, $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$. □

End LE07 2026/05/08

§04.14 **Lemma**. *For $i \in [\![2]\!]$ let $\mu_i \in \mathfrak{M}_\sigma(\mathscr{S}_i)$ be a σ-finite measure on $(\mathcal{S}_i, \mathscr{S}_i)$. Then, for all $\mathcal{E} \in \mathscr{S}_1 \otimes \mathscr{S}_2$, the map $\mu_2(\mathcal{E}_\bullet) : s_1 \mapsto \mu_2(\mathcal{E}_{s_1})$ and $\mu_1(\mathcal{E}^\bullet) : s_2 \mapsto \mu_1(\mathcal{E}^{s_2})$ defined on $\mathcal{S}_1$ respectively $\mathcal{S}_2$ is positive numerical, that is, $\mu_2(\mathcal{E}_\bullet) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S}_1)$ and $\mu_1(\mathcal{E}^\bullet) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S}_2)$.*

§04.15 **Proof** of Lemma §04.14. is given in the lecture. □

§04.16 **Theorem** *(Existence of a product measure)*. *For $i \in [\![2]\!]$ let $\mu_i \in \mathfrak{M}_\sigma(\mathscr{S}_i)$ be a σ-finite measure on $(\mathcal{S}_i, \mathscr{S}_i)$. Then there exists a unique product measure $\mu_{[\![2]\!]}$ on $(\mathcal{S}_{[\![2]\!]}, \mathscr{S}_{[\![2]\!]})$. Moreover, $\mu_{[\![2]\!]} \in \mathfrak{M}_\sigma(\mathscr{S}_{[\![2]\!]})$ is also σ-finite and $\mu_1\big(\mu_2(\mathcal{E}_\bullet)\big) = \mu_{[\![2]\!]}(\mathcal{E}) = \mu_2\big(\mu_1(\mathcal{E}^\bullet)\big)$ for all $\mathcal{E} \in \mathscr{S}_{[\![2]\!]}$.*

§04.17 **Proof** of Theorem §04.16. is given in the lecture. □

§04.18 **Remark**. The last statement can easily be extended to a finite product measure. It should be noted that the parentheses in the products can be arbitrarily rearranged. Formally we identify the product sets $\mathcal{S}_{[\![n-1]\!]} \times \mathcal{S}_n$ und $\mathcal{S}_{[\![n]\!]}$ as usual with help of the bijection $(s_{[\![n-1]\!]}, s_n) \mapsto s_{[\![n]\!]} = (s_i)_{i\in[\![n]\!]}$. The agreed equality of the sets implies then directly the equality of the corresponding products of σ-algebras $\mathscr{S}_{[\![n-1]\!]} \otimes \mathscr{S}_n$ and $\mathscr{S}_{[\![n]\!]}$ and the associative property $(\bigotimes_{i\in[\![m]\!]} \mathscr{S}_i) \otimes (\bigotimes_{i\in[\![n-m]\!]} \mathscr{S}_{m+i}) = \bigotimes_{i\in[\![n]\!]} \mathscr{S}_i$ for $m \in [\![n-1]\!]$. □

§04.19 **Corollary** *(Existence of product measures)*. *For $i \in [\![n]\!]$ let $\mu_i \in \mathfrak{M}_\sigma(\mathscr{S}_i)$ be a σ-finite measure on $(\mathcal{S}_i, \mathscr{S}_i)$. Then there exists a unique σ-finite product measure $\mu_{[\![n]\!]} \in \mathfrak{M}_\sigma(\mathscr{S}_{[\![n]\!]})$ on $(\mathcal{S}_{[\![n]\!]}, \mathscr{S}_{[\![n]\!]})$.*

§04.20 **Proof** of Corollary §04.19. is given in the lecture. □

§04.21 **Remark**. For measures that are not necessarily σ-finite, it is still possible to prove the existence, but no longer the uniqueness, of a product measure. □

§04|02 Projective family

§04.22 **Reminder**. If (Ω_i, τ_i), $i \in \mathcal{I}$, are topological spaces, then the product topology τ on $\Omega_{\mathcal{I}}$ is the coarsest topology with respect to which all coordinate maps $\Pi_i : \Omega_{\mathcal{I}} \to \Omega_i$ are continuous. □

§04.23 **Lemma**. *Let $\mathcal{I}$ be countable, for every $i \in \mathcal{I}$ let $\mathcal{S}_i$ be a separable, complete metric space (Polish) with Borel σ-algebra $\mathscr{B}_i := \mathscr{B}_{\mathcal{S}_i}$ and let $\mathscr{B}_{\mathcal{S}_\mathcal{I}}$ be the Borel σ-algebra with respect to the product topology on $\mathcal{S}_\mathcal{I} = \times_{i\in\mathcal{I}} \mathcal{S}_i$. Then $\mathcal{S}_\mathcal{I}$ is Polish and $\mathscr{B}_{\mathcal{S}_\mathcal{I}} = \mathscr{B}_\mathcal{I} = \bigotimes_{i\in\mathcal{I}} \mathscr{B}_i$. In particular, $\mathscr{B}_{\mathbb{R}^n} = \mathscr{B}^n$ for $n \in \mathbb{N}$.*

§04.24 **Proof** of Lemma §04.23. We refer to Klenke (2020, Theorem 14.8, p.303) or Bauer (1992, Theorem 22.1, p.151). □

§04.25 **Definition**. Let $\mathcal{I}$ be an arbitrary nonempty index set and for any $\mathcal{J} \subseteq \mathcal{I}$ let $\Pi_\mathcal{J}$ be the canonical projection on $(\mathcal{S}_\mathcal{I}, \mathscr{S}_\mathcal{I})$. For any $\mathcal{E} \in \mathscr{S}_\mathcal{J}$, $\Pi_\mathcal{J}^{-1}(\mathcal{E}) \in \mathscr{S}_\mathcal{I}$ is called a *cylinder set* with base $\mathcal{J}$. The set of such cylinder sets is denoted by $\mathcal{Z}_\mathcal{J} := \{\Pi_\mathcal{J}^{-1}(\mathcal{E})\colon \mathcal{E} \in \mathscr{S}_\mathcal{J}\} \subseteq \mathscr{S}_\mathcal{I}$. In particular, if $\mathcal{E}_\mathcal{J} = \times_{j\in\mathcal{J}} \mathcal{E}_j \in \mathscr{S}_\mathcal{J}$, then $\Pi_\mathcal{J}^{-1}(\mathcal{E}) \in \mathscr{S}_\mathcal{I}$ is called a *rectangular cylinder* with base $\mathcal{J}$. The set of such rectangular cylinders will be denoted by $\mathcal{Z}_\mathcal{J}^R := \{\Pi_\mathcal{J}^{-1}(\mathcal{E}_\mathcal{J})\colon \mathcal{E}_\mathcal{J} = \times_{j\in\mathcal{J}} \mathcal{E}_j \in \mathscr{S}_\mathcal{J}\} \subseteq \mathscr{S}_\mathcal{I}$. For every $i \in \mathcal{I}$ let $\mathscr{E}_i \subseteq \mathscr{S}_i$. The set of rectangular cylinders for which in addition $\mathcal{E}_j \in \mathscr{E}_j$ for all $j \in \mathcal{J}$ holds will be denoted by $\mathcal{Z}_\mathcal{J}^{\mathscr{E},R} := \{\Pi_\mathcal{J}^{-1}(\mathcal{E}_\mathcal{J})\colon \mathcal{E}_\mathcal{J} = \times_{j\in\mathcal{J}} \mathcal{E}_j, \mathcal{E}_j \in \mathscr{E}_j, j \in \mathcal{J}\} \subseteq \mathscr{S}_\mathcal{I}$. Write $\mathcal{Z} := \bigcup\{\mathcal{Z}_\mathcal{J}\colon \mathcal{J} \subseteq \mathcal{I} \text{ finite}\}$ and similarly define $\mathcal{Z}^R$ and $\mathcal{Z}^{\mathscr{E},R}$. □

§04.26 **Remark**. Every $\mathcal{Z}_\mathcal{J}$ is a σ-algebra, and $\mathcal{Z}$ is a algebra where $\mathscr{S}_\mathcal{I} = \sigma(\mathcal{Z})$. Moreover, if every $\mathscr{E}_i$ is $\cap$-closed, then $\mathcal{Z}^{\mathscr{E},R}$ is also $\cap$-closed (Exercise). □

§04.27 **Lemma**. *For any $i \in \mathcal{I}$ let $\mathscr{E}_i \subseteq \mathscr{S}_i$ be a generator of $\mathscr{S}_i$.*

(i) $\mathscr{S}_\mathcal{J} = \sigma(\times_{j\in\mathcal{J}} \mathcal{E}_j : \mathcal{E}_j \in \mathscr{E}_j, j \in \mathcal{J})$ *for every finite $\mathcal{J} \subseteq \mathcal{I}$.*

(ii) $\mathscr{S}_\mathcal{I} = \sigma(\mathcal{Z}^R) = \sigma(\mathcal{Z}^{\mathscr{E},R})$.

(iii) *Let* (A1) $\mu \in \mathfrak{M}_\sigma(\mathscr{S}_\mathcal{I})$ *be a σ-finite measure on $(\mathcal{S}_\mathcal{I}, \mathscr{S}_\mathcal{I})$, assume* (A2) *every $\mathscr{E}_i$ is $\cap$-closed, and* (A3) *there is a sequence $(\mathcal{E}_n)_{n\in\mathbb{N}}$ in $\mathcal{Z}^{\mathscr{E},R}$ with $\mathcal{E}_n \uparrow \mathcal{S}_\mathcal{I}$ and $\mu(\mathcal{E}_n) \in \mathbb{R}_{\geqslant 0}$ for all $n \in \mathbb{N}$. Then μ is uniquely determined by the values $\mu(A)$ for all $A \in \mathcal{Z}^{\mathscr{E},R}$.*

§04.28 **Proof** of Lemma §04.27. (i) and (ii) are exercises. (iii) follows immediately from §01.28, since $\mathcal{Z}^{\mathscr{E},R}$ is a $\cap$-closed generator of $\mathcal{S}_\mathcal{I}$. □

§04.29 **Comment**. The condition (A3) in Lemma §04.27 (iii) is fulfilled, if $\mu \in \mathfrak{M}_{\mathrm{fi}}(\mathscr{S}_\mathcal{I})$ is finite and $\mathcal{S}_i \in \mathscr{E}_i$ for every $i \in \mathcal{I}$ (compare Lemma §01.28). □

§04.30 **Notation**. For $\mathcal{J} \subseteq \mathcal{K} \subseteq \mathcal{I}$ the map $\Pi_\mathcal{J}^\mathcal{K} : \mathcal{S}_\mathcal{K} \ni s_\mathcal{K} = (s_k)_{k\in\mathcal{K}} \mapsto (s_j)_{j\in\mathcal{J}} = s_\mathcal{J} \in \mathcal{S}_\mathcal{J}$ is called *canonical projection*, where evidently $\Pi_\mathcal{J} = \Pi_\mathcal{J}^\mathcal{I}$. □

§04.31 **Definition**. For every finite $\mathcal{J} \subseteq \mathcal{I}$ let $\mathbb{P}_\mathcal{J} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{S}_\mathcal{J})$ be a probability measure on $(\mathcal{S}_\mathcal{J}, \mathscr{S}_\mathcal{J})$. The familiy $\{\mathbb{P}_\mathcal{J}\colon \mathcal{J} \subseteq \mathcal{I} \text{ finite}\}$ is called *projective* or *consistent* if $\mathbb{P}_\mathcal{J} = \mathbb{P}_\mathcal{K} \circ (\Pi_\mathcal{J}^\mathcal{K})^{-1}$ for all finite $\mathcal{J} \subseteq \mathcal{K} \subseteq \mathcal{I}$. □

§04.32 **Remark**. Let $\mathbb{P} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{S}_\mathcal{I})$ be a probability measure on $(\mathcal{S}_\mathcal{I}, \mathscr{S}_\mathcal{I})$. Since $\Pi_\mathcal{J} = \Pi_\mathcal{J}^\mathcal{K} \circ \Pi_\mathcal{K}$, the family $\{\mathbb{P}_\mathcal{J} := \mathbb{P} \circ \Pi_\mathcal{J}^{-1}\colon \mathcal{J} \subseteq \mathcal{I} \text{ finite}\}$ is consistent. Thus, consistency is a necessary condition for the existence of a measure $\mathbb{P}$ on the product space with $\mathbb{P}_\mathcal{J} := \mathbb{P} \circ \Pi_\mathcal{J}^{-1}$. If all the measurable spaces are Polish spaces then this condition is also sufficient. □

§04.33 **Kolmogorov's extension theorem**. *Let $\mathcal{I}$ be an arbitrary nonempty index set and let $\mathcal{S}_i$ be a separable and complete metric space (Polish) with Borel σ algebra $\mathscr{B}_i := \mathscr{B}_{\mathcal{S}_i}$ for all $i \in \mathcal{I}$. Let $\{\mathbb{P}_\mathcal{J}\colon \mathcal{J} \subseteq \mathcal{I} \text{ finite}\}$ be a consistent family of probability measures. Then there exists a unique probability measure $\mathbb{P} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{B}_\mathcal{I})$ on $(\mathcal{S}_\mathcal{I}, \mathscr{B}_\mathcal{I})$ with $\mathbb{P}_\mathcal{J} = \mathbb{P} \circ \Pi_\mathcal{J}^{-1}$ for all finite $\mathcal{J} \subseteq \mathcal{I}$. $\mathbb{P}$ is called* projective limit.

§04.34 **Proof** of Theorem §04.33. We refer to Klenke (2020, Theorem 14.39, p.319) □

§04.35 **Definition**. Let $\mathbb{P}_i \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{S}_i)$ be a probability measure on $(\mathcal{S}_i, \mathscr{S}_i)$ for all $i \in \mathcal{I}$. A probability measure $\mathbb{P}_{\mathcal{I}} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{S}_{\mathcal{I}})$ on $(\mathcal{S}_{\mathcal{I}}, \mathscr{S}_{\mathcal{I}})$ is called *product measure* of the $\mathbb{P}_i$, $i \in \mathcal{I}$, if

$$\mathbb{P}_{\mathcal{I}}\Big(\bigcap_{j\in\mathcal{J}} \Pi_j^{-1}(\mathcal{E}_j)\Big) = \prod_{j\in\mathcal{J}} \mathbb{P}_j(\mathcal{E}_j) \quad \text{for} \quad \mathcal{E}_j \in \mathscr{S}_j,\ j \in \mathcal{J} \subseteq \mathcal{I} \text{ finite.}$$

In this case we write $\bigotimes_{i\in\mathcal{I}} \mathbb{P}_i := \mathbb{P}_{\mathcal{I}}$. If $\mathbb{P}_i = \mathbb{P}$ for all $i \in \mathcal{I}$ then $\mathbb{P}^{\mathcal{I}} := \mathbb{P}_{\mathcal{I}}$ and $\mathbb{P}^n := \mathbb{P}_{\mathcal{I}}$ in case $n := |\mathcal{I}| \in \mathbb{N}$. □

§04.36 **Remark**. Let $\mathcal{I}$ be an arbitrary nonempty index set. For every $i \in \mathcal{I}$ let $\mathcal{S}_i$ be a separable and complete metric space (Polish) with Borel σ-algebra $\mathscr{B}_i := \mathscr{B}_{\mathcal{S}_i}$ and $\mathbb{P}_i \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{B}_i)$ be a probability measure on $(\mathcal{S}_i, \mathscr{B}_i)$. For every finite $\mathcal{J} \subseteq \mathcal{I}$ let $\mathbb{P}_{\mathcal{J}} := \bigotimes_{j\in\mathcal{J}} \mathbb{P}_j$ be the finite product measure of the $\mathbb{P}_j$, $j \in \mathcal{J}$. Evidently, the family $\{\mathbb{P}_{\mathcal{J}} \colon \mathcal{J} \subseteq \mathcal{I} \text{ finite}\}$ is *projective*. Making use of Theorem §04.33 there exists a unique product measure $\mathbb{P}_{\mathcal{I}} := \bigotimes_{i\in\mathcal{I}} \mathbb{P}_i \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{B}_{\mathcal{I}})$ on $(\mathcal{S}_{\mathcal{I}}, \mathscr{B}_{\mathcal{I}})$. Considering the canonical process $(\Pi_i)_{i\in\mathcal{I}}$ under $\mathbb{P}_{\mathcal{I}}$, all coordinate maps Π_i are independent, i.e. $\perp\!\!\!\perp_{i\in\mathcal{I}} \Pi_i$. □

§04|03 Integration with respect to product measures

§04.37 **Notation**. Let $h : \mathcal{S}_1 \times \mathcal{S}_2 \to \mathcal{S}_3$ be a map. For all $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$ we denote by $h_{s_1} : \mathcal{S}_2 \to \mathcal{S}_3$ with $s_2 \mapsto h_{s_1}(s_2) := h(s_1, s_2)$ and $h^{s_2} : \mathcal{S}_1 \to \mathcal{S}_3$ with $s_1 \mapsto h^{s_2}(s_1) := h(s_1, s_2)$, respectively, the *$s_1$-slice* and *$s_2$-slice* of h. □

§04.38 **Lemma**. *For $i \in [\![3]\!]$, let $(\mathcal{S}_i, \mathscr{S}_i)$ be a measurable space. For all $h \in \mathcal{M}(\mathscr{S}_1 \otimes \mathscr{S}_2, \mathscr{S}_3)$, $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$ we have $h_{s_1} \in \mathcal{M}(\mathscr{S}_2, \mathscr{S}_3)$ and $h^{s_2} \in \mathcal{M}(\mathscr{S}_1, \mathscr{S}_3)$.*

§04.39 **Proof** of Lemma §04.38. is given in the lecture. □

§04.40 **Theorem** *(Tonelli)*. *For $i \in [\![2]\!]$ let $\mu_i \in \mathfrak{M}_\sigma(\mathscr{S}_i)$ be a σ-finite measure on $(\mathcal{S}_i, \mathscr{S}_i)$. Then, for every $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S}_1 \otimes \mathscr{S}_2)$ the map $\mu_1(h^\bullet) : s_2 \mapsto \mu_1(h^{s_2})$ and $\mu_2(h_\bullet) : s_1 \mapsto \mu_2(h_{s_1})$ defined on $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively, is positive numerical, that is, $\mu_1(h^\bullet) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S}_2)$ and $\mu_2(h_\bullet) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S}_1)$. Moreover, it holds*

$$\begin{aligned}(\mu_1 \otimes \mu_2)(h) = \mu_2(\mu_1(h^\bullet)) = \int \mu_1(h^{s_2})\mu_2(\mathrm{d}s_2) &= \int\int h(s_1, s_2)\mu_1(\mathrm{d}s_1)\mu_2(\mathrm{d}s_2) \\ &= \int \mu_2(h_{s_1})\mu_1(\mathrm{d}s_1) = \mu_1(\mu_2(h_\bullet))\end{aligned}$$

§04.41 **Proof** of Theorem §04.40. is given in the lecture. □

§04.42 **Definition**. Let $(\Omega, \mathscr{A}, \mu)$ be a measure space, $(\mathcal{S}, \mathscr{S})$ be a measurable space, and $N \in \mathscr{A}$ be an μ-null set. A function $h : N^c := \Omega \backslash N \to \mathcal{S}$ is called *μ-almost everywhere defined* and *$\mathscr{A}$-$\mathscr{S}$-measurable* if $h^{-1}(\mathscr{S}) \subseteq \mathscr{A}$ or in equal $\tilde{h} \in \mathcal{M}(\mathscr{A}, \mathscr{S})$ with $\tilde{h}|_{N^c} = h$ holds.

§04.43 **Remark**. If $h, g \in \overline{\mathcal{M}}(\mathscr{A})$ and $h, g \in \mathbb{R}$ μ-almost everywhere, then the function $g - h$ is μ-almost everywhere defined and $\mathscr{A}$-$\overline{\mathscr{B}}$-measurable. This holds in particular if g and h are μ-integrable. Now, if f is $\overline{\mathbb{R}}$-valued, μ-almost everywhere defined with μ-null set N and $\mathscr{A}$-$\overline{\mathscr{B}}$-measurable, then there exists $\tilde{f} \in \overline{\mathcal{M}}(\mathscr{A})$ with $\tilde{f}|_{N^c} = f$. Evidently, for any other $\tilde{f}_2 \in \overline{\mathcal{M}}(\mathscr{A})$ with $\tilde{f}_2|_{N^c} = f$ we have $\tilde{f} = \tilde{f}_2$ μ-a.e.. Consequently, if $\tilde{f}$, and hence $\tilde{f}_2$, is furthermore μ-integrable, then we define for f the integral $\mu(f) = \int f \,\mathrm{d}\mu := \mu(\tilde{f}) = \mu(\tilde{f}_2)$. Finally, if $h, g \in \overline{\mathcal{M}}(\mathscr{A})$ are μ-integrable, then $h - g \in \overline{\mathcal{M}}(\mathscr{A})$ is μ-almost everywhere defined, $\mathscr{A}$-$\overline{\mathscr{B}}$-measurable, and $\mu(h - g) = \mu(h) - \mu(g)$. □

§04.44 **Corollary** *(Fubini's theorem). Let $(\mathcal{S}_i, \mathscr{S}_i, \mu_i)$, $i \in [\![2]\!]$, be σ-finite measure spaces and $h \in \mathcal{L}^1(\mu_1 \otimes \mu_2)$. Then $\mu_2(h_\bullet) : s_1 \mapsto \mu_2(h_{s_1})$ is μ_1-almost everywhere defined and $\mathscr{S}_1$-$\overline{\mathscr{B}}$-measurable, and $\mu_1(h^\bullet) : s_2 \mapsto \mu_1(h^{s_2})$ is μ_2-almost everywhere defined and $\mathscr{S}_2$-$\overline{\mathscr{B}}$-measurable. It holds that*

$$\mu_2(\mu_1(h^\bullet)) = \int \mu_1(h^{s_2})\mu_2(\mathrm{d}s_2) = (\mu_1 \otimes \mu_2)(h) = \int \mu_2(h_{s_1})\mu_1(\mathrm{d}s_1) = \mu_1(\mu_2(h_\bullet)).$$

§04.45 **Proof** of Corollary §04.44. is given in the lecture. □

§04.46 **Remark**. The last statements can be easily extended to finite product measures, as in Remark §04.18. □

End LE08 2026/05/13

§04.47 **Theorem**. *For each $i \in [\![n]\!]$, let $(\mathcal{S}_i, \mathscr{S}_i, \mu_i)$ be a σ-finite measure space, $\mathfrak{f}_i \in \mathcal{M}_{\geqslant 0}(\mathscr{S}_i)$, and $\nu_i := \mathfrak{f}_i\mu_i \in \mathfrak{M}_\sigma(\mathscr{S})$. Then the product measure $\nu_{[\![n]\!]} = \bigotimes_{i\in[\![n]\!]} \nu_i \in \mathfrak{M}_\sigma(\mathscr{S}_{[\![n]\!]})$ is σ-finite and absolutely continuous with respect to the product measure $\mu_{[\![n]\!]} = \bigotimes_{i\in[\![n]\!]} \mu_i \in \mathfrak{M}_\sigma(\mathscr{S}_{[\![n]\!]})$ with product density $\bigotimes_{i\in[\![n]\!]} \mathfrak{f}_i := \prod_{i\in[\![n]\!]} \mathfrak{f}_i(\Pi_i) \in \mathcal{M}_{\geqslant 0}(\mathscr{S}_{[\![n]\!]})$, meaning $\nu_{[\![n]\!]} = \bigotimes_{i\in[\![n]\!]} \nu_i = \bigotimes_{i\in[\![n]\!]} (\mathfrak{f}_i\mu_i) = (\bigotimes_{i\in[\![n]\!]} \mathfrak{f}_i)\, \mu_{[\![n]\!]}$.*

§04.48 **Proof** of Theorem §04.47. is given in the lecture. □

§04.49 **Reminder**. Now, let $\nu = \mathbb{P}_0$ and $\mu = \mathbb{P}_1$ be probability measures on $(\mathcal{S}, \mathscr{S})$, where it is not necessarily the case that $\mathbb{P}_0 \ll \mathbb{P}_1$. Then any positive, measurable function $\mathrm{L} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$ with $\mathbb{P}_0 = \mathrm{L}\mathbb{P}_1 + \mathbb{1}_{\mathrm{L}=\infty}\mathbb{P}_0$ and $\mathbb{P}_1(\mathrm{L} \in \mathbb{R}_{\geqslant 0}) = 1$ is is a $\mathbb{P}_1$-likelihood ratio of $\mathbb{P}_0$ (cf. Definition §03.14). Let $\mu \in \mathfrak{M}_\sigma(\mathscr{S})$ denote a σ-finite measure such that $\mathbb{P}_i \ll \mu$, $i \in [\![2]\!]$, (for example, the finite measure $\mu = \mathbb{P}_0 + \mathbb{P}_1$), and let $\mathfrak{f}_i \in \mathcal{M}_{\geqslant 0}(\mathscr{S})$ be a μ-density of $\mathbb{P}_i$, $i \in [\![2]\!]$. Then

$$\mathrm{L}_\star := \frac{\mathfrak{f}_0}{\mathfrak{f}_1}\mathbb{1}_{\{\mathfrak{f}_1\in\mathbb{R}_{>0}\}} + \infty\mathbb{1}_{\{\mathfrak{f}_1=0\}\cap\{\mathfrak{f}_0\in\mathbb{R}_{>0}\}} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$$

is a specific choice of the $\mathbb{P}_1$-likelihood ratio of $\mathbb{P}_0$. A $\mathbb{P}_0$-likelihood ratio of $\mathbb{P}_1$ is given by

$$(1/\mathrm{L}_\star) = \frac{\mathfrak{f}_1}{\mathfrak{f}_0}\mathbb{1}_{\{\mathfrak{f}_0\in\mathbb{R}_{>0}\}} + \infty\mathbb{1}_{\{\mathfrak{f}_0=0\}\cap\{\mathfrak{f}_1\in\mathbb{R}_{>0}\}} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$$

In the special case where $\mathbb{P}_0 \ll \mathbb{P}_1$, the $\mathbb{P}_1$-likelihood ratio of $\mathbb{P}_0$ is a $\mathbb{P}_1$-density of $\mathbb{P}_0$ and is $\mathbb{P}_1$-determined. □

§04.50 **Lemma**. *For each $i \in [\![n]\!]$, let $\mathbb{P}_{0|i}, \mathbb{P}_{1|i} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{S}_i)$ be probability measures on $(\mathcal{S}_i, \mathscr{S}_i)$ with $\mathbb{P}_{1|i}$-density ratio L_i of $\mathbb{P}_{0|i}$. Then the product $\mathrm{L} := \bigotimes_{i\in[\![n]\!]} \mathrm{L}_i = \prod_{i\in[\![n]\!]} \mathrm{L}_i(\Pi_i)$ is a density ratio of $\mathbb{P}_0 := \bigotimes_{i\in[\![n]\!]} \mathbb{P}_{0|i}$ with respect to $\mathbb{P}_1 := \bigotimes_{i\in[\![n]\!]} \mathbb{P}_{1|i}$.*

§04.51 **Proof** of Lemma §04.50. is given in the lecture. □

§04|04 Integration with respect to transition kernels

§04.52 **Definition**. Let $(\Omega, \mathscr{A})$ and $(\mathcal{S}, \mathscr{S})$ be two measurable spaces. A map $\boxed{\kappa} : \Omega \times \mathscr{S} \to \overline{\mathbb{R}}_{\geqslant 0}$ is called a $(\sigma$-$)$finite *transition kernel* from $(\Omega, \mathscr{A})$ to $(\mathcal{S}, \mathscr{S})$ if it satisfies the following two conditions:

(tK1) For all $\omega \in \Omega$ the ω-slice $\kappa_\omega : \mathscr{S} \to \overline{\mathbb{R}}_{\geqslant 0}$ with $B \mapsto \kappa_\omega(B) := \kappa(\omega, B)$ is a $(\sigma$-$)$finite measure on $(\mathcal{S}, \mathscr{S})$, i.e. $\boxed{\kappa_\bullet} : \Omega \to \mathfrak{M}_{\mathrm{fi}}(\mathscr{S})$, respectively, $\boxed{\kappa_\bullet} : \Omega \to \mathfrak{M}_\sigma(\mathscr{S})$, with $\omega \mapsto \kappa_\omega$.

(tK2) For all $B \in \mathscr{S}$ the B-slice $\kappa^B : \Omega \to \overline{\mathbb{R}}_{\geqslant 0}$ with $\omega \mapsto \kappa^B(\omega) := \kappa(\omega, B)$ is positive, numerical and $\mathscr{A}$-measurable, i.e. $\boxed{\kappa^\bullet} : \mathscr{S} \to \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ with $B \mapsto \kappa^B$.

If for every $\omega \in \Omega$, the measure in (tK1) is a probability measure, i.e. $\kappa_{\bullet} : \Omega \to \mathfrak{M}_{\mathrm{pr}}(\mathscr{S})$, and hence $\kappa^{\bullet} : \mathscr{S} \to \mathcal{M}_{[0,1]}(\mathscr{A})$, then κ is called a *Markov kernel*.

If there exists σ-finite measure $\nu \in \mathfrak{M}_{\sigma}(\mathscr{S})$ and a function $\mathbb{f} \in \mathcal{M}_{\geqslant 0}(\mathscr{A} \otimes \mathscr{S})$ such that for every $\omega \in \Omega$ the ω-slice $\mathbb{f}_{\omega} \in \mathcal{M}_{\geqslant 0}(\mathscr{S})$ is a ν-density of κ_{ω} then $\mathbb{f}$ is called *ν-density* of the transition kernel κ and we write $\kappa_{\bullet} = \mathbb{f}_{\bullet}\nu : \Omega \ni \omega \mapsto \kappa_{\omega} = \mathbb{f}_{\omega}\nu \in \mathfrak{M}_{\sigma}(\mathscr{S})$. □

§04.53 **Remark**. It suffices to verify condition (tK2) for sets from a $\cap$-closed generator $\mathscr{E}$ of $\mathscr{S}$ only, which contains $\mathcal{S}$ or a sequence $(\mathcal{E}_n)_{n \in \mathbb{N}}$ of sets such that $\mathcal{E}_n \uparrow \mathcal{S}$. Then $\mathscr{D} := \{B \in \mathscr{S} : \kappa^B \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})\}$ is a Dynkin system (exercise) with $\mathscr{E} \subseteq \mathscr{D} \subseteq \mathscr{S}$, and from the π-λ-Theorem §01.11, it follows that $\mathscr{D} = \sigma(\mathscr{E}) = \mathscr{S}$. □

§04.54 **Lemma**. *Let κ be a finite transition kernel from $(\Omega, \mathscr{A})$ to $(\mathcal{S}, \mathscr{S})$, and let $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A} \otimes \mathscr{S})$ be positive numerical. Then the function $\kappa_{\bullet}(h_{\bullet}) : \Omega \to \overline{\mathbb{R}}_{\geqslant 0}$ defined by*

$$\omega \mapsto \kappa_{\omega}(h_{\omega}) = \int h_{\omega} \, \mathrm{d}\kappa_{\omega} = \int h_{\omega}(s) \kappa_{\omega}(\mathrm{d}s)$$

belongs to $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$. If $\mathbb{f}$ is ν-density of the transition kernel κ, then

$$\kappa_{\bullet}(h_{\bullet}) = \mathbb{f}_{\bullet}\nu(h_{\bullet}) = \nu(\mathbb{f}_{\bullet} h_{\bullet}) \text{ with } \omega \mapsto \nu(\mathbb{f}_{\omega} h_{\omega}) = \int h_{\omega} \mathbb{f}_{\omega} \, \mathrm{d}\nu = \int h_{\omega}(s) \mathbb{f}_{\omega}(s) \nu(\mathrm{d}s).$$

§04.55 **Proof** of Lemma §04.54. is given in the lecture. □

§04.56 **Notation**. For $\mathbb{1}_D \in \mathcal{M}_{\geqslant 0}(\mathscr{A} \otimes \mathscr{S})$, that is, $D \in \mathscr{A} \otimes \mathscr{S}$, according to Lemma §04.54, the function $\kappa_{\bullet}(D_{\bullet}) = \kappa_{\bullet}\big((\mathbb{1}_D)_{\bullet}\big) : \Omega \to \overline{\mathbb{R}}_{\geqslant 0}$ defined by

$$\omega \mapsto \kappa_{\omega}(D_{\omega}) = \kappa_{\omega}\big((\mathbb{1}_D)_{\omega}\big) = \int \mathbb{1}_D(\omega, s) \kappa_{\omega}(\mathrm{d}s)$$

belongs to $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$. □

§04.57 **Lemma**. *Let $(\Omega, \mathscr{A}, \mu)$ be a finite measure space, $(\mathcal{S}, \mathscr{S})$ be a measurable space, and κ be a finite transition kernel from $(\Omega, \mathscr{A})$ to $(\mathcal{S}, \mathscr{S})$. Then there exists a uniquely determined σ-finite measure $\mu \odot \kappa \in \mathfrak{M}_{\sigma}(\mathscr{A} \otimes \mathscr{S})$ on the product space $(\Omega \times \mathcal{S}, \mathscr{A} \otimes \mathscr{S})$ such that*

$$(\mu \odot \kappa)(D) = \mu(\kappa_{\bullet}(D_{\bullet})) \quad \text{for } D \in \mathscr{A} \otimes \mathscr{S},$$

where for all $A \in \mathscr{A}$ and $B \in \mathscr{S}$, we have

$$(\mu \odot \kappa)(A \times B) = \mu\big(\mathbb{1}_A \kappa^B\big) = \int_A \kappa^B \, \mathrm{d}\mu = \int_A \kappa(\omega, B) \mu(\mathrm{d}\omega).$$

If κ is a Markov kernel and μ is a probability measure, then $\mu \odot \kappa$ is a probability measure.

§04.58 **Proof** of Lemma §04.57. is given in the lecture. □

§04.59 **Theorem** *(Tonelli/Fubini for transition kernel)*. *Let $(\Omega, \mathscr{A}, \mu)$ be a finite measure space, $(\mathcal{S}, \mathscr{S})$ be a measurable space, and κ be a finite transition kernel from $(\Omega, \mathscr{A})$ to $(\mathcal{S}, \mathscr{S})$. If $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A} \otimes \mathscr{S})$ or $h \in \mathcal{L}^1(\mu \odot \kappa)$ then*

$$(\mu \odot \kappa)(h) = \mu(\kappa_{\bullet}(h_{\bullet})) = \int \kappa_{\omega}(h_{\omega}) \mu(\mathrm{d}\omega) = \int \Big(\int h_{\omega} \, \mathrm{d}\kappa_{\omega} \Big) \mu(\mathrm{d}\omega)$$
$$= \int \int h(\omega, s) \kappa(\omega, \mathrm{d}s) \mu(\mathrm{d}\omega).$$

§04.60 **Proof** of Theorem §04.59. is given in the lecture. □

§04.61 **Notation.** Consider a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, a measurable space $(\mathcal{S}, \mathscr{S})$, and a Markov kernel κ from $(\Omega, \mathscr{A})$ to $(\mathcal{S}, \mathscr{S})$. Due to Lemma §04.57, $\mathbb{P} \odot \kappa \in \mathfrak{M}_{\text{pr}}(\mathscr{A} \otimes \mathscr{S})$ is an uniquely determined probability measure on $(\Omega \times \mathcal{S}, \mathscr{A} \otimes \mathscr{S})$. Then we denote by

$$(\mathbb{P}\kappa)(S) := \mathbb{P}(\kappa^S) = \int \kappa^S \, d\mathbb{P} = \int \kappa(\omega, S)\mathbb{P}(d\omega), \quad \text{for } S \in \mathscr{S}$$

the marginal distribution $\mathbb{P}\kappa \in \mathfrak{M}_{\text{pr}}(\mathscr{S})$ on $(\mathcal{S}, \mathscr{S})$ induced by $\mathbb{P} \odot \kappa \in \mathfrak{M}_{\text{pr}}(\mathscr{A} \otimes \mathscr{S})$. □

Chapter 2

Conditional expectation

§05 Discretely or continuously distributed random variables

§05.01 **Reminder**. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $B \in \mathscr{A}$ with $\mathbb{P}(B) \in \mathbb{R}_{>0}$. For any $A \in \mathscr{A}$

$$\mathbb{P}\big(A\big|B\big) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \in \mathbb{R}_{\geqslant 0}$$

is called the *conditional probability of A given B*. □

§05.02 **Property**.

(i) *Let (S, X) be a discretely distributed random variable assuming values in $(\mathcal{S} \times \mathcal{X}, 2^{\mathcal{S}\times\mathcal{X}})$ and admitting a joint probability mass function $\mathbb{p}^{S,X} : \mathcal{S} \times \mathcal{X} \to [0, 1]$ with respect to the product counting measure $\zeta_{\mathcal{S}} \otimes \zeta_{\mathcal{X}}$ on $\mathcal{S} \times \mathcal{X}$. For any $s \in \mathcal{S}$ with strictly positive marginal probability mass function $\mathbb{p}^{S}(s) \in \mathbb{R}_{>0}$ of S at the point s the map*

$$\mathbb{p}^{X|S=s} : \mathcal{X} \to [0, 1] \text{ with } x \mapsto \mathbb{p}^{X|S=s}(x) := \frac{\mathbb{p}^{S,X}(x, s)}{\mathbb{p}^{S}(s)} \qquad (05.01)$$

is a probability mass function with respect to the counting measure $\zeta_{\mathcal{X}}$ on $\mathcal{X}$.

(ii) *Let (S, X) be a continuously distributed random vector with joint probability density $\mathbb{f}^{S,X} : \mathbb{R}^{m+n} \to \mathbb{R}_{\geqslant 0}$ with respect to the Lebesgue measure λ^{m+n} on $\mathbb{R}^{m+n}$. For any $s \in \mathbb{R}^{m}$ with strictly positive marginal density $\mathbb{f}^{S}(s) \in \mathbb{R}_{>0}$ of S at the point s the map*

$$\mathbb{f}^{X|S=s} : \mathbb{R}^{n} \to \mathbb{R}_{\geqslant 0} \text{ with } x \mapsto \mathbb{f}^{X|S=s}(x) := \frac{\mathbb{f}^{S,X}(s, x)}{\mathbb{f}^{S}(s)} \qquad (05.02)$$

is a probability density with respect to the Lebesgue measure λ^{n} on $\mathbb{R}^{n}$. □

§05.03 **Remark**. Since the sets $\{s \in \mathcal{S} : \mathbb{p}^{S}(s) = 0\}$ and $\{s \in \mathbb{R}^{m} : \mathbb{f}^{S}(s) = 0\}$ are null sets with respect to the distribution $\mathbb{P}^{S}$ of S, the functions $\mathbb{p}^{X|S=s}$ and $\mathbb{f}^{X|S=s}$ are indeed for $\mathbb{P}^{S}$-almost every $s \in \mathcal{S}$ and $s \in \mathbb{R}^{m}$ defined, i.e., defined outside a $\mathbb{P}^{S}$-null set only. We may arbitrarily choose densities on those null sets, for example, the marginal densities $\mathbb{p}^{X|S=s} = \mathbb{p}^{X}$ and $\mathbb{f}^{X|S=s} = \mathbb{f}^{X}$. □

§05.04 **Definition**.

(a) Let (S, X) be a discrete distributed random variable with values in $(\mathcal{S} \times \mathcal{X}, 2^{\mathcal{S}\times\mathcal{X}})$. For each $s \in \mathcal{S}$ with strictly positive marginal probability mass function of S at the point s, the distribution denoted by $\boxed{\mathbb{P}^{X|S=s}}$ with probability mass function $\boxed{\mathbb{p}^{X|S=s}}$ given in (05.01) is called the *conditional distribution* (with *conditional probability mass function*) of X given $S = s$. For $h \in \overline{\mathcal{M}}_{\geqslant 0}(2^{\mathcal{X}})$ or $h \in \mathcal{L}^{1}(\mathbb{P}^{X|S=s})$, the quantity

$$\boxed{\mathbb{E}^{X|S=s}(h)} := \mathbb{P}^{X|S=s}(h) = \sum_{x \in \mathcal{X}} h(x)\mathbb{p}^{X|S=s}(x) = \int h\, \mathbb{p}^{X|S=s}\, \mathrm{d}\zeta_{\mathcal{X}},$$

is called the *conditional expectation* of $h(X)$ given $S = s$.

(b) Let (S, X) be a continuous distributed random variable with values in $\mathbb{R}^{n+m}$. For each $s \in \mathbb{R}^m$ with strictly positive marginal density of S at the point s, the distribution denoted by $\mathbb{P}^{X|S=s}$ with probability density $\mathbb{f}^{X|S=s}$ given in (05.02) is called the *conditional distribution* (with *conditional probability density*) of X given $S = s$. For $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{B}^n)$ or $h \in \mathcal{L}^1(\mathbb{P}^{X|S=s})$, the quantity

$$\mathbb{E}^{X|S=s}(h) := \mathbb{P}^{X|S=s}(h) = \int h(x)\mathbb{f}^{X|S=s}(x)\lambda^n(\mathrm{d}x) = \int h\, \mathbb{f}^{X|S=s}\, \mathrm{d}\lambda^n$$

is called the *conditional expectation* of $h(X)$ given $S = s$. □

§05.05 **Remark.** Let's consider the continuous case (statements for the discrete case follow analogously). For each $h \in \mathcal{L}^1(\mathbb{P}^X)$ we have $\int |h|\mathbb{f}^{S,X}_s \,\mathrm{d}\lambda^n \in \mathbb{R}_{\geqslant 0}$ for $\mathbb{P}^S$-almost every $s \in \mathcal{S}$. Consequently, the conditional expectation $\mathbb{E}^{X|S=s}(h) = \mathbb{P}^{X|S=s}(h)$ of $h(X)$ given $S = s$ as well as the conditional density $\mathbb{f}^{X|S=s}$ exist for $\mathbb{P}^S$-almost every $s \in \mathcal{S}$, while on the corresponding $\mathbb{P}^S$-null set we may define them arbitrarily. Thus, different versions (specifications) only differ on a $\mathbb{P}^S$-null set. We see later that we can choose a version in such a way that the map $s \mapsto \varphi(s) := \mathbb{E}^{X|S=s}(h) = \mathbb{P}^{X|S=s}(h)$ is measurable. In this case, for all $B \in \mathscr{B}^m$:

$$\begin{aligned}\mathbb{E}^S(\mathbb{1}_B\varphi) = \mathbb{P}^S(\mathbb{1}_B\varphi) &= \int \mathbb{1}_B\varphi\, \mathbb{f}^S\, \mathrm{d}\lambda^m = \int_{\mathbb{R}^m} \mathbb{1}_B(s) \int_{\mathbb{R}^n} h(x)\, \mathbb{f}^{S,X}(s,x)\lambda^n(\mathrm{d}x)\lambda^m(\mathrm{d}s)\\ &= \int_{\mathbb{R}^{m+n}} \mathbb{1}_B(s)h(x)\, \mathbb{f}^{S,X}(s,x)\, \lambda^{m+n}\,(\mathrm{d}(s,x)) = \mathbb{P}(\mathbb{1}_{B}(S)h(X)) = \mathbb{E}(\mathbb{1}_{B}(S)h(X)).\end{aligned}$$

The measurability and the last equality are the starting point for the following more general approach, which allows us to cover both the discrete and continuous cases. The subsequent results and properties can then be transferred to the conditional densities, distributions, and expectations introduced so far. □

§05.06 **Example.**

(a) Let $X = |W_1 - W_2|$ and $S = W_1 + W_2$, where X represents the absolute difference and S represents the sum of the face values of two independent fair dices (W_1, W_2). Then, (S, X) is discretely distributed with values in $[\![0, 5]\!] \times [\![2, 12]\!]$, and the conditional probability mass functions of X given $S = s$ are as follows:

	$\mathbb{p}^{X\mid S=s}(x)$	s: 2	3	4	5	6	7	8	9	10	11	12	$\mathbb{p}^X(x)$
x	0	1	0	$\frac{1}{3}$	0	$\frac{1}{5}$	0	$\frac{1}{5}$	0	$\frac{1}{3}$	0	1	$\frac{6}{36}$
	1	0	1	0	$\frac{1}{2}$	0	$\frac{1}{3}$	0	$\frac{1}{2}$	0	1	0	$\frac{10}{36}$
	2	0	0	$\frac{2}{3}$	0	$\frac{2}{5}$	0	$\frac{2}{5}$	0	$\frac{2}{3}$	0	0	$\frac{8}{36}$
	3	0	0	0	$\frac{1}{2}$	0	$\frac{1}{3}$	0	$\frac{1}{2}$	0	0	0	$\frac{6}{36}$
	4	0	0	0	0	$\frac{2}{5}$	0	$\frac{2}{5}$	0	0	0	0	$\frac{4}{36}$
	5	0	0	0	0	0	$\frac{1}{3}$	0	0	0	0	0	$\frac{2}{36}$
	$\mathbb{p}^S(s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

(b) Let X and S be random vectors with

$$Y := \begin{pmatrix} S \\ X \end{pmatrix} \sim \mathrm{N}_{(\mu,\Sigma)} \text{ with } \mu = \begin{pmatrix} \mu_S \\ \mu_X \end{pmatrix} \in \mathbb{R}^{n+m} \text{ and}$$

$$\Sigma = \begin{pmatrix} \Sigma_S & \Sigma_{SX} \\ \Sigma_{XS} & \Sigma_X \end{pmatrix} > 0 \qquad \text{(strictly positive definit)},$$

hence $X \sim \mathrm{N}_{(\mu_X,\Sigma_X)}$, $S \sim \mathrm{N}_{(\mu_S,\Sigma_S)}$ and $\mathbb{C}\mathrm{ov}(X,S) = \Sigma_{XS} = \Sigma_{SX}^{\mathrm{T}}$. The conditional distribution of X given $S = s$ is a $\mathrm{N}_{(\mu_{X|S=s},\Sigma_{X|S=s})}$ distribution with

$$\mu_{X|S=s} := \mu_X + \Sigma_{XS}\Sigma_S^{-1}(s-\mu_S) \in \mathbb{R}^m \text{ and } \Sigma_{X|S=s} := \Sigma_X - \Sigma_{XS}\Sigma_S^{-1}\Sigma_{SX} > 0.$$

By definition, $\mu_{X|S=s}$ is precisely the conditional expectation of X given $S = s$. To prove this statement, we use the fact that for $\Sigma > 0$ we have $\Sigma_X > 0$, $\Sigma_S > 0$ and also $E := \Sigma_X - \Sigma_{XS}\Sigma_S^{-1}\Sigma_{SX} > 0$, setting $F := \Sigma_{XS}\Sigma_S^{-1}$ it follows

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_S^{-1} + F^{\mathrm{T}}E^{-1}F & -F^{\mathrm{T}}E^{-1} \\ -E^{-1}F & E^{-1} \end{pmatrix} > 0.$$

For the joint density $\mathbb{f}_{\mathrm{N}_{(\mu,\Sigma)}}$ of $Y = (S^{\mathrm{T}}, X^{\mathrm{T}})^{\mathrm{T}}$ and the marginal density $\mathbb{f}_{\mathrm{N}_{(\mu_S,\Sigma_S)}}$ of S, with $y = (x^{\mathrm{T}}, s^{\mathrm{T}})^{\mathrm{T}}$ and normalisation constant c, we have:

$$\begin{aligned}\frac{\mathbb{f}_{\mathrm{N}_{(\mu,\Sigma)}}(y)}{\mathbb{f}_{\mathrm{N}_{(\mu_S,\Sigma_S)}}(s)} &= c\exp\big\{-\tfrac{1}{2}(y-\mu)^{\mathrm{T}}\Sigma^{-1}(y-\mu) + \tfrac{1}{2}(s-\mu_S)^{\mathrm{T}}\Sigma^{-1}(s-\mu_S)\big\} \\ &= c\exp\big\{-\tfrac{1}{2}(x-(\mu_X+F(s-\mu_S)))^{\mathrm{T}}E^{-1}(x-(\mu_X+F(s-\mu_S)))\big\}.\end{aligned}$$

Since $\mu_{X|S=s} = \mu_X + F(s-\mu_S)$ and $\Sigma_{X|S=s} = E$ the claim follows. X^{-1} □

§06 Positive numerical random variables

In the following, let $(\Omega, \mathscr{A}, \mathbb{P})$ denote a probability space, $\mathbb{E}$ denote the expectation with respect to $\mathbb{P}$ and $\mathscr{F} \subseteq \mathscr{A}$ denotes a sub-σ-algebra of $\mathscr{A}$.

§06.01 **Reminder**. $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ denotes the set of all positive numerical random variables defined on $(\Omega, \mathscr{A})$. Since $\mathscr{F} \subseteq \mathscr{A}$, $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ is a subset of $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, and $\mathbb{E}(Y)$ is thus defined for any $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$. □

§06.02 **Theorem**. *For every $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, there exists a solution $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ to the Radon-Nikodym equation $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ for every $F \in \mathscr{F}$, where Y is unique up to $\mathbb{P}$-a.e. equality.*

End LE09 2026/05/15

§06.03 **Proof** of Theorem §06.02. is given in the lecture. □

§06.04 **Definition**. For $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, any function $Y : \Omega \to \overline{\mathbb{R}}_{\geqslant 0}$ is called a *version (specification) of the conditional expectation* of X given $\mathscr{F}$, symbolically $\boxed{\mathbb{E}(X|\mathscr{F})} := Y$, if it satisfies the following two conditions:

(cE1) Y is $\mathscr{F}$-$\overline{\mathscr{B}}_{\geqslant 0}$-measurable, i.e. $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ and

(cE2) $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ for every $F \in \mathscr{F}$. (*Radon-Nikodym equation*)

Any function

$$\boxed{\mathbb{E}(\bullet|\mathscr{F})} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \ni X \mapsto \mathbb{E}(X|\mathscr{F}) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$$

is called a *version (specification) of the conditional expectation* with respect to $\mathbb{P}$ given $\mathscr{F}$. The map implied by $\mathbb{E}(\bullet|\mathscr{F})$

$$\boxed{\mathbb{P}(\bullet|\mathscr{F})} : \Omega \times \mathscr{A} \to \overline{\mathbb{R}}_{\geqslant 0} \text{ with } (\omega, A) \mapsto \mathbb{P}(A|\mathscr{F})(\omega) := \mathbb{E}(\mathbb{1}_A|\mathscr{F})(\omega) \qquad (06.01)$$

is called a *version of the conditional distribution* of $\mathbb{P}$ given $\mathscr{F}$. For each $A \in \mathscr{A}$, we denote by $\boxed{\mathbb{P}(A|\mathscr{F})}$ the A-slice of $\mathbb{P}(\bullet|\mathscr{F})$, that is, $\mathbb{P}(A|\mathscr{F}) = \mathbb{E}(\mathbb{1}_A|\mathscr{F}) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ due to (cE1). By (cE2), every version $\mathbb{P}(A|\mathscr{F})$ is solution to the Radon-Nikodym equation

$$\mathbb{P}(F \cap A) = \mathbb{P}(\mathbb{1}_F\mathbb{P}(A|\mathscr{F})) = \int_F \mathbb{P}(A|\mathscr{F})(\omega)\mathbb{P}(\mathrm{d}\omega) \text{ for all } F \in \mathscr{F}. \qquad \square$$

§06.05 **Remark.** According to Theorem §06.02, for every $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, versions of the conditional expectation of X given $\mathscr{F}$ differ only on a $\mathbb{P}$-null set. This property generally does not transfer to versions of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$, as for every $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, we obtain a set of exceptions, and their union is generally not a $\mathbb{P}$-null set anymore. Considering a version $\mathbb{P}(\bullet|\mathscr{F})$ of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$, a desirable property is that for every $\omega \in \Omega$ the corresponding ω-slice, i.e. the mapping $\mathbb{P}(\bullet|\mathscr{F})(\omega) : \mathscr{A} \to [0, 1]$ with $A \mapsto \mathbb{P}(A|\mathscr{F})(\omega)$, is a probability measure on $(\Omega, \mathscr{A})$, symbolically $\mathbb{P}(\bullet|\mathscr{F})(\omega) \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$. $\square$

§06.06 **Reminder** *(compare Definition §04.52).* Let $(\mathcal{S}, \mathscr{S})$ and $(\Omega, \mathscr{A})$ be measurable spaces. A mapping $\boxed{\kappa} : \mathcal{S} \times \mathscr{A} \to \overline{\mathbb{R}}_{\geqslant 0}$ is called *Markov kernel* from $(\mathcal{S}, \mathscr{S})$ to $(\Omega, \mathscr{A})$, if it satisfies the following two conditions:

(Mk1) For all $s \in \mathcal{S}$ the s-slice $\kappa_s : \mathscr{A} \to [0, 1]$ with $A \mapsto \kappa_s(A) := \kappa(s, A)$ is a probability measure on $(\Omega, \mathscr{A})$, symbollically $\boxed{\kappa_\bullet} : \mathcal{S} \to \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ with $s \mapsto \kappa_s$.

(Mk2) For all $A \in \mathscr{A}$ the A-slice $\kappa^A : \mathcal{S} \to [0, 1]$ with $s \mapsto \kappa^A(s) := \kappa(s, A)$ is $\mathscr{S}$-$\mathscr{B}_{[0,1]}$-measurable, symbolically $\boxed{\kappa^\bullet} : \mathscr{A} \to \mathcal{M}_{[0,1]}(\mathscr{S})$ with $A \mapsto \kappa^A$. $\square$

§06.07 **Notation** *(compare §04.61).* Consider the probability space $(\Omega, \mathscr{A}, \mathbb{P})$, the measurable space $(\Omega, \mathscr{F})$ with sub-σ-algebra $\mathscr{F} \subseteq \mathscr{A}$, the restriction $\mathbb{P}_{\mathscr{F}} := \mathbb{P}|_{\mathscr{F}}$ of $\mathbb{P}$ to $\mathscr{F}$ and an arbitrary Markov kernel κ from $(\Omega, \mathscr{F}, \mathbb{P}_{\mathscr{F}})$ to $(\Omega, \mathscr{A})$. Then due to Lemma §04.57 there exists an uniquely determined probability measure $\boxed{\mathbb{P}_{\mathscr{F}} \odot \kappa}$ on $(\Omega \times \Omega, \mathscr{F} \otimes \mathscr{A})$ satisfying

$$(\mathbb{P}_{\mathscr{F}} \odot \kappa)(F \times A) = \mathbb{P}_{\mathscr{F}}(\mathbb{1}_F\kappa^A) = \int_F \kappa(\omega, A)\mathbb{P}_{\mathscr{F}}(\mathrm{d}\omega), \quad \text{for all } A \in \mathscr{A}, F \in \mathscr{F}.$$

If $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F} \otimes \mathscr{A})$ or $h \in \mathcal{L}^1(\mathbb{P}_{\mathscr{F}} \odot \kappa)$ then

$$(\mathbb{P}_{\mathscr{F}} \odot \kappa)(h) = \mathbb{P}_{\mathscr{F}}(\kappa_\bullet(h_\cdot)) = \int_\Omega \int_\Omega h(\omega_1, \omega_2)\kappa(\omega_1, \mathrm{d}\omega_2)\mathbb{P}_{\mathscr{F}}(\mathrm{d}\omega_1)$$

(see Theorem §04.59). Moreover, we denote by $\boxed{\mathbb{P}_{\mathscr{F}}\kappa} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ with

$$(\mathbb{P}_{\mathscr{F}}\kappa)(A) = (\mathbb{P}_{\mathscr{F}} \odot \kappa)(\Omega \times A) = \mathbb{P}_{\mathscr{F}}(\kappa^A) = \int_\Omega \kappa(\omega, A)\mathbb{P}_{\mathscr{F}}(\mathrm{d}\omega), \quad \text{for all } F \in \mathscr{F},$$

the marginal distribution on $(\Omega, \mathscr{A})$ induced by $\mathbb{P}_{\mathscr{F}} \odot \kappa \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{F} \otimes \mathscr{A})$. If the Markov kernel κ satisfies

$$\mathbb{P}(F \cap A) = (\mathbb{P}_{\mathscr{F}} \odot \kappa)(F \times A) = \mathbb{P}_{\mathscr{F}}(\mathbb{1}_F\kappa^A) = \mathbb{P}(\mathbb{1}_F\kappa^A), \quad \text{for all } A \in \mathscr{A}, F \in \mathscr{F}, \qquad (06.02)$$

and hence $\mathbb{P}_{\mathscr{F}}\kappa = \mathbb{P} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ then $\mathbb{P}(\bullet|\mathscr{F}) := \kappa$ is a version of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$. Indeed, (cE1) holds since $\mathbb{P}(A|\mathscr{F}) = \kappa^A \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ for all $A \in \mathscr{A}$, and by (06.02) $\mathbb{P}(A|\mathscr{F}) = \kappa^A$ satisfies (cE2). In this case we write $\mathbb{P}_{\mathscr{F}} \odot \mathbb{P}(\bullet|\mathscr{F})$ where $\mathbb{P}_{\mathscr{F}}\mathbb{P}(\bullet|\mathscr{F}) = \mathbb{P}$. $\square$

§06.08 **Definition.**

(a) $\mathbb{P}(\bullet|\mathscr{F})$ is called *regular version* of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$, if $\mathbb{P}(\bullet|\mathscr{F})$ is a Markov kernel from $(\Omega,\mathscr{F})$ to $(\Omega,\mathscr{A})$ fulfilling (06.02), that is, the map $\kappa := \mathbb{P}(\bullet|\mathscr{F}) : \Omega\times\mathscr{A}\to\overline{\mathbb{R}}_{\geqslant 0}$ with $(\omega, A)\mapsto \kappa(\omega, A) := \mathbb{P}(A|\mathscr{F})(\omega)$ satisfies the conditions (Mk1) and (Mk2), and every A-slice $\kappa^A = \mathbb{P}(A|\mathscr{F})$ is solution to the Radon-Nikodym equation (cE2).

(b) $\mathbb{E}(\bullet|\mathscr{F})$ is called *regular version* of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$, if the by (06.01) induced version $\mathbb{P}(\bullet|\mathscr{F})$ of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$ is regular, and for every $\omega\in\Omega$ the map $\mathbb{E}(\bullet|\mathscr{F})(\omega) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})\to\overline{\mathbb{R}}_{\geqslant 0}$ is the expectation with respect to the probability measure $\mathbb{P}(\bullet|\mathscr{F})(\omega)\in\mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$. □

§06.09 **Notation.** If $\mathbb{E}(\bullet|\mathscr{F})$ is a regular version of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$, and $\mathbb{P}(\bullet|\mathscr{F})$ is the by (06.01) induced regular version of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$, then we use for $X\in\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ both notations $\mathbb{E}(X|\mathscr{F}) = \mathbb{P}(X|\mathscr{F})$ as before. □

§06.10 **Theorem.**

(i) *To every regular version of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$, there exists a corresponding regular version of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$ that it implies.*

(ii) *For any probability measure $\mathbb{P}$ on $(\mathbb{R}^k,\mathscr{B}^k)$ and sub-σ-algebra $\mathscr{F}\subseteq\mathscr{B}^k$, there exists a regular version of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$.*

§06.11 **Proof** of Theorem §06.10. (i) using the proof strategy of approximation with simple random variables, (ii) Klenke (2020, Theorem 8.37, p.206) □

§06.12 **Remark.** Before we establish properties of the conditional expectation, we will construct versions of the conditional expectation or distribution in simple situations. It will be evident that even a regular version of the conditional distribution is not unique, but in many situations of interest to us, it is uniquely determined up to almost everywhere equality. □

§06|01 Simple conditions

We first consider a sub-σ-algebra $\mathscr{F}$ of $\mathscr{A}$ generated by a single measurable set $B\in\mathscr{A}$, that is, $\mathscr{F} = \{\emptyset, B, B^c, \Omega\} = \sigma(\{B\})$.

§06.13 **Reminder.** Let $(\Omega,\mathscr{A},\mathbb{P})$ be a probability space. According to Theorem §02.02, there exists a uniquely determined expectation $\mathbb{E} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})\to\overline{\mathbb{R}}_{\geqslant 0}$ with respect to $\mathbb{P}$, which is (I1) linear, (I2) monotone convergent and (I3) normed. □

§06.14 **Conditional expectation.** *Let $B\in\mathscr{A}$ with $\mathbb{P}(B)\in\mathbb{R}_{>0}$. Then*

$$\mathbb{P}(\bullet|B) : \mathscr{A}\to[0,1] \text{ with } A\mapsto \mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} \qquad (06.03)$$

is a probability measure on $(\Omega,\mathscr{A})$. Correspondingly, there exists a uniquely determined expectation with respect to $\mathbb{P}(\bullet|B)$, which we denote by $\mathbb{E}(\bullet|B)$. Due to (I3) *it satisfies*

$$\forall A\in\mathscr{A} \;:\; \mathbb{E}(\mathbb{1}_A|B) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{E}(\mathbb{1}_A\mathbb{1}_B)}{\mathbb{P}(B)} = \mathbb{P}(\mathbb{1}_A|B).$$

Since for every $X\in\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ it still holds that $X\mathbb{1}_B\in\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, the expectation $\mathbb{E}(\bullet|B)$ also satisfies

$$\forall X\in\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \;:\; \mathbb{E}(X|B) = \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)} = \mathbb{P}(X|B).$$

If $B^c = \Omega \backslash B \in \mathscr{A}$ with $\mathbb{P}(B^c) \in \mathbb{R}_{>0}$, then we can analogously consider the probability measure $\mathbb{P}(\bullet|B^c)$ on $(\Omega, \mathscr{A})$ and the corresponding expectation $\mathbb{E}(\bullet|B^c)$ with respect to $\mathbb{P}(\bullet|B^c)$. For every $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ is then

$$\mathbb{E}(X|\mathscr{F}) : \Omega \to \overline{\mathbb{R}}_{\geqslant 0} \text{ with } \omega \mapsto \mathbb{E}(X|\mathscr{F})(\omega) := \mathbb{E}(X|B)\mathbb{1}_B(\omega) + \mathbb{E}(X|B^c)\mathbb{1}_{B^c}(\omega)$$

measurable with respect to the sub-σ-algebra $\mathscr{F} := \sigma(\{B\}) = \{\emptyset, B, B^c, \Omega\} \subseteq \mathscr{A}$, that is, $\mathbb{E}(X|\mathscr{F})$ is a simple numerical random variable on $(\Omega, \mathscr{F})$, symbolically $\mathbb{E}(X|\mathscr{F}) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$. For every $\omega \in \Omega$, either $\omega \in B$ and $\mathbb{E}(X|\mathscr{F})(\omega) = \mathbb{E}(X|B)$ or $\omega \in B^c$ and $\mathbb{E}(X|\mathscr{F})(\omega) = \mathbb{E}(X|B^c)$, such that

$$\mathbb{E}(\bullet|\mathscr{F}) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}) \text{ with } X \mapsto \mathbb{E}(X|\mathscr{F}) := \mathbb{E}(X|B)\mathbb{1}_B + \mathbb{E}(X|B^c)\mathbb{1}_{B^c}. \quad (06.04)$$

We note that given $\omega \in B$, $\mathbb{E}(\bullet|B) = \mathbb{E}(\bullet|\mathscr{F})(\omega)$. For every $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, $\mathbb{E}(X|\mathscr{F})$ is particularly a solution of the Radon-Nikodym equation (simple verification!)

$$\forall F \in \mathscr{F} \ : \ \mathbb{E}(\mathbb{1}_F \mathbb{E}(X|\mathscr{F})) = \mathbb{E}(\mathbb{1}_F X). \quad (06.05)$$

Thus, the mapping $\mathbb{E}(\bullet|\mathscr{F})$ in (06.04) satisfies conditions (cE1) and (cE2), and it is therefore a version of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$. □

§06.15 **Regular version**. *We further note that the version of the conditional expectation $\mathbb{E}(\bullet|\mathscr{F})$ given in (06.04) implies by (06.01) a regular version $\mathbb{P}(\bullet|\mathscr{F})$ of the conditional distribution. Indeed, the map $\kappa := \mathbb{P}(\bullet|\mathscr{F}) : \Omega \times \mathscr{A} \to \overline{\mathbb{R}}_{\geqslant 0}$ given by*

$$(\omega, A) \mapsto \kappa(\omega, A) := \mathbb{P}(A|\mathscr{F})(\omega) = \mathbb{P}(A|B)(\mathbb{1})[B]\omega + \mathbb{P}(A|B^c)\mathbb{1}_{B^c}(\omega),$$

satisfies the conditions (Mk1) and (Mk2), and hence it is a Markov kernel from $(\Omega, \mathscr{F})$ to $(\Omega, \mathscr{A})$. Consider (Mk1). For $\omega \in \Omega$ either $\omega \in B$, and thus $\kappa_\omega = \mathbb{P}(\bullet|\mathscr{F})(\omega) = \mathbb{P}(\bullet|B) \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$, or $\omega \in B^c$ and $\kappa_\omega = \mathbb{P}(\bullet|\mathscr{F})(\omega) = \mathbb{P}(\bullet|B^c) \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$. Consequently, $\kappa_\bullet : \Omega \to \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ with $\omega \mapsto \kappa_\omega = \mathbb{P}(\bullet|\mathscr{F})(\omega)$, i.e. every ω-slice $\kappa_\omega = \mathbb{P}(\bullet|\mathscr{F})(\omega)$ is a probability measure on $(\Omega, \mathscr{A})$. On the other hand, condition (Mk2) is fulfilled since every A-slice satisfies $\kappa^A = \mathbb{P}(A|\mathscr{F}) = \mathbb{E}(\mathbb{1}_A|\mathscr{F}) \in \mathcal{M}_{[0,1]}(\mathscr{F})$, i.e. $\kappa^\bullet : \mathscr{A} \to \mathcal{M}_{[0,1]}(\mathscr{F})$ with $A \mapsto \kappa^A = \mathbb{P}(A|\mathscr{F})$. Moreover, by (06.05), $\kappa^A = \mathbb{P}(A|\mathscr{F}) = \mathbb{E}(\mathbb{1}_A|\mathscr{F})$ is solution to the Radon-Nikodym equation (cE2) for every $A \in \mathscr{A}$. Consequently, the implied version $\mathbb{P}(\bullet|\mathscr{F})$ of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$ given in (06.01) is regular. On the other hand, by construction, for every $\omega \in \Omega$ the mapping $\mathbb{E}(\bullet|\mathscr{F})(\omega)$ is the expectation with respect to the probability measure $\mathbb{P}(\bullet|\mathscr{F})(\omega)$ so that the version $\mathbb{E}(\bullet|\mathscr{F})$ in (06.04) is also regular. □

§06.16 **Conditional distribution given a $\mathbb{P}$-null set** B. *For $B \in \mathscr{A}$ with $\mathbb{P}(B) = 0$, $\mathbb{P}(\bullet|B)$ and hence also $\mathbb{E}(\bullet|\mathscr{F})$ as well as $\mathbb{P}(\bullet|\mathscr{F})$ are yet not defined. Here and subsequently, we agree on $\mathbb{P}(\bullet|B) := \mathbb{P}$ and hence $\mathbb{E}(\bullet|B) = \mathbb{E}$. On the other hand side, since $\mathbb{P}(B^c) = 1$ for all $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ we have*

$$\mathbb{E}(X|B^c) = \frac{\mathbb{E}(X\mathbb{1}_{B^c})}{\mathbb{P}(B^c)} = \mathbb{E}(X).$$

The mapping (06.04) becomes then

$$\mathbb{E}(\bullet|\mathscr{F}) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \ni X \mapsto \mathbb{E}(X|\mathscr{F}) := \mathbb{E}(X) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}) \quad (06.06)$$

satisfying for all $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ the conditions (cE1) and (cE2), that is, $\mathbb{E}(X|\mathscr{F}) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ and $\mathbb{E}(\mathbb{1}_F\mathbb{E}(X|\mathscr{F})) = \mathbb{E}(X)\mathbb{P}(F) = \mathbb{E}(\mathbb{1}_F X)$ for all $F \in \mathscr{F}$. Consequently, $\mathbb{E}(\bullet|\mathscr{F}) = \mathbb{E}$ as in

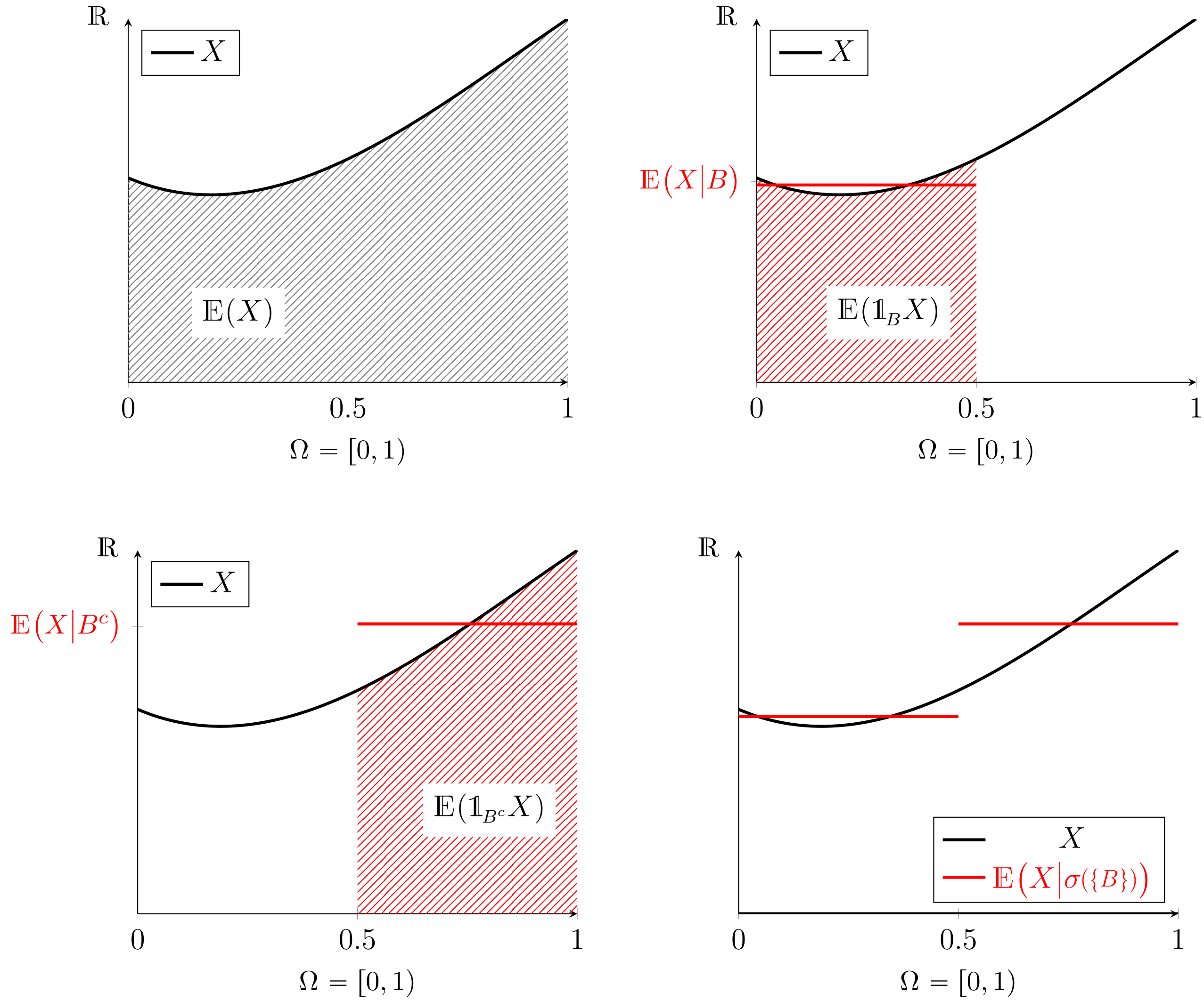
(06.06) *is a version of the conditional expectation with respect to* $\mathbb{P}$ *given* $\mathscr{F}$ *and the implied version of the conditional distribution given in* (06.01) *is regular. Indeed, the map*

$$\kappa = \mathbb{P}\big(\bullet\big|\mathscr{F}\big) : \Omega \times \mathscr{A} \ni (\omega, A) \mapsto \kappa(\omega, A) := \mathbb{P}\big(A\big|\mathscr{F}\big)(\omega) = \mathbb{P}(A) \in [0, 1],$$

is a Markov kernel from $(\Omega, \mathscr{F})$ *to* $(\Omega, \mathscr{A})$ *satisfying* (Mk1) *and* (Mk2)*. Moreover, every* A*-slice* $\kappa^A = \mathbb{P}(A)$ *is solution to the Radon-Nikodym equation* (cE2)*. Consequently,* $\mathbb{P}\big(\bullet\big|\mathscr{F}\big) = \mathbb{P}$ *is a regular version of the conditional distribution of* $\mathbb{P}$ *given* $\mathscr{F}$*. By construction, for every* $\omega \in \Omega$*, the map* $\mathbb{E}\big(\bullet\big|\mathscr{F}\big)(\omega) = \mathbb{E}$ *is the expectation with respect to the probability measure* $\mathbb{P}\big(\bullet\big|\mathscr{F}\big)(\omega) = \mathbb{P}$*, and hence the version* $\mathbb{E}\big(\bullet\big|\mathscr{F}\big)$ *given in* (06.06) *is regular.* □

§06.17 **Comment**. All statements made so far regarding $\mathbb{E}\big(\bullet\big|\mathscr{F}\big)$ and $\mathbb{P}\big(\bullet\big|\mathscr{F}\big)$ apply to any sub-σ-algebra of $\mathscr{A}$ of the form $\mathscr{F} = \sigma(\{B\})$ with $B \in \mathscr{A}$. Obviously, in the situation where $\mathbb{P}(B) = 0$ the choice $\mathbb{P}\big(\bullet\big|B\big) = \mathbb{P}$ is arbitrary. The statements hold equally if we set $\mathbb{P}\big(\bullet\big|B\big) := \tilde{\mathbb{P}}$ and $\mathbb{E}\big(\bullet\big|B\big) := \tilde{\mathbb{E}}$ for any $\tilde{\mathbb{P}} \in \mathfrak{M}_{\text{pr}}(\mathscr{A})$ and expectation $\tilde{\mathbb{E}}$ with respect to $\tilde{\mathbb{P}}$. It is important to note that in the situation where $\mathbb{P}(B) = 0$, a different choice only changes $\mathbb{E}\big(\bullet\big|\mathscr{F}\big)$ and $\mathbb{P}\big(\bullet\big|\mathscr{F}\big)$ on the null set B with respect to $\mathbb{P}$, so $\mathbb{E}\big(\bullet\big|\mathscr{F}\big)$ and $\mathbb{P}\big(\bullet\big|\mathscr{F}\big)$ are unique up to equality almost everywhere with respect to $\mathbb{P}$. □

§06.18 **Illustration**. Consider the probability space $([0, 1), \mathscr{B}_{[0,1)}, \mathrm{U}_G[0, 1))$. The following graphs illustrate a random variable $X \in \mathcal{M}(\mathscr{B}_{[0,1)})$ as well as the regular version of the conditional expectation of X given $\mathscr{F} = \sigma(\{B\})$ with $B = [0, 0.5) \in \mathscr{B}_{[0,1)}$.



§06|02 Countable conditions

We now consider a sub-σ-algebra $\mathscr{F} = \sigma(\mathcal{P})$ of $\mathscr{A}$, generated by a countable, measurable partition $\mathcal{P} = \{B_i : i \in \mathcal{I}\} \subseteq \mathscr{A}_{\neq\emptyset}$ of Ω.

§06.19 **Conditional expectation**. *Let $\mathcal{P} = \{B_i : i \in \mathcal{I}\} \subseteq \mathscr{A}_{\neq\emptyset}$ be a countable, measurable partition of Ω. For any $B \in \mathcal{P}$ with $\mathbb{P}(B) \in \mathbb{R}_{>0}$ we define $\mathbb{P}(\bullet|B)$ as in (06.03), and in case $\mathbb{P}(B) = 0$ we set as before $\mathbb{P}(\bullet|B) := \mathbb{P}$. Moreover, let $\mathbb{E}(\bullet|B)$ denote the expectation with respect to $\mathbb{P}(\bullet|B)$. For $\mathscr{F} = \sigma(\mathcal{P})$ the map*

$$\mathbb{E}(\bullet|\mathscr{F}) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \ni X \mapsto \mathbb{E}(X|\mathscr{F}) := \sum_{B\in\mathcal{P}} \mathbb{E}(X|B)\mathbb{1}_B \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}) \tag{06.07}$$

satisfies the conditions (cE1) and (cE2). The function $f : \Omega \to \mathcal{I}$ with $\omega \mapsto f(\omega) = i :\Leftrightarrow \omega \in B_i$ is $\mathscr{F}$-$2^{\mathcal{I}}$-measurable, and for every $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, the function $g : \mathcal{I} \to \overline{\mathbb{R}}_{\geqslant 0}$ with $i \mapsto \mathbb{E}(X|B_i)$ is $2^{\mathcal{I}}$-$\overline{\mathscr{B}}_{\geqslant 0}$-measurable. Consequently, $\mathbb{E}(X|\mathscr{F}) = g \circ f \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ is a $\mathscr{F}$-$\overline{\mathscr{B}}_{\geqslant 0}$-measurable function and (cE1) is fulfilled. Moreover, for each $F \in \mathscr{F}$ exists $\mathcal{J} \subseteq \mathcal{I}$ with $F = \biguplus_{j\in\mathcal{J}} B_j$ and

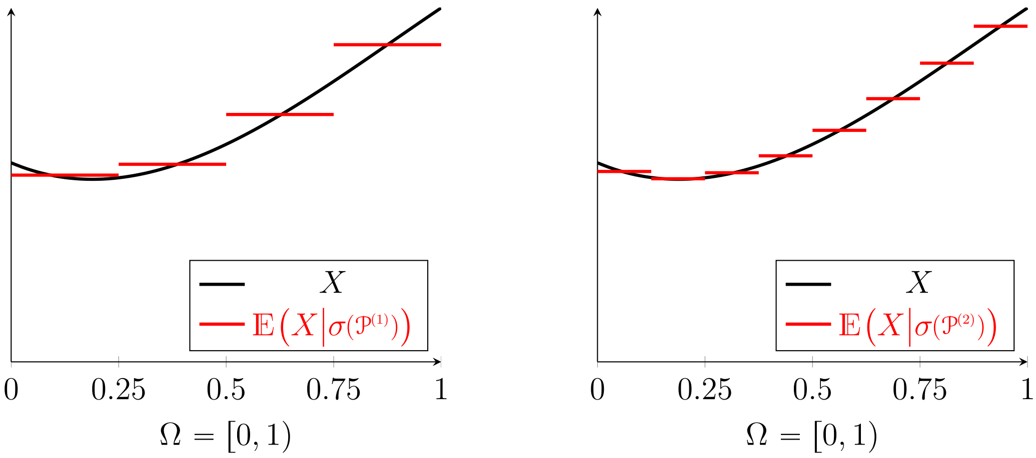
$$\begin{aligned}\mathbb{E}(\mathbb{1}_F\mathbb{E}(X|\mathscr{F})) &= \sum_{j\in\mathcal{J}} \mathbb{E}(\mathbb{1}_{B_j}\mathbb{E}(X|\mathscr{F})) = \sum_{j\in\mathcal{J}} \mathbb{E}(\mathbb{1}_{B_j}\mathbb{E}(X|B_j)) \\ &= \sum_{j\in\mathcal{J}} \mathbb{P}(B_j)\mathbb{E}(X|B_j) = \sum_{j\in\mathcal{J}} \mathbb{E}(\mathbb{1}_{B_j}X) = \mathbb{E}(\mathbb{1}_F X),\end{aligned} \tag{06.08}$$

hence, (cE2) is satisfied. Consequently, $\mathbb{E}(\bullet|\mathscr{F})$ as in (06.07) is a version of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$. Furthermore, the implied version in (06.01) of the conditional distribution is regular. Indeed, the map $\kappa = \mathbb{P}(\bullet|\mathscr{F}) : \Omega \times \mathscr{A} \to \overline{\mathbb{R}}_{\geqslant 0}$ given by

$$(\omega, A) \mapsto \kappa(\omega, A) := \mathbb{P}(A|\mathscr{F})(\omega) = \mathbb{E}(\mathbb{1}_A|\mathscr{F})(\omega) = \sum_{B\in\mathcal{P}} \mathbb{P}(A|B)\mathbb{1}_B(\omega)$$

satisfies the conditions (Mk1) and (Mk2), and hence it is a Markov kernel from $(\Omega, \mathscr{F})$ to $(\Omega, \mathscr{A})$. For each $\omega \in \Omega$ there exists a unique $B \in \mathcal{P}$ with $\omega \in B$ and the corresponding ω-slice satisfies $\kappa_\omega = \mathbb{P}(\bullet|\mathscr{F})(\omega) = \mathbb{P}(\bullet|B) \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$. Consequently, $\kappa_\bullet : \Omega \to \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ with $\omega \mapsto \kappa_\omega = \mathbb{P}(\bullet|\mathscr{F})(\omega)$, and (Mk1) is fulfilled. On the other hand, every A-slice satisfies $\kappa^A = \mathbb{P}(A|\mathscr{F}) = \mathbb{E}(\mathbb{1}_A|\mathscr{F}) \in \mathcal{M}_{[0,1]}(\mathscr{F})$, i.e. $\kappa^\bullet : \mathscr{A} \to \mathcal{M}_{[0,1]}(\mathscr{F})$ with $A \mapsto \kappa^A = \mathbb{P}(A|\mathscr{F})$, and (Mk2) is fulfilled. Moreover, by (06.08), $\kappa^A = \mathbb{P}(A|\mathscr{F}) = \mathbb{E}(\mathbb{1}_A|\mathscr{F})$ is solution to the Radon-Nikodym equation (cE2) for every $A \in \mathscr{A}$. Consequently, $\mathbb{P}(\bullet|\mathscr{F})$ is a regular version of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$. By construction, for every $\omega \in \Omega$, the map $\mathbb{E}(\bullet|\mathscr{F})(\omega)$ is the expectation with respect to the probability measure $\mathbb{P}(\bullet|\mathscr{F})(\omega)$, and hence the version $\mathbb{E}(\bullet|\mathscr{F})$ given in (06.08) is regular. □

§06.20 **Illustration**. Consider as in Illustration §06.18 the probability space $([0,1), \mathscr{B}_{[0,1)}, \mathrm{U}_G[0,1))$. The following graphs illustrate a random variable $X \in \mathcal{M}(\mathscr{B}_{[0,1)})$ as well as the regular version of the conditional expectation of X given $\mathscr{F} = \sigma(\mathcal{P}^{(j)})$, $j \in [\![2]\!]$, for the two measurable partitions $\mathcal{P}^{(1)} = \{[\frac{i-1}{4}, \frac{i}{4}) : i \in [\![4]\!]\}$ and $\mathcal{P}^{(2)} = \{[\frac{i-1}{8}, \frac{i}{8}) : i \in [\![8]\!]\}$ of $\Omega = [0,1)$.



□

§06|03 Induced conditional distribution

§06.21 **Reminder**. If $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$ is a random variable on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{X}, \mathscr{X})$, then the image probability measure $\mathbb{P}^X := \mathbb{P} \circ X^{-1} \in \mathfrak{M}_{\text{pr}}(\mathscr{X})$ is also called the distribution on $(\mathcal{X}, \mathscr{X})$ induced by X. □

§06.22 **Definition**. Let $X \in \mathcal{M}(\mathscr{A}, \mathscr{X})$ be a random variable on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{X}, \mathscr{X})$ and let $\mathscr{F} \subseteq \mathscr{A}$ be a sub-σ-algebra of $\mathscr{A}$.

(a) For each version $\mathbb{P}(\bullet|\mathscr{F})$ of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$, the map

$$\boxed{\mathbb{P}^X(\bullet|\mathscr{F})} : \Omega \times \mathscr{X} \ni (\omega, B) \mapsto \mathbb{P}^X(B|\mathscr{F})(\omega) := \mathbb{P}\big(X^{-1}(B)\big|\mathscr{F}\big)(\omega) \in \overline{\mathbb{R}}_{\geqslant 0}$$

induced by X is called *version of the conditional distribution of $\mathbb{P}^X$ given $\mathscr{F}$*. By (cE2), for each $B \in \mathscr{X}$ every version $\mathbb{P}^X(B|\mathscr{F})$ is solution to the Radon-Nikodym equation

$$\mathbb{P}(X^{-1}(B) \cap F) = \mathbb{P}(\mathbb{P}(X^{-1}(B)|\mathscr{F})\mathbb{1}_F) = \mathbb{P}(\mathbb{P}^X(B|\mathscr{F})\mathbb{1}_F) \text{ for all } F \in \mathscr{F}.$$

If in addition $\mathbb{P}^X(\bullet|\mathscr{F})$ is a Markov kernel from $(\Omega, \mathscr{F})$ to $(\mathcal{X}, \mathscr{X})$, then the version $\mathbb{P}^X(\bullet|\mathscr{F})$ is called *regular*.

(b) Given any version $\mathbb{E}(\bullet|\mathscr{F})$ of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$, for each $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X})$ the random variable $\mathbb{E}^X(h|\mathscr{F}) := \mathbb{E}\big(h(X)\big|\mathscr{F}\big) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ is called *version of the conditional expectation of $h(X)$ given $\mathscr{F}$* and the mapping induced by X

$$\boxed{\mathbb{E}^X(\bullet|\mathscr{F})} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X}) \ni h \mapsto \mathbb{E}^X(h|\mathscr{F}) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$$

is called the *conditional expectation with respect to $\mathbb{P}^X$ given $\mathscr{F}$*. By (cE2), every version $\mathbb{E}^X(h|\mathscr{F})$ is solution to the Radon-Nikodym equation

$$\mathbb{E}(h(X)\mathbb{1}_F) = \mathbb{E}(\mathbb{E}(h(X)|\mathscr{F})\mathbb{1}_F) = \mathbb{E}(\mathbb{E}^X(h|\mathscr{F})\mathbb{1}_F) \text{ for all } F \in \mathscr{F}.$$

The version $\mathbb{E}(\bullet|\mathscr{F})$ is called *regular*, if the implied version

$$\mathbb{P}^X(\bullet|\mathscr{F}) : \Omega \times \mathscr{X} \ni (\omega, B) \mapsto \mathbb{P}^X(B|\mathscr{F})(\omega) := \mathbb{E}^X(\mathbb{1}_B|\mathscr{F})(\omega) \in \overline{\mathbb{R}}_{\geqslant 0} \qquad (06.09)$$

of the conditional distribution of $\mathbb{P}^X$ given $\mathscr{F}$ is regular, and for each $\omega \in \Omega$, the map $\mathbb{E}^X(\bullet|\mathscr{F})(\omega) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X}) \to \overline{\mathbb{R}}_{\geqslant 0}$ is the expectation with respect to $\mathbb{P}^X(\bullet|\mathscr{F})(\omega)$. □

§06.23 **Remark**. The map $\mathbb{P}^X(\bullet|\mathscr{F})$ satisfies the condition (Mk2) for any version $\mathbb{P}(\bullet|\mathscr{F})$ by definition (cE1). In addition let $\mathbb{P}(\bullet|\mathscr{F})$ be a regular version, then every ω-slice $\mathbb{P}(\bullet|\mathscr{F})(\omega) \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ is a probability measure on $(\Omega, \mathscr{A})$ and by definition $\mathbb{P}^X(\bullet|\mathscr{F})(\omega) = \mathbb{P}(\bullet|\mathscr{F})(\omega) \circ X^{-1} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{X})$ is the image probability measure of $\mathbb{P}(\bullet|\mathscr{F})(\omega)$ under X. Therefore, $\mathbb{P}^X(\bullet|\mathscr{F})$ satisfies also (Mk1), and hence, it is a Markov kernel. Consequently, the version $\mathbb{P}^X(\bullet|\mathscr{F})$ is regular, since every A-slice $\mathbb{P}(A|\mathscr{F}) = \mathbb{E}(\mathbb{1}_A|\mathscr{F})$ is also solution to the Radon-Nikodym equation (cE2). For any regular version $\mathbb{E}(\bullet|\mathscr{F})$ the induced version $\mathbb{P}(\bullet|\mathscr{F})$, and hence $\mathbb{P}^X(\bullet|\mathscr{F})$, is also regular. For each $\omega \in \Omega$ by the transformation formula (see Lemma §02.27) the induced version $\mathbb{E}^X(\bullet|\mathscr{F})(\omega)$ is the expectation with respect to the image probability measure $\mathbb{P}^X(\bullet|\mathscr{F})(\omega)$. □

§06|04 Conditional expectation given a random variable

§06.24 **Reminder**. Let $\mathscr{F} = \sigma(S)$ be the sub-σ-algebra in $\mathscr{A}$ generated by a random variable S defined on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{S}, \mathscr{S})$, i.e., $S \in \mathcal{M}(\mathscr{A}, \mathscr{S})$. We denote as before by $\mathbb{P}^S$ the image probability measure of $\mathbb{P}$ under S, and by $\mathbb{E}^S$ the corresponding expectation. If $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\sigma(S))$ is a positive numerical, $\sigma(S)$-measurable, random variable, then there exists $\varphi \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$ with $Y = \varphi \circ S$, that is, $Y(\omega) = \varphi(S(\omega))$ for all $\omega \in \Omega$. The function φ is uniquely determined by Y and S only on $S(\Omega) \subseteq \mathcal{S}$. □

§06.25 **Definition**. For $S : (\Omega, \mathscr{A}, \mathbb{P}) \to (\mathcal{S}, \mathscr{S})$ let $\mathbb{E}(X|\sigma(S)) \in \overline{\mathcal{M}}_{\geqslant 0}(\sigma(S))$ be a version of the conditional expectation of $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ given $\sigma(S)$ and let $\varphi \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$ fulfil $\mathbb{E}(X|\sigma(S)) = \varphi(S)$ as in §06.24. $\boxed{\mathbb{E}^{|S}(X)} := \varphi$ is called *version of the conditional expectation of X given S*. By (cE2), every version $\mathbb{E}^{|S}(X)$ is solution to the Radon-Nikodym equation

$$\mathbb{E}(\mathbb{1}_{S^{-1}(B)}X) = \mathbb{E}(\mathbb{1}_{S^{-1}(B)}\mathbb{E}(X|\sigma(S))) = \mathbb{E}^S(\mathbb{1}_B \mathbb{E}^{|S}(X)) \text{ for all } B \in \mathscr{S}.$$

Any map

$$\boxed{\mathbb{E}^{|S}} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S}) \text{ with } X \mapsto \mathbb{E}^{|S}(X)$$

is called *version of the conditional expectation with respect to $\mathbb{P}$ given S*. The induced mapping

$$\boxed{\mathbb{P}^{|S}} : \mathcal{S} \times \mathscr{A} \to \overline{\mathbb{R}}_{\geqslant 0} \text{ with } (s, A) \mapsto \mathbb{P}^{|S}(A)(s) := \mathbb{E}^{|S}(\mathbb{1}_A)(s)$$

is called *version of the conditional distribution of $\mathbb{P}$ given S*. By (cE2), for each $A \in \mathscr{A}$ every version $\mathbb{P}^{|S}(A) = \mathbb{E}^{|S}(\mathbb{1}_A)$ is solution to the Radon-Nikodym equation

$$\mathbb{P}(S^{-1}(B) \cap A) = \mathbb{P}^S(\mathbb{1}_B \mathbb{P}^{|S}(A)) = \int_B \mathbb{P}^{|S}(A) \mathrm{d}\mathbb{P}^S \text{ for all } B \in \mathscr{S}.$$

If in addition $\mathbb{P}^{|S}$ is a Markov kernel from $(\mathcal{S}, \mathscr{S})$ to $(\Omega, \mathscr{A})$, then the version $\mathbb{P}^{|S}$ is called *regular*. Similarly, the version $\mathbb{E}^{|S}$ is called *regular*, if the induced version $\mathbb{P}^{|S}$ is regular, and for each $s \in \mathcal{S}$, the map $\mathbb{E}^{|S}(\bullet)(s) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathbb{R}}_{\geqslant 0}$ is the expectation with respect to $\mathbb{P}^{|S}(\bullet)(s) \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$. □

§06.26 **Notation**. Let $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ and $A \in \mathscr{A}$. For any version of the conditional expectation, respectively, distribution with respect to $\mathbb{P}$ given S we write also shortly $\boxed{\mathbb{E}^{|S=s}(X)} := \mathbb{E}^{|S}(X)(s)$ and $\boxed{\mathbb{P}^{|S=s}(A)} := \mathbb{P}^{|S}(A)(s)$ for $s \in \mathcal{S}$. A s-slice of $\mathbb{P}^{|S}$ is thus denoted by $\mathbb{P}^{|S=s} : \mathscr{A} \to \overline{\mathbb{R}}_{\geqslant 0}$ with $A \mapsto \mathbb{P}^{|S=s}(A)$ while $\mathbb{P}^{|S}(A)$ denotes an A-slice of $\mathbb{P}^{|S}$. □

§06.27 **Property**. *For $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ let $\varphi_1(S)$ and $\varphi_2(S)$ with $\varphi_1, \varphi_2 \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$ be two versions of the conditional expectation $\mathbb{E}\big(X\big|\sigma(S)\big)$ of X given $\sigma(S)$. By Theorem §06.02, we have $\varphi_1(S) = \varphi_2(S)$ $\mathbb{P}$-a.e. and hence $\varphi_1 = \varphi_2$ $\mathbb{P}^S$-a.e. In opposite to $\mathbb{E}\big(X\big|\sigma(S)\big)$, any version φ does not only depend on $\sigma(S)$ but also on the specific choice of the random variable S.* □

§06.28 **Remark**. Let $\mathbb{P}\big(\,\bullet\,\big|\sigma(S)\big)$ be a version of the conditional distribution of $\mathbb{P}$ given $\sigma(S)$ and let $\mathbb{P}^{|S}$ be an induced version of the conditional distribution of $\mathbb{P}$ given S. If $\mathbb{P}^{|S}$ is regular, i.e. $\kappa := \mathbb{P}^{|S}$ is a Markov kernel from $(\mathcal{S}, \mathscr{S})$ to $(\Omega, \mathscr{A})$. Then the map

$$(\omega, A) \mapsto \kappa(S(\omega), A) = \mathbb{P}^{|S=S(\omega)}(A) = \mathbb{P}\big(A\big|\sigma(S)\big)(\omega)$$

is a Markov kernel from $(\Omega, \sigma(S))$ to $(\Omega, \mathscr{A})$, and hence $\mathbb{P}\big(\,\bullet\,\big|\sigma(S)\big)$ is regular. Evidently, (Mk1) and (Mk2) are satisfied, since $\kappa_{S(\omega)} = \mathbb{P}^{|S=S(\omega)} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ for all $\omega \in \Omega$ and the composition of measurable functions is measurable. If S is surjective, then the converse also holds, because in this case $\mathbb{P}\big(\,\bullet\,\big|\sigma(S)\big)$ determines uniquely $\mathbb{P}^{|S}$ (compare Reminder §06.24). □

§06.29 **Example**. Let $S : (\Omega, \mathscr{A}, \mathbb{P}) \to (\{0,1\}, 2^{\{0,1\}})$ be a simple random variable, that is, $\{S = 1\} = S^{-1}(\{1\}) \in \mathscr{A}$. Since $\sigma(S) = \sigma(\{S^{-1}(\{1\})\})$ we follow Subsection §06|01 and a version of the conditional expectation with respect to $\mathbb{P}$ given $\sigma(S)$ is given by

$$\mathbb{E}\big(\,\bullet\,\big|\sigma(S)\big) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \to \overline{\mathcal{M}}_{\geqslant 0}(\sigma(S)) \text{ with}$$
$$X \mapsto \mathbb{E}\big(X\big|\sigma(S)\big) = \mathbb{E}\big(X\big|S^{-1}(\{1\})\big)\mathbb{1}_{S^{-1}(\{1\})} + \mathbb{E}\big(X\big|S^{-1}(\{0\})\big)\mathbb{1}_{S^{-1}(\{0\})} = \varphi(S)$$

and $\varphi \in \overline{\mathcal{M}}_{\geqslant 0}(2^{\{0,1\}})$ with

$$s \mapsto \varphi(s) = \mathbb{E}\big(X\big|S^{-1}(\{1\})\big)\mathbb{1}_{\{1\}}(s) + \mathbb{E}\big(X\big|S^{-1}(\{0\})\big)\mathbb{1}_{\{0\}}(s).$$

Consequently, $\mathbb{E}^{|S} = \mathbb{E}\big(\,\bullet\,\big|S^{-1}(\{1\})\big)\mathbb{1}_{\{1\}} + \mathbb{E}\big(\,\bullet\,\big|S^{-1}(\{0\})\big)\mathbb{1}_{\{0\}}$ is a *version of the conditional expectation with respect to $\mathbb{P}$ given S* where in particular $\mathbb{E}^{|S=1} = \mathbb{E}\big(\,\bullet\,\big|S^{-1}(\{1\})\big)$ and $\mathbb{E}^{|S=0} = \mathbb{E}\big(\,\bullet\,\big|S^{-1}(\{0\})\big)$. □

§06|05 Induced conditional distribution given a random variable

§06.30 **Reminder**. Let (S, X) be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in the product space $\mathcal{S} \times \mathcal{X}$ equipped with the product σ-algebra $\mathscr{S} \otimes \mathscr{X}$. We denote by $\mathbb{P}^{S,X}$ the joint distribution on $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$ induced by (S, X), and by $\Pi_{\mathcal{X}} : \mathcal{S} \times \mathcal{X} \to \mathcal{X}$ and $\Pi_{\mathcal{S}} : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$ the coordinate maps, where $(s, x) \mapsto \Pi_{\mathcal{X}}(s, x) := x$ and $(s, x) \mapsto \Pi_{\mathcal{S}}(s, x) := s$. The marginal distribution of X and S satisfies $\mathbb{P}^X = \mathbb{P} \circ X^{-1} = \mathbb{P} \circ (\Pi_{\mathcal{X}}(S, X))^{-1} = \mathbb{P}^{S,X} \circ \Pi_{\mathcal{X}}^{-1}$ and $\mathbb{P}^S = \mathbb{P}^{X,S} \circ \Pi_{\mathcal{S}}^{-1}$, respectively. The random variables $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{S}}$ are evidently surjective. Our goal is to introduce a conditional distribution of X given S, using Definitions §06.22 and §06.25. Consider first a version $\mathbb{P}\big(\,\bullet\,\big|\sigma(S)\big) : \Omega \times \mathscr{A} \to \overline{\mathbb{R}}_{\geqslant 0}$ of the conditional distribution of $\mathbb{P}$ given $\sigma(S)$ implying by Definition §06.22 a version $\mathbb{P}^X\big(\,\bullet\,\big|\sigma(S)\big) : \Omega \times \mathscr{X} \to \overline{\mathbb{R}}_{\geqslant 0}$ of the conditional distribution of $\mathbb{P}^X$ given $\sigma(S)$ induced by X, which in turn induces by Definition §06.25 a version $\mathbb{P}^{X|S} : \mathcal{S} \times \mathscr{X} \to \overline{\mathbb{R}}_{\geqslant 0}$ of the conditional distribution of $\mathbb{P}^X$ given S. Note that $\mathbb{P}^X\big(\,\bullet\,\big|\sigma(S)\big)$ is regular if $\mathbb{P}\big(\,\bullet\,\big|\sigma(S)\big)$ is regular (see Remark §06.28), but in general regularity of $\mathbb{P}^{X|S}$ is not implied. However, if $\mathbb{P}^{(S,X)}\big(\bullet\big|\sigma(\Pi_{\mathcal{S}})\big)$ is a regular version on $(\Omega := \mathcal{S} \times \mathcal{X}, \mathscr{A} := \mathscr{S} \otimes \mathscr{X})$ of the conditional distribution of $\mathbb{P}^{(S,X)}$ given $\sigma(\Pi_{\mathcal{S}})$, then by Definition §06.22 the version $\mathbb{P}^X\big(\,\bullet\,\big|\sigma(\Pi_{\mathcal{S}})\big)$ of the conditional distribution of $\mathbb{P}^X$ given $\sigma(\Pi_{\mathcal{S}})$ induced by $X := \Pi_{\mathcal{X}}$ is regular, and by Definition §06.25 the version $\mathbb{P}^{X|S}$ of the conditional distribution of $\mathbb{P}^X$ given $S := \Pi_{\mathcal{S}}$ is also regular since $\Pi_{\mathcal{S}}$ is surjective. □

§06.31 **Definition**. Let $\mathbb{P}^{S,X}$ be the joint distribution of the coordinate maps $(S := \Pi_{\mathcal{S}}, X := \Pi_{\mathcal{X}})$ on $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$.

(a) For each version $\mathbb{P}^{S,X}\big(\bullet\,\big|\sigma(\Pi_s)\big)$ of the conditional distribution of $\mathbb{P}^{S,X}$ given $\sigma(\Pi_s)$, the map

$$\boxed{\mathbb{P}^{X|S}} : \mathcal{S} \times \mathscr{X} \to \overline{\mathbb{R}}_{\geqslant 0} \text{ with } (s, B) \mapsto \mathbb{P}^{X|S}(s, B) := \varphi(s) \text{ and}$$
$$\mathbb{P}^{X}\big(B\big|\sigma(\Pi_s)\big) = \mathbb{P}^{S,X}\big(\Pi_{\mathcal{X}}^{-1}(B)\big|\sigma(\Pi_s)\big) = \varphi(\Pi_s)$$

is called *version of the conditional distribution of $\mathbb{P}^X$ given S*. For $s \in \mathcal{S}$ and $B \in \mathscr{X}$ we denote by $\mathbb{P}^{X|S=s}$ and $\mathbb{P}^{X|S}(B)$ the s-slice and B-slice of $\mathbb{P}^{X|S}$, respectively, where every B-slice satisfies $\mathbb{P}^{X|S}(B) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$ due to (cE1). Moreover, by (cE2), $\mathbb{P}^{X|S}(B)$ is solution to the Radon-Nikodym equation

$$\mathbb{P}^{S,X}(A \times B) = \mathbb{P}^{S}(\mathbb{1}_A\mathbb{P}^{X|S}(B)) = \int_A \mathbb{P}^{X|S}(B)\,\mathrm{d}\mathbb{P}^{S} \text{ for all } A \in \mathscr{S}.$$

If in addition $\mathbb{P}^{X|S}$ is a Markov kernel from $(\mathcal{S}, \mathscr{S})$ to $(\mathcal{X}, \mathscr{X})$, then the version $\mathbb{P}^{X|S}$ is called *regular* and $\mathbb{P}^{S} \odot \mathbb{P}^{X|S} = \mathbb{P}^{S,X}$ (see Notation §06.07). If every s-slice $\mathbb{P}^{X|S=s} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{X})$ of a regular version $\mathbb{P}^{X|S}$ has a finite first absolute moment, i.e. $\mathbb{P}^{X|S=s} \in \mathfrak{M}^1_{\mathrm{pr}}(\mathscr{X})$, then we write

$$\boxed{\mathbb{P}^{|S=s}(X)} := \mathbb{P}^{X|S=s}(\mathrm{id}_{\mathcal{X}}) = \int_{\mathcal{X}} x\mathbb{P}^{X|S=s}(\mathrm{d}x).$$

(b) For each version $\mathbb{E}^{S,X}\big(\bullet\,\big|\sigma(\Pi_s)\big)$ of the conditional expectation with respect to $\mathbb{P}^{S,X}$ given $\sigma(\Pi_s)$, the map

$$\boxed{\mathbb{E}^{X|S}} : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X}) \to \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S}) \text{ with } h \mapsto \mathbb{E}^{X|S}(h) := \varphi \text{ and}$$
$$\mathbb{E}^{X}\big(h\big|\sigma(\Pi_s)\big) = \mathbb{E}^{S,X}\big(h(\Pi_{\mathcal{X}})\big|\sigma(\Pi_s)\big) = \varphi(\Pi_s)$$

is called *version of the conditional expectation with respect to $\mathbb{P}^X$ given S*. By (cE2), for all $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X})$ any version $\mathbb{E}^{X|S}(h) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{S})$ is solution to the Radon-Nikodym equation

$$\mathbb{E}^{S,X}\big(h(\Pi_{\mathcal{X}})\mathbb{1}_{\Pi_s^{-1}(B)}\big) = \mathbb{E}^{S,X}\big(\mathbb{1}_{\Pi_s^{-1}(B)}\mathbb{E}^{S,X}\big(h(\Pi_{\mathcal{X}})\big|\sigma(\Pi_s)\big)\big)$$
$$= \mathbb{E}^{S}(\mathbb{1}_B\mathbb{E}^{X|S}(h)) \text{ for all } B \in \mathscr{S}.$$

The version $\mathbb{E}^{X|S}$ is called *regular*, if the implied version

$$\mathbb{P}^{X|S} : \mathcal{S} \times \mathscr{X} \to \overline{\mathbb{R}}_{\geqslant 0} \text{ with } (s, B) \mapsto \mathbb{P}^{X|S}(s, B) := \mathbb{E}^{X|S}(\mathbb{1}_B)(s)$$

of the conditional distribution of $\mathbb{P}^X$ given S is *regular*, and for each $s \in \mathcal{S}$, the map $\mathbb{E}^{X|S=s} := \mathbb{E}^{X|S}(\bullet)(s) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{X}) \to \overline{\mathbb{R}}_{\geqslant 0}$ is the expectation with respect to $\mathbb{P}^{X|S=s} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{X})$. □

§06|06 Properties

§06.32 **Preliminary**. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space, $\mathscr{F} \subseteq \mathscr{A}$ a sub-σ-algebra and $X \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$. A random variable $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ is solution to the Radon-Nikodym equation

(cE2) $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ for every $F \in \mathscr{F}$

if and only if $\mathbb{E}(ZY) = \mathbb{E}(ZX)$ for all $Z \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$. Moreover, for $\tilde{X} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ with $X = \tilde{X}$ $\mathbb{P}$-a.e., $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ satisfies (cE2) if and only if $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F \tilde{X})$ for every $F \in \mathscr{F}$, and hence $\mathbb{E}\big(X\big|\mathscr{F}\big) = \mathbb{E}\big(\tilde{X}\big|\mathscr{F}\big)$ $\mathbb{P}$-a.e.. Let $\Omega \in \mathscr{E} \subseteq \mathscr{F}$ be $\cap$-closed generator of $\mathscr{F} = \sigma(\mathscr{E})$ and $\mathbb{E}(X) \in \mathbb{R}_{\geqslant 0}$. $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ satisfies (cE2) if and only if $\mathbb{E}(\mathbb{1}_E Y) = \mathbb{E}(\mathbb{1}_E X)$ for every $E \in \mathscr{E}$. By Lemma §01.28 (note that $\Omega \in \mathscr{E}$ and $X\mathbb{P}(\Omega) = \mathbb{E}(X) \in \mathbb{R}_{\geqslant 0}$) the finite measures $X\mathbb{P}$ and $Y\mathbb{P}$ agree also on $\mathscr{A}$, i.e., (cE2) is satisfied, if they agree on $\mathscr{E}$. □

§06.33 **Proposition**. *Let $\mathscr{G} \subseteq \mathscr{F} \subseteq \mathscr{A}$ be sub-σ-algebras, $X, Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ and let $\mathbb{E}(\bullet|\mathscr{F})$ and $\mathbb{E}(\bullet|\mathscr{G})$ be versions of the conditional expectation with respect to $\mathbb{P} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ given $\mathscr{F}$ and $\mathscr{G}$, respectively. Then:*

(i) $\mathbb{E}(aX + bY|\mathscr{F}) = a\mathbb{E}(X|\mathscr{F}) + b\mathbb{E}(Y|\mathscr{F})$ $\mathbb{P}$-a.e. *for all* $a, b \in \mathbb{R}_{\geqslant 0}$; (linear) End LE10 2026/05/20

(ii) $\mathbb{E}(X|\mathscr{F}) \leqslant \mathbb{E}(Y|\mathscr{F})$ $\mathbb{P}$-a.e. *if* $X \leqslant Y$ $\mathbb{P}$-a.e.; (monotone)

(iii) $\sup_{n\in\mathbb{N}} \mathbb{E}(X_n|\mathscr{F}) = \mathbb{E}(X|\mathscr{F})$ $\mathbb{P}$-a.e. *for all* $X_n \uparrow X$ *in* $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$; (monotone convergence)

(iv) $\mathbb{E}(\liminf_{n\to\infty} X_n|\mathscr{F}) \leqslant \liminf_{n\to\infty} \mathbb{E}(X_n|\mathscr{F})$ $\mathbb{P}$-a.e. *for all* $(X_n)_{n\in\mathbb{N}}$ *in* $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$; (Fatou)

(v) *for all* $Z \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ *we have* $\mathbb{E}(ZX|\mathscr{F}) = Z\mathbb{E}(X|\mathscr{F})$ $\mathbb{P}$-a.e. *and* $\mathbb{E}(Z|\mathscr{F}) = Z$ $\mathbb{P}$-a.e., *in particular* $\mathbb{E}(1|\mathscr{F}) = 1$ $\mathbb{P}$-a.e.

(vi) $\mathbb{E}(\mathbb{E}(X|\mathscr{F})|\mathscr{G}) = \mathbb{E}(\mathbb{E}(X|\mathscr{G})|\mathscr{F}) = \mathbb{E}(X|\mathscr{G})$ $\mathbb{P}$-a.e.; (tower property)

(vii) $\mathbb{E}(X|\mathscr{F}) = \mathbb{E}(X)$ $\mathbb{P}$-a.e. *if* $\sigma(X)$ *and* $\mathscr{F}$ *are independent;* (independence)

(viii) $\mathbb{E}(X|\mathscr{T}^{\mathbb{P}}) = \mathbb{E}(X)$ $\mathbb{P}$-a.e. *for* $\mathscr{T}^{\mathbb{P}} := \{A \in \mathscr{A} : \mathbb{P}(A) \in \{0,1\}\}$ *and hence* $\mathbb{E}(X|\mathscr{T}) = \mathbb{E}(X)$ $\mathbb{P}$-a.e. *for* $\mathscr{T} = \{\varnothing, \Omega\}$;

(ix) $\mathbb{E}(\mathbb{E}(X|\mathscr{F})) = \mathbb{E}(X)$ (total probability)

§06.34 **Proof** of Proposition §06.33. is given in the lecture. □

§06.35 **Theorem**. *Let $\mathbb{P}_0, \mathbb{P}_1 \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ be probability measures with $\mathbb{P}_0 \ll \mathbb{P}_1$, $S \in \mathcal{M}(\mathscr{A}, \mathscr{S})$ and let $\mathbb{E}_0^{|S}$ and $\mathbb{E}_1^{|S}$ be versions of the conditional expectation with respect to $\mathbb{P}_0$ and $\mathbb{P}_1$ given S. Then:*

(i) $\mathbb{P}_0^S \ll \mathbb{P}_1^S$;

(ii) $\frac{d\mathbb{P}_0^S}{d\mathbb{P}_1^S} = \mathbb{E}_1^{|S}(\frac{d\mathbb{P}_0}{d\mathbb{P}_1})$ $\mathbb{P}_1^S$-a.e.;

(iii) *If $h \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ is $\mathbb{P}_0$-quasiintegrable, then*

$$\mathbb{E}_1^{|S}(h\tfrac{d\mathbb{P}_0}{d\mathbb{P}_1}) = \mathbb{E}_0^{|S}(h)\frac{d\mathbb{P}_0^S}{d\mathbb{P}_1^S}\ \mathbb{P}_1^S\text{-a.e.}\quad \text{and}\quad \mathbb{E}_0^{|S}(h) = \frac{\mathbb{E}_1^{|S}(h\frac{d\mathbb{P}_0}{d\mathbb{P}_1})}{\mathbb{E}_1^{|S}(\frac{d\mathbb{P}_0}{d\mathbb{P}_1})}\ \mathbb{P}_0^S\text{-a.e.}$$

Let $\widetilde{\mathbb{P}}_0 \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ be another probability measure with $\widetilde{\mathbb{P}}_0 \ll \mathbb{P}_1$, let L and L^S be respectively a likelihood ratio (LR) of $\widetilde{\mathbb{P}}_0$ with respect to $\mathbb{P}_0$ and $\widetilde{\mathbb{P}}_0^S$ with respect to $\mathbb{P}_0^S$, where $\{\mathrm{L} = \infty\}$ and $\{\mathrm{L}^S = \infty\}$ is respectively the singular area of $\widetilde{\mathbb{P}}_0$ with respect to $\mathbb{P}_0$ and $\widetilde{\mathbb{P}}_0^S$ with respect to $\mathbb{P}_0^S$. Then setting $\varphi := \widetilde{\mathbb{P}}_0^{|S}(\mathrm{L} \in \mathbb{R}_{\geqslant 0})$ it holds that

(iv) $\widetilde{\mathbb{P}}_0(\mathrm{L} = \infty) = \widetilde{\mathbb{P}}_0^S(\mathrm{L}^S = \infty) + \mathbb{P}_0(\mathrm{L}^S(S)(1 - \varphi(S)))$;

(v) $\mathbb{E}_0^{|S}(\mathrm{L}) = \mathrm{L}^S\varphi$ $\mathbb{P}_0^S$-a.e.

§06.36 **Proof** of Theorem §06.35. (i)-(iii) is given in the lecture, for (iv) and (v) we refer to Witting (1985, Satz 1.121, S.121) □

§07 Integrable random variables

§07.01 **Reminder**. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space, $\mathscr{F} \subseteq \mathscr{A}$ a sub-σ-algebra and $X \in \overline{\mathcal{M}}(\mathscr{A})$ a numerical random variable. Making use of the decomposition $X = X^+ - X^-$ with $X^+ = X \vee 0, X^- := (-X) \vee 0 \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ the expectation $\mathbb{E}(X) := \mathbb{P}(X) = \mathbb{P}(X^+) - \mathbb{P}(X^-)$ is defined for any X with $\mathbb{P}(X^+) \wedge \mathbb{P}(X^-) \in \mathbb{R}_{\geqslant 0}$. In the sequel let $\overline{\mathcal{L}}^1(\mathscr{A}, \mathbb{P}) := \overline{\mathcal{L}}^1(\Omega, \mathscr{A}, \mathbb{P}) = \{X \in \overline{\mathcal{M}}(\mathscr{A}) : \mathbb{P}(|X|) \in \mathbb{R}_{\geqslant 0}\}$ and we denote by $\mathbb{E} : \overline{\mathcal{L}}^1(\mathscr{A}, \mathbb{P}) \to \mathbb{R}$ the unique expectation with respect to $\mathbb{P}$. Since $\overline{\mathcal{M}}(\mathscr{F}) \subseteq \overline{\mathcal{M}}(\mathscr{A})$ we have $\overline{\mathcal{L}}^1(\mathscr{F}, \mathbb{P}) \subseteq \overline{\mathcal{L}}^1(\mathscr{A}, \mathbb{P})$. If $X \in \overline{\mathcal{L}}^1(\mathscr{A}, \mathbb{P})$ then $\mathbb{E}(X^+) \in \mathbb{R}_{\geqslant 0}$

and any version of the conditional expectation of X^+ given $\mathscr{F}$ satisfies $\mathbb{E}(X^+|\mathscr{F}) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ due to (cE1) and by (cE2), any version $\mathbb{E}(X^+|\mathscr{F})$ is solution to the Radon-Nikodym equation $\mathbb{E}(\mathbb{1}_F\mathbb{E}(X^+|\mathscr{F})) = \mathbb{E}(\mathbb{1}_F X^+)$ for all $F \in \mathscr{F}$. Note that with $F = \Omega$ we have in particular $\mathbb{E}(\mathbb{E}(X^+|\mathscr{F})) = \mathbb{E}(X^+) \in \mathbb{R}_{\geqslant 0}$. Consequently, we can choose $\mathbb{E}(X^+|\mathscr{F}) \in \mathcal{M}_{\geqslant 0}(\mathscr{F})$ and similarly $\mathbb{E}(X^-|\mathscr{F}) \in \mathcal{M}_{\geqslant 0}(\mathscr{F})$. In this case, the random variable $Y := \mathbb{E}(X^+|\mathscr{F}) - \mathbb{E}(X^-|\mathscr{F}) \in \mathcal{M}(\mathscr{F})$ is again solution to the Radon-Nikodym equation (cE2), where evidently $Y \in \bar{\mathcal{L}}^1(\mathscr{F}, \mathbb{P})$. □

§07.02 **Definition**. For $X \in \bar{\mathcal{L}}^1(\mathscr{A}, \mathbb{P})$, any function $Y : \Omega \to \overline{\mathbb{R}}$ is called *version (specification) of the conditional expectation* of X given $\mathscr{F}$, symbolically $\boxed{\mathbb{E}(X|\mathscr{F})} := Y$, if it satisfies the following two conditions:

(cE1) $Y \in \bar{\mathcal{L}}^1(\mathscr{F}, \mathbb{P})$, i.e. $Y \in \overline{\mathcal{M}}(\mathscr{F})$ is $\mathbb{P}$-integrable

(cE2) $\mathbb{E}(\mathbb{1}_F Y) = \mathbb{E}(\mathbb{1}_F X)$ for every $F \in \mathscr{F}$. (*Radon-Nikodym equation*)

Any mapping

$$\boxed{\mathbb{E}(\bullet|\mathscr{F})} : \bar{\mathcal{L}}^1(\mathscr{A}, \mathbb{P}) \ni X \mapsto \mathbb{E}(X|\mathscr{F}) \in \bar{\mathcal{L}}^1(\mathscr{F}, \mathbb{P})$$

is called a *version (specification) of the conditional expectation* with respect to $\mathbb{P}$ given $\mathscr{F}$. $\mathbb{E}(\bullet|\mathscr{F})$ is called *regular version* of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$, if the implied version $\mathbb{P}(\bullet|\mathscr{F})$ of the conditional distribution of $\mathbb{P}$ given $\mathscr{F}$ as in (06.01) is regular, and for every $\omega \in \Omega$ the map $\mathbb{E}(\bullet|\mathscr{F})(\omega) : \bar{\mathcal{L}}^1(\mathscr{A}, \mathbb{P}) \to \mathbb{R}$ is the expectation with respect to the probability measure $\mathbb{P}(\bullet|\mathscr{F})(\omega)$. □

§07.03 **Remark**.

(a) Let X be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{X}, \mathscr{X})$ and let $\mathscr{F} \subseteq \mathscr{A}$ be a sub-σ-algebra. We define similarly to Definition §06.22 (b) for $h \in \bar{\mathcal{L}}^1(\mathscr{X}, \mathbb{P}^X)$ a version $\mathbb{E}^X(h|\mathscr{F}) \in \bar{\mathcal{L}}^1(\mathscr{F}, \mathbb{P})$ of the conditional expectation of $h(X)$ given $\mathscr{F}$ as well as a *(regular) version*

$$\boxed{\mathbb{E}^X(\bullet|\mathscr{F})} : \bar{\mathcal{L}}^1(\mathscr{X}, \mathbb{P}^X) \ni h \mapsto \mathbb{E}^X(h|\mathscr{F}) \in \bar{\mathcal{L}}^1(\mathscr{F}, \mathbb{P})$$

of the *conditional expectation with respect to $\mathbb{P}^X$ given $\mathscr{F}$*.

(b) Let S be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable space $(\mathcal{S}, \mathscr{S})$. We define similarly to Definition §06.25 for $X \in \bar{\mathcal{L}}^1(\mathscr{A}, \mathbb{P})$ a version $\mathbb{E}^{|S}(X) \in \bar{\mathcal{L}}^1(\mathscr{S}, \mathbb{P}^S)$ of the conditional expectation of X given S as well as a *(regular) version*

$$\boxed{\mathbb{E}^{|S}} : \bar{\mathcal{L}}^1(\mathscr{A}, \mathbb{P}) \ni X \mapsto \mathbb{E}^{|S}(X) \in \bar{\mathcal{L}}^1(\mathscr{S}, \mathbb{P}^S)$$

of the *conditional expectation with respect to $\mathbb{P}$ given S*.

(c) Let (S, X) be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a measurable product space $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$. We define similarly to Definition §06.31 (b) for $h \in \bar{\mathcal{L}}^1(\mathscr{X}, \mathbb{P}^X)$ a version $\mathbb{E}^{X|S}(h) \in \bar{\mathcal{L}}^1(\mathscr{S}, \mathbb{P}^S)$ of the conditional expectation of $h(X)$ given S as well as a *(regular) version*

$$\boxed{\mathbb{E}^{X|S}} : \bar{\mathcal{L}}^1(\mathscr{X}, \mathbb{P}^X) \ni h \mapsto \mathbb{E}^{X|S}(h) \in \bar{\mathcal{L}}^1(\mathscr{S}, \mathbb{P}^S)$$

of the *conditional expectation with respect to $\mathbb{P}^X$ given S*. If similar to Definition §06.31 (a) $\mathbb{P}^{X|S}$ *is a regular version of the conditional distribution of $\mathbb{P}^X$ given S*, and if a s-slice $\mathbb{P}^{X|S=s} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{X})$ has, for example, a finite first absolute moment, then we write also $\boxed{\mathbb{P}^{|S=s}(X)} := \mathbb{P}^{X|S=s}(\mathrm{id}_{\mathcal{X}}) = \int_{\mathcal{X}} x\mathbb{P}^{X|S=s}(\mathrm{d}x)$. □

§07.04 **Theorem**. *Let $X, Y \in \bar{\mathcal{L}}^1(\mathscr{A}, \mathbb{P})$ and $\mathscr{F} \subseteq \mathscr{A}$ be a sub-σ-algebra. For every version $\mathbb{E}(\bullet|\mathscr{F})$ of the conditional expectation with respect to $\mathbb{P}$ given $\mathscr{F}$ the following statements hold $\mathbb{P}$-a.e.*

(i) *If $aX + bY \in \overline{\mathcal{M}}(\mathscr{A})$ for $a, b \in \mathbb{R}$, then $\mathbb{E}(aX + bY|\mathscr{F}) = a\mathbb{E}(X|\mathscr{F}) + b\mathbb{E}(Y|\mathscr{F})$;* (linear)

(ii) $\mathbb{E}(X|\mathscr{F}) \leqslant \mathbb{E}(Y|\mathscr{F})$ *if $X \leqslant Y$;* (monotone)

(iii) $|\mathbb{E}(X|\mathscr{F})| \leqslant \mathbb{E}(|X||\mathscr{F})$; (triangle inequality)

(iv) $\mathbb{P}(|S| = \infty) = 0$ *for every $S \in \overline{\mathcal{M}}(\mathscr{A})$ with $\mathbb{E}(|S||\mathscr{F}) \in \mathbb{R}_{\geqslant 0}$ $\mathbb{P}$-a.e.;* (finite)

(v) $\phi(\mathbb{E}(X|\mathscr{F})) \leqslant \mathbb{E}((\phi(X))|\mathscr{F})$ *for every convex $\phi : \mathbb{R} \to \mathbb{R}$ with $X, \phi(X) \in \mathcal{L}^1(\mathscr{A}, \mathbb{P})$;* (Jensen's inequality)

(vi) $\lim_{n\to\infty} \mathbb{E}(X_n|\mathscr{F}) = \mathbb{E}(X|\mathscr{F})$ $\mathbb{P}$-a.e. *and in $\mathcal{L}^1(\mathscr{A}, \mathbb{P})$ for all $X_n \to X$ $\mathbb{P}$-a.e. with $\sup_{n\in\mathbb{N}} |X_n| \leqslant Y$.* (dominated convergence)

If the version is regular, and hence, $\mathbb{E}(\bullet|\mathscr{F})(\omega)$ is the expectation with respect to the probability measure $\mathbb{P}(\bullet|\mathscr{F})(\omega)$ for all $\omega \in \Omega$, then the statements (i)-(vi) hold for all $\omega \in \Omega$.

§07.05 **Proof** of Theorem §07.04. is given in the lecture. For a regular version (i)-(vi) are shown in the lecture EWS. □

§07.06 **Theorem**. *Let $X, Y \in \bar{\mathcal{L}}^1(\mathscr{A}, \mathbb{P})$ and let $\mathscr{G} \subseteq \mathscr{F} \subseteq \mathscr{A}$ be sub-σ-algebras. For every version $\mathbb{E}(\bullet|\mathscr{F})$ and $\mathbb{E}(\bullet|\mathscr{G})$ of the conditional expectation with respect to $\mathbb{P} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ given $\mathscr{F}$ and $\mathscr{G}$, respectively, the following statements hold*

(i) $\mathbb{E}(XZ|\mathscr{F}) = Z\mathbb{E}(X|\mathscr{F})$ $\mathbb{P}$-a.e. *for all $Z \in \mathcal{M}(\mathscr{F})$ with $\mathbb{E}(|XZ|) \in \mathbb{R}_{\geqslant 0}$ and $\mathbb{E}(Z|\mathscr{F}) = \mathbb{E}(Z|\sigma(Z)) = Z$ $\mathbb{P}$-a.e. for all $Z \in \bar{\mathcal{L}}^1(\mathscr{F}, \mathbb{P})$;*

(ii) $\mathbb{E}(\mathbb{E}(X|\mathscr{F})|\mathscr{G}) = \mathbb{E}(\mathbb{E}(X|\mathscr{G})|\mathscr{F}) = \mathbb{E}(X|\mathscr{G})$ $\mathbb{P}$-a.e.; (tower property)

(iii) $\mathbb{E}(X|\mathscr{F}) = \mathbb{E}(X)$ $\mathbb{P}$-a.e. *if $\sigma(X)$ and $\mathscr{F}$ are independent;* (independence)

(iv) $\mathbb{E}(X|\overline{\mathscr{T}}) = \mathbb{E}(X)$ $\mathbb{P}$-a.e. *for $\overline{\mathscr{T}} := \{A \in \mathscr{A} \mid \mathbb{P}(A) \in \{0, 1\}\}$;*

(v) $\mathbb{E}(\mathbb{E}(X|\mathscr{F})) = \mathbb{E}(X)$. (total probability)

In addition let $S : (\Omega, \mathscr{A}) \to (\mathcal{S}, \mathscr{S})$, $T : (\Omega, \mathscr{A}) \to (\mathcal{T}, \mathscr{T})$. Then

(vi) $\mathbb{E}^{|S}(h(S)X) = h\mathbb{E}^{|S}(X)$ $\mathbb{P}^S$-a.e. *for $h \in \mathcal{M}(\mathscr{S})$ with $\mathbb{E}(|h(S)X|) \in \mathbb{R}_{\geqslant 0}$;*

(vii) $\mathbb{E}^{|S,T}(XY) = \mathbb{E}^{|S}(X)\mathbb{E}^{|T}(Y)$ $\mathbb{P}^{S,T}$-a.e. *if (X, S) and (Y, T) are independent.*

§07.07 **Proof** of Theorem §07.06. is given in the lecture. □

End LE11 2026/05/22

§07|01 Regular versions

§07.08 **Reminder**. Let (S, X) be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in the measurable product space $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$. A regular version $\mathbb{P}^{X|S}$ of the conditional distribution of $\mathbb{P}^X$ given S, if it exists, is a Markov kernel from $(\mathcal{S}, \mathscr{S})$ to $(\mathcal{X}, \mathscr{X})$, which is a solution to the Radon-Nikodym equation $\mathbb{P}^S \odot \mathbb{P}^{X|S} = \mathbb{P}^{S,X}$, that is, for every $B \in \mathscr{X}$ hold

$$\mathbb{P}(S^{-1}(A) \cap X^{-1}(B)) = \mathbb{P}^{S,X}(A \times B) = \mathbb{P}^S(\mathbb{1}_A \mathbb{P}^{X|S}(B)) = \int_A \mathbb{P}^{X|S}(B)\, d\mathbb{P}^S \quad \forall\, A \in \mathscr{S}.$$

Often it is useful to consider, in addition to $\mathbb{P}^{X|S}$, a regular version $\mathbb{P}^{S,X|S}$ of the conditional distribution of $\mathbb{P}^{S,X}$ given S defined as a Markov kernel $(s, D) \mapsto \mathbb{P}^{S,X|S=s}(D)$ from $(\mathcal{S}, \mathscr{S})$ to $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$, which is for every $D \in \mathscr{S} \otimes \mathscr{X}$ solution to the Radon-Nikodym equation

$$\mathbb{P}((S, X)^{-1}(D) \cap S^{-1}(B)) = \mathbb{P}^{S,X}(D \cap (B \times \mathcal{X})) = \mathbb{P}^S(\mathbb{P}^{S,X|S}(D)\mathbb{1}_B) = \int_B \mathbb{P}^{S,X|S}(D)\, d\mathbb{P}^S \quad \forall\, B \in \mathscr{S}.$$

Then the conditional distribution $\mathbb{P}^{X|S}$ arises again from the conditional distribution $\mathbb{P}^{S,X|S}$ as the marginal distribution according to $\mathbb{P}^{X|S}(B) = \mathbb{P}^{S,X|S}(\mathcal{S} \times B)$, $B \in \mathscr{X}$. We further denote with $D_s := \{x \in \mathcal{X} : (s,x) \in D\}$ or $h_s(x) := h(s,x)$ the *s-slice* of $D \subseteq \mathcal{S} \times \mathcal{X}$ or $h : \mathcal{S} \times \mathcal{X} \to \mathbb{R}$, respectively, for a given $s \in \mathcal{S}$. □

§07.09 **Theorem**. *Let (S,X) be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$. Let $\mathbb{P}^{X|S}$ be a regular version of the conditional distribution of $\mathbb{P}^X$ given S.*

(i) *The map $(s,D) \mapsto \mathbb{P}^{S,X|S=s}(D) := \mathbb{P}^{X|S=s}(D_s)$ is a regular version of the conditional distribution of $\mathbb{P}^{S,X}$ given S.*

(ii) *If $h \in \mathcal{M}(\mathscr{S} \otimes \mathscr{X})$ is $\mathbb{P}^{S,X}$-quasiintegrable, then*

$$\mathbb{P}^{S,X|S=s}(h) = \int h_s \,\mathrm{d}\mathbb{P}^{X|S=s} = \mathbb{P}^{X|S=s}(h_s) \quad \text{for } \mathbb{P}^S\text{-a.e. } s \in \mathcal{S},$$

as well as for all $A \in \mathscr{S}$ and $B \in \mathscr{X}$

$$\mathbb{P}^{S,X}(h\mathbb{1}_{A\times B}) = \mathbb{P}^S\big(\mathbb{1}_A \mathbb{P}^{S,X|S}(\mathbb{1}_B h)\big) = \int_A \Big(\int_B h_s \,\mathrm{d}\mathbb{P}^{X|S=s}\Big)\mathbb{P}^S(\mathrm{d}s).$$

§07.10 **Proof** of Theorem §07.09. is given in the lecture. □

§07.11 **Remark**. Intuitively, it is tempting to write $\mathbb{P}^{X|S}(f_S)$ for $\mathbb{P}^{(X,S)|S}(f)$. However, this is not justified because $\mathbb{P}^{X|S}(f_S)$ cannot be obtained as an $\mathscr{S}$-measurable solution of a Radon-Nikodym equation, and thus cannot be formulated in the usual way as a conditional expectation (see Witting (1985, S.125) for further discussion). On the other hand, for simplification in computation, $f(S,X)$ can be replaced by $f(s,X)$ in the integrand, so the symbolic notation $S = s$ can be understood as substitution. The justification for such substitution follows from the fact that $(s,D) \mapsto \mathbb{P}^{X|S=s}(D_s)$ is a regular version of the conditional expectation $\mathbb{P}^{(S,X)|S}$ (under the assumptions of Theorem §07.09). For $D = \{s\} \times \mathcal{X}$, it follows that $\mathbb{P}^{(S,X)|S=s}(D) = \mathbb{P}^{X|S=s}(\mathcal{X}) = 1$ for $\mathbb{P}^S$-a.e. $s \in \mathcal{S}$. Thus, the conditional distribution $\mathbb{P}^{(S,X)|S}$ concentrates on the set $\{S = s\}$ for $\mathbb{P}^S$-a.e. $s \in \mathcal{S}$. □

§07.12 **Example**.

(a) Let $X, S \in \mathcal{L}^1(\mathscr{A}, \mathbb{P})$ be independent. Then $\mathbb{P}$-a.e. hold

$$\mathbb{E}\big(X + S\big|\sigma(S)\big) = \mathbb{E}\big(X\big|\sigma(S)\big) + \mathbb{E}\big(S\big|\sigma(S)\big) = \mathbb{E}(X) + S.$$

More general, let $f \in \mathcal{M}(\mathscr{S} \otimes \mathscr{X})$ be $\mathbb{P}^{S,X}$-quasiintegrable. Since $X \perp\!\!\!\perp S$ (see also Theorem §07.13 below) the Markov kernel $(s,B) \mapsto \mathbb{P}^X(B)$ from $(\mathcal{S}, \mathscr{S})$ to $(\mathcal{X}, \mathscr{X})$ is a regular version of the conditional distribution of $\mathbb{P}^X$ given S. Making use of Theorem §07.09 (ii) it follows that $\mathbb{E}^{|S=s}(f(S,X)) = \mathbb{P}^{S,X|S=s}(f) = \mathbb{P}^X(f_s) = \mathbb{P}\big(f(s,X)\big)$ for $\mathbb{P}^S$-a.e. $s \in \mathcal{S}$.

(b) Let $(X_i)_{i\in[\![n]\!]}$ be an independent family in $\mathcal{L}^1(\mathscr{A}, \mathbb{P})$ with $\mathbb{E}(X_i) = 0$, $i \in [\![n]\!]$. For $m \in [\![n]\!]$ set $\mathscr{F}_m := \sigma((X_i)_{i\in[\![m]\!]})$ and $S_m := \sum_{i\in[\![m]\!]} X_i$, then $\mathbb{P}$-a.e. holds

$$\mathbb{E}\big(S_n\big|\mathscr{F}_m\big) = \sum_{i\in[\![n]\!]} \mathbb{E}\big(X_i\big|\mathscr{F}_m\big) = \sum_{i\in[\![m]\!]} X_i + \sum_{i\in(\![m,n]\!]} \mathbb{E}(X_i) = S_m.$$

Since $\sigma(S_m) \subseteq \mathscr{F}_m$ from Theorem §07.06 (ii) follows $\mathbb{P}$-a.e.

$$\mathbb{E}\big(S_n\big|\sigma(S_m)\big) = \mathbb{E}\big(\mathbb{E}\big(S_n\big|\mathscr{F}_m\big)\big|\sigma(S_m)\big) = \mathbb{E}\big(S_m\big|\sigma(S_m)\big) = S_m.$$ □

§07.13 **Theorem**. *Let (S,X) be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$. If there exists a regular version $\mathbb{P}^{X|S}$ of the conditional distribution of $\mathbb{P}^X$ given S, then the following three statements are equivalent:*

(i) *X and S are independent under $\mathbb{P}$, symbolically $X \perp\!\!\!\perp S$.*

(ii) *There exists a version $\mathbb{P}^{X|S}$ which is independent of S.*

(iii) $\mathbb{P}^{X|S} = \mathbb{P}^X \quad \mathbb{P}^S$-a.e..

If $X \perp\!\!\!\perp S$ then for each $g : (\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X}) \to (\mathcal{T}, \mathscr{T})$ there exists a $\mathbb{P}^S$-null set N such that $\mathbb{P}^{g(S,X)|S=s}(C) = \mathbb{P}^{g(s,X)|S=s}(C)$ for all $C \in \mathscr{T}$ and $s \in N^c$.

§07.14 **Proof** of Theorem §07.13. (i)⇔(ii)⇔(iii) is an exercise, for the addendum we refer to Witting (1985, Satz 1.129, S.130). □

§07|02 Conditional density

§07.15 **Preliminary**. An explicit representation of a regular version of the conditional distribution $\mathbb{P}^{X|S}$ of X given S, and thus a way to demonstrate its existence, can be provided when the joint distribution has a density with respect to a product measure of σ-finite measures. We still denote by $D_s := \{x \in \mathcal{X} : (s,x) \in D\}$ or $f_s(x) := f(s,x)$ the *s-slice* of $D \subseteq \mathcal{S} \times \mathcal{X}$ or $f : \mathcal{S} \times \mathcal{X} \to \overline{\mathbb{R}}$, respectively, for a given $s \in \mathcal{S}$, and by $f^x(s) := f(s,x)$ the *x-slice* of f for a given $x \in \mathcal{X}$. □

§07.16 **Definition**. Let (S, X) be a random variable on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$ and $\mathbb{P}^{(S,X)} \ll (\nu \otimes \mu)$ with $(\nu \otimes \mu)$-density $\mathbb{f} := \mathrm{d}\mathbb{P}^{(S,X)}/\mathrm{d}(\nu \otimes \mu) \in \mathcal{M}_{\geqslant 0}(\mathscr{S} \otimes \mathscr{X})$ for σ-finite measures $\mu \in \mathfrak{M}_\sigma(\mathscr{X})$ and $\nu \in \mathfrak{M}_\sigma(\mathscr{S})$. We denote by $\mathbb{f}^X \in \mathcal{M}_{\geqslant 0}(\mathscr{X})$ and $\mathbb{f}^S \in \mathcal{M}_{\geqslant 0}(\mathscr{S})$ version of the marginal μ-density and ν-density, respectively, that is, $\mathbb{f}^S(s) = \int \mathbb{f}_s \,\mathrm{d}\mu = \mu(\mathbb{f}_s)$ for ν-a.e. $s \in \mathcal{S}$ and $\mathbb{f}^X(x) = \nu(\mathbb{f}^x)$ for μ-a.e. $x \in \mathcal{X}$. The $(\mathscr{S} \otimes \mathscr{X})$-$\mathscr{B}_{\geqslant 0}$-measurable map

$$\boxed{\mathbb{f}^{X|S}} := (\mathbb{f}/\mathbb{f}^S)\mathbb{1}_{\{\mathbb{f}^S \in \mathbb{R}_{>0}\}} + \mathbb{f}^X \mathbb{1}_{\{\mathbb{f}^S = 0\}} : \mathcal{S} \times \mathcal{X} \to \mathbb{R}_{\geqslant 0} \text{ with}$$

$$(s,x) \mapsto \mathbb{f}^{X|S}(s,x) := \frac{\mathbb{f}(s,x)}{\mathbb{f}^S(s)} \mathbb{1}_{\{\mathbb{f}^S(s) \in \mathbb{R}_{>0}\}} + \mathbb{f}^X(x) \mathbb{1}_{\{\mathbb{f}^S(s) = 0\}} \quad (07.01)$$

is called *regular version of the conditional μ-density* of $\mathbb{P}^X$ given S. A s-slice of $\mathbb{f}^{X|S} \in \mathcal{M}_{\geqslant 0}(\mathscr{S} \otimes \mathscr{X})$ is again denoted by $\mathbb{f}^{X|S=s} \in \mathcal{M}_{\geqslant 0}(\mathscr{X})$. □

§07.17 **Theorem**. *Under the assumptions of Definition §07.16 we have*

(i) *The Markov kernel $(s, B) \mapsto \mathbb{P}^{X|S=s}(B) = \mathbb{f}^{X|S=s}\mu(B) = \mu(\mathbb{1}_B \mathbb{f}^{X|S=s})$ and*

$$(s, D) \mapsto \mathbb{P}^{S,X|S=s}(D) = \mathbb{P}^{X|S=s}(D_s) = \mathbb{f}^{X|S=s}\mu(D_s) = \mu(\mathbb{1}_{D_s}\mathbb{f}^{X|S=s}) = \int_{D_s} \mathbb{f}^{X|S=s}\,\mathrm{d}\mu$$

define regular versions $\mathbb{P}^{X|S}$ and $\mathbb{P}^{(S,X)|S}$ of the conditional distribution of $\mathbb{P}^X$ and $\mathbb{P}^{(S,X)}$ given S. In particular, $\mathbb{f}^{X|S}$ is a μ-density of the Markov kernel $\mathbb{P}^{X|S}$, i.e. $\mathbb{P}^{X|S} = \mathbb{f}^{X|S}\mu$.

(ii) *If $h \in \overline{\mathcal{M}}(\mathscr{S} \otimes \mathscr{X})$ is $\mathbb{P}^{(S,X)}$-quasiintegrable, then*

$$\mathbb{P}^{S,X|S=s}(h) = \mu(h_s \mathbb{f}^{X|S=s}) = \mathbb{f}^{X|S=s}\mu(h_s) \quad \textit{for } \mathbb{P}^S\text{-a.e. } s \in \mathcal{S},$$

as well as for all $B \in \mathscr{X}$ and $A \in \mathscr{S}$

$$\mathbb{P}^{S,X}(h\mathbb{1}_{A \times B}) = \nu\big(\mathbb{1}_A \mathbb{f}^S \mathbb{P}^{X|S}(\mathbb{1}_B h_\cdot)\big) = \int_A \mu(\mathbb{1}_B h_s \mathbb{f}^{X|S=s})\, \mathbb{f}^S(s)\nu(\mathrm{d}s).$$

§07.18 **Proof** of Theorem §07.17. is given in the lecture. □

§07.19 **Remark**. The terms introduced in Definition §05.04 (a) and (b) can now be incorporated into the more general approach. To this end, let (S, X) be a random variable with a joint distribution $\mathbb{P}^{S,X}$ on $(\mathcal{S} \times \mathcal{X}, \mathscr{S} \otimes \mathscr{X})$.

(a) Let $\mathcal{S} \times \mathcal{X}$ be countable, hence (S, X) is *discretely distributed*, as in Definition §05.04 (a). The mass function $\mathbb{p}^{X|S} : 2^{\mathcal{Z}} \otimes 2^{\mathcal{X}} \to [0,1]$ given by

$$(z, x) \mapsto \mathbb{p}^{X|S}(s, x) = \frac{\mathbb{p}^{S,X}(s,x)}{\mathbb{p}^{S}(s)} \mathbb{1}_{\{\mathbb{p}^S(s) \in \mathbb{R}_{>0}\}} + \mathbb{p}^{X}(x) \mathbb{1}_{\{\mathbb{p}^S(s)=0\}}$$

defines a *regular version* $\mathbb{p}^{X|S}$ *of the conditional* $\zeta_{\mathcal{X}}$*-mass function of* $\mathbb{P}^X$ *given* S. A s-slice of $\mathbb{p}^{X|S} \in \mathcal{M}_{\geqslant 0}(2^{\mathcal{Z}} \otimes 2^{\mathcal{X}})$ is again denoted by $\mathbb{p}^{X|S=s} \in \mathcal{M}_{\geqslant 0}(2^{\mathcal{X}})$. The Markov kernel $\mathbb{P}^{X|S} = \mathbb{p}^{X|S}\zeta_{\mathcal{X}}$ from $(\mathcal{S}, 2^{\mathcal{S}})$ to $(\mathcal{X}, 2^{\mathcal{X}})$ with $\zeta_{\mathcal{X}}$-mass function $\mathbb{p}^{X|S}$ satisfies (cE1), i.e. $\mathbb{P}^{X|S}(B) \in \mathcal{M}_{[0,1]}(2^{\mathcal{S}})$, and (cE2), because for every $B \in 2^{\mathcal{X}}$ and $A \in 2^{\mathcal{S}}$

$$\mathbb{P}^{S,X}(A \times B) = \sum_{s \in A} \sum_{x \in B} \mathbb{p}^{S,X}(x, s) = \sum_{s \in A} \mathbb{p}^{S}(s) \mathbb{P}^{X|S=s}(B) = \mathbb{P}^{S}(\mathbb{1}_A \mathbb{P}^{X|S}(B)).$$

(b) Let (S, X) be *continuous distributed* with values in $(\mathbb{R}^{m+n}, \mathscr{B}^{m+n})$, as in Definition §05.04 (b). The $\mathscr{B}^{m+n}$-$\mathscr{B}_{\geqslant 0}$-measurable map

$$(s, x) \mapsto \mathbb{f}^{X|S}(s, x) = \frac{\mathbb{f}^{S,X}(s,x)}{\mathbb{f}^{S}(s)} \mathbb{1}_{\{\mathbb{f}^S(s) \in \mathbb{R}_{>0}\}} + \mathbb{f}^{X}(x) \mathbb{1}_{\{\mathbb{f}^S(s)=0\}}$$

defines a *regular version* $\mathbb{f}^{X|S}$ *of the conditional* λ^n*-density of* $\mathbb{P}^X$ *given* S. A s-slice of $\mathbb{f}^{X|S} \in \mathcal{M}_{\geqslant 0}(\mathscr{B}^{m+n})$ is again denoted by $\mathbb{f}^{X|S=s} \in \mathcal{M}_{\geqslant 0}(\mathscr{B}^n)$. The Markov kernel $\mathbb{P}^{X|S} = \mathbb{f}^{X|S}\lambda^n$ from $(\mathbb{R}^m, \mathscr{B}^m)$ to $(\mathbb{R}^n, \mathscr{B}^n)$ with λ^n-density $\mathbb{f}^{X|S}$ satisfies (cE1), i.e. $\mathbb{P}^{X|S}(B) \in \mathcal{M}_{[0,1]}(\mathscr{B}^m)$, and (cE2), because for every $B \in \mathscr{B}^n$ and $A \in \mathscr{B}^m$

$$\mathbb{P}^{S,X}(A \times B) = \int_A \int_B \mathbb{f}^{S,X}(s, x)\, \lambda^n(\mathrm{d}x)\, \lambda^m(\mathrm{d}s)$$
$$= \int_A \mathbb{f}^{S}(s) \mathbb{P}^{X|S=s}(B)\, \lambda^m(\mathrm{d}s) = \mathbb{P}^{S}(\mathbb{1}_A \mathbb{P}^{X|S}(B)) \quad \square$$

§07.20 **Corollary**. *Under the assumptions of Definition §07.16 X and S are independent under $\mathbb{P}$, symbolically $X \perp\!\!\!\perp S$, if and only if, $\mathbb{f}^{X|S=s} = \mathbb{f}^{X}$ μ-a.e. for $\mathbb{P}^S$-a.e. $s \in \mathcal{S}$.*

§07.21 **Proof** of Corollary §07.20. The claim follows directly from Theorems §07.13 and §07.17. $\square$

§07.22 **Example**.

(a) By Corollary §07.20 in Example §05.06 (a) X and S are not independent, since, for instance, $\mathbb{p}^{X|S=2} \neq \mathbb{p}^{X}$ with $\mathbb{P}^{S}(\{2\}) = \mathbb{p}^{S}(2) = 1/36 \in \mathbb{R}_{>0}$. On the other hand, in Example §05.06 (b) X and S are independent, i.e., $\mathbb{p}^{X|S} = \mathbb{p}^{X}$ $\mathbb{P}^S$-a.e., if and only if $\Sigma_{XS} = 0$, that is, if X and S are uncorrelated.

(b) For $a \in \mathbb{R}_{>0}$ let $X \sim \mathrm{Exp}_a$ and $Y \sim \mathrm{Exp}_a$ be independent. Then X and $S = X + Y$ are not independent, since $\mathbb{P}^{X|S=s} = \mathrm{U}_{[0,s]}$ for λ-a.e. $s \in \mathbb{R}_{\geqslant 0}$. In contrast, $W := X/(X+Y)$ and S are independent, since $\mathbb{P}^{W|S} = \mathrm{U}_{[0,1]}$ λ-a.e. (exercise). $\square$

§07|03 Best prediction

§07.23 **Preliminary**. Let $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}})$ be a Hilbert space, let $\|\cdot\|_{\mathbb{H}}$ denote the induced norm, and let $\mathbb{U}$ be a closed linear subspace of $\mathbb{H}$. For every element $h \in \mathbb{H}$ there exists a uniquely determined element $u_h \in \mathbb{U}$ with $\|h - u_h\|_{\mathbb{H}} = \inf_{u \in \mathbb{U}} \|h - u\|_{\mathbb{H}}$. The map $\Pi_{\mathbb{U}} : \mathbb{H} \to \mathbb{U}$ with $h \mapsto \Pi_{\mathbb{U}}(h) := u_h$ is called *orthogonal projection*. If $\mathbb{U}^{\perp}$ denotes the orthogonal complement of $\mathbb{U}$ in $\mathbb{H}$, then for all $h, g \in \mathbb{H}$ hold:

(a) $\Pi_{\mathbb{U}} \circ \Pi_{\mathbb{U}} = \Pi_{\mathbb{U}}$, (*projection*)

(b) $h - \Pi_{\mathbb{U}}(h) \in \mathbb{U}^{\perp}$, (*orthogonal*)

(c) $h = \Pi_{\mathbb{U}}(h) + (h - \Pi_{\mathbb{U}}(h))$ unique decomposition of h into an orthogonal direct sum of elements from $\mathbb{U}$ and $\mathbb{U}^{\perp}$,

(d) $\Pi_{\mathbb{U}}$ is linear and self-adjoint, that is, $\langle \Pi_{\mathbb{U}}(h), g\rangle_{\mathbb{H}} = \langle h, \Pi_{\mathbb{U}}(g)\rangle_{\mathbb{H}}$. □

§07.24 **Property**. *$\mathcal{L}^2(\mathscr{F}, \mathbb{P})$ is a closed linear subspace of $\mathcal{L}^2(\mathscr{A}, \mathbb{P})$ for every sub-σ-algebra $\mathscr{F} \subseteq \mathscr{A}$.* □

§07.25 **Lemma**. *Let $\mathscr{F} \subseteq \mathscr{A}$ be a sub-σ-algebra and let $\mathbb{E}\big(\bullet\,\big|\mathscr{F}\big)$ be a version of the conditional expectation. Then $\mathbb{E}\big(\bullet\,\big|\mathscr{F}\big) : \mathcal{L}^2(\mathscr{A}, \mathbb{P}) \to \mathcal{L}^2(\mathscr{F}, \mathbb{P})$ is a orthogonal projection, that is, for all $X \in \mathcal{L}^2(\mathscr{A}, \mathbb{P})$ and $Y \in \mathcal{L}^2(\mathscr{F}, \mathbb{P})$ we have*

$$\|X - Y\|^2_{\mathcal{L}^2} = \mathbb{E}(|X - Y|^2) \geqslant \mathbb{E}(|X - \mathbb{E}\big(X\big|\mathscr{F}\big)|^2) = \|X - \mathbb{E}\big(X\big|\mathscr{F}\big)\|^2_{\mathcal{L}^2},$$

where equality holds if and only if $Y = \mathbb{E}\big(X\big|\mathscr{F}\big)$ $\mathbb{P}$-a.e.

§07.26 **Proof** of Lemma §07.25. Exercise. □

§07.27 **Comment**. Let $S : (\Omega, \mathscr{A}) \to (\mathcal{S}, \mathscr{S})$ and $\varphi \in \overline{\mathcal{M}}(\sigma(S))$ such that $\mathbb{E}\big(X\big|\mathscr{S}\big) = \varphi(S)$. By Lemma §07.25 we have $\mathbb{E}\big(|X - \varphi(S)|^2\big) \leqslant \mathbb{E}\big(|X - h(S)|^2\big)$ for any $h \in \mathcal{L}^2(\mathscr{S}, \mathbb{P}^S)$. $\varphi(S)$ is called *best prediction* of X by S. In the lecture EWS we have shown for any $S \in \mathcal{L}^2$ with $\mathbb{V}\mathrm{ar}(S) \in \mathbb{R}_{>0}$ that $a^*S + b^* = \mathbb{E}(X) + \frac{\mathbb{C}\mathrm{ov}(X,S)}{\mathbb{V}\mathrm{ar}(S)}(S - \mathbb{E}(S))$ satisfies $\mathbb{E}\big(|X - (a^*S + b^*)|^2\big) \leqslant \mathbb{E}\big(|X - (aS + b)|^2\big)$ for all $a, b \in \mathbb{R}$. $a^*S + b^*$ is called *best linear prediction* of X by S. Evidently, any best prediction, which is also (affine) linear, is a best linear prediction, but the converse does generally not apply. □

§07.28 **Example**.

(a) Let X and S be jointly normally distributed (as in Example §05.06 (b) with $n = m = 1$). If $\mathbb{V}\mathrm{ar}(S) \in \mathbb{R}_{>0}$ then the best prediction and the best linear prediction coincide, that is, $\mathbb{E}\big(X\big|\sigma(S)\big) = \mu_X + \Sigma_{XS}\Sigma_S^{-1}(S - \mu_S)$.

(b) Let U and S be independent and identically $\mathrm{N}_{(0,1)}$-distributed, and let $X := S^2 + S + U$. Then $\mathbb{E}\big(X\big|\sigma(S)\big) = S^2 + S$ is the best prediction of X by S (use Example §07.12 (a)). Its mean squared prediction error satisfies $\mathbb{E}\big(|X - \mathbb{E}\big(X\big|\sigma(S)\big)|^2\big) = \mathbb{E}(U)^2 = 1$. The best linear prediction of X by S is in this case $Z^* = S + 1$, since $\mathbb{C}\mathrm{ov}(X, S) = 1$, $\mathbb{E}(X) = 1$, $\mathbb{V}\mathrm{ar}(S) = 1$ and $\mathbb{E}(S) = 0$. Therefore, we have $\mathbb{E}\big(|X - Z^*|^2\big) = \mathbb{E}(S^4) + \mathbb{E}(U^2) - 2\mathbb{E}(S^2) + 1 = 3 > 1 = \mathbb{E}\big(|X - \mathbb{E}\big(X\big|\sigma(S)\big)|^2\big)$. □

§07.29 **Lemma**. *Let $p \in \overline{\mathbb{R}}_{\geqslant 1}$, let $\mathscr{F} \subseteq \mathscr{A}$ be a sub-σ-algebra and let $\mathbb{E}\big(\bullet\,\big|\mathscr{F}\big)$ be a version of the conditional expectation. Then the linear map $\mathbb{E}\big(\bullet\,\big|\mathscr{F}\big) : \mathcal{L}^p(\mathscr{A}, \mathbb{P}) \to \mathcal{L}^p(\mathscr{F}, \mathbb{P})$ is a contraction, that is $\|\mathbb{E}\big(X\big|\mathscr{F}\big)\|_{\mathcal{L}^p} \leqslant \|X\|_{\mathcal{L}^p}$ for all $X \in \mathcal{L}^p(\mathscr{A}, \mathbb{P})$, and therefore it is bounded and continuous. Consequently, $(\mathbb{E}\big(X_n\big|\mathscr{F}\big))_{n\in\mathbb{N}}$ converges in $\mathcal{L}^p(\mathscr{F}, \mathbb{P})$ for every in $\mathcal{L}^p(\mathscr{A}, \mathbb{P})$ convergent $(X_n)_{n\in\mathbb{N}}$.*

§07.30 **Proof** of Lemma §07.29. is given in the lecture. □

§07.31 **Reminder**. A family $(X_i)_{i\in\mathcal{I}}$ of random variables in $\mathcal{L}^1(\mathscr{A}, \mathbb{P})$ with arbitrary nonempty index set $\mathcal{I}$ satisfying

$$\inf\{\sup_{i\in\mathcal{I}} \mathbb{P}\big(|X_i|\mathbb{1}_{\{|X_i|\geqslant a\}}\big) : a \in \mathbb{R}_{\geqslant 0}\} = 0$$

is called *uniformly $\mathbb{P}$-integrable* (see Definition §02.39). □

§07.32 **Lemma**. *For any nonempty index set $\mathcal{I}$ and $\mathcal{J}$ let $(X_i)_{i\in\mathcal{I}}$ be an uniformly $\mathbb{P}$-integrable family of random variables in $\mathcal{L}^1(\mathscr{A},\mathbb{P})$ and let $(\mathscr{F}_j)_{j\in\mathcal{J}}$ by a family of sub-σ-algebra of $\mathscr{A}$. Then the family $\big(\mathbb{E}(X_i|\mathscr{F}_j)\big)_{i\in\mathcal{I},j\in\mathcal{J}}$ is uniformly $\mathbb{P}$-integrable in $\mathcal{L}^1(\mathscr{A},\mathbb{P})$. In particular, for $X\in\mathcal{L}^1(\mathscr{A},\mathbb{P})$ and $\mathcal{J}:=\{\mathscr{F}: \mathscr{F}\text{ is sub-}\sigma\text{-algebra of }\mathscr{A}\}$, the family $\big(\mathbb{E}(X|\mathscr{F})\big)_{\mathscr{F}\in\mathcal{J}}$ is uniformly $\mathbb{P}$-integrable in $\mathcal{L}^1(\mathscr{A},\mathbb{P})$.*

§07.33 **Proof** of Lemma §07.32. Klenke (2020, Corollary 8.22, p.199) □

§08 Bayes approach

§08.01 **Preliminary**. Consider a statistical experiment $(\mathcal{X},\mathscr{X},\mathbb{P}_\Theta := (\mathbb{P}_\theta)_{\theta\in\Theta})$, where $\mathbb{P}_\theta\in\mathfrak{M}_{\text{pr}}(\mathscr{X})$ for every $\theta\in\Theta$ by definition. In addition let $(\Theta,\mathscr{T})$ be a measurable space. Consider the map

$$\mathbb{P}_\bullet : \Theta\times\mathscr{X}\to[0,1] \text{ with } (\theta,B)\mapsto\mathbb{P}_\theta(B).$$

Evidently, for each $\theta\in\Theta$ the map $\mathbb{P}_\theta:\mathscr{X}\to[0,1]$ (the θ-slice of $\mathbb{P}_\bullet$) is a probability measure, i.e. $\mathbb{P}_\theta\in\mathfrak{M}_{\text{pr}}(\mathscr{X})$. If for each $B\in\mathscr{X}$ the B-slice $\mathbb{P}_\bullet(B):\Theta\to[0,1]$ of $\mathbb{P}_\bullet$ with $\theta\mapsto\mathbb{P}_\theta(B)$ is $\mathscr{T}$-$\mathscr{B}_{[0,1]}$-measurable, i.e. $\mathbb{P}_\bullet(B)\in\mathcal{M}_{[0,1]}(\mathscr{X})$, then $\mathbb{P}_\bullet$ satisfies the conditions (Mk1) and (Mk2) (see Reminder §06.06), and hence it is a Markov kernel from $(\Theta,\mathscr{T})$ to $(\mathcal{X},\mathscr{X})$. □

§08.02 **Definition**. Let $(\mathcal{X},\mathscr{X},\mathbb{P}_\Theta)$ be a statistical experiment, $(\Theta,\mathscr{T})$ a measurable space and $\mathbb{P}_\bullet$ with $(\theta,F)\mapsto\mathbb{P}_\theta(F)$ a Markov kernel from $(\Theta,\mathscr{T})$ to $(\mathcal{X},\mathscr{X})$. Let $\boldsymbol{\vartheta}$ be a $(\Theta,\mathscr{T})$-valued random variable such that the parameter $\theta\in\Theta$ can be seen as realisation of $\boldsymbol{\vartheta}$. The probability measure $\mathbb{P}^{\boldsymbol{\vartheta}}\in\mathfrak{M}_{\text{pr}}(\mathscr{T})$ of $\boldsymbol{\vartheta}$ on $(\Theta,\mathscr{T})$ is called *a-priori distribution* of the parameter θ and we denote by $\mathbb{E}^{\boldsymbol{\vartheta}}$ the expectation with respect to $\mathbb{P}^{\boldsymbol{\vartheta}}$. □

§08.03 **Reminder**. A Markov kernel $\mathbb{P}_\bullet$ from $(\Theta,\mathscr{T})$ to $(\mathcal{X},\mathscr{X})$ and an a-priori distribution $\mathbb{P}^{\boldsymbol{\vartheta}}\in\mathfrak{M}_{\text{pr}}(\mathscr{T})$ determine uniquely (compare Lemma §04.57) the joint distribution $\mathbb{P}^{\boldsymbol{\vartheta},X}=\mathbb{P}^{\boldsymbol{\vartheta}}\odot\mathbb{P}_\bullet\in\mathfrak{M}_{\text{pr}}(\mathscr{T}\otimes\mathscr{X})$ on the measurable space $(\Theta\times\mathcal{X},\mathscr{T}\otimes\mathscr{X})$ where for all $A\in\mathscr{T}$ and $B\in\mathscr{X}$ hold

$$\mathbb{P}^{\boldsymbol{\vartheta},X}(A\times B)=(\mathbb{P}^{\boldsymbol{\vartheta}}\odot\mathbb{P}_\bullet)(A\times B)=\mathbb{P}^{\boldsymbol{\vartheta}}\big(\mathbb{1}_A\mathbb{P}_\bullet(B)\big)=\int_A\mathbb{P}_\theta(B)\mathbb{P}^{\boldsymbol{\vartheta}}(\mathrm{d}\theta).$$

The marginal distribution $\mathbb{P}^X:=\mathbb{P}^{\boldsymbol{\vartheta}}\mathbb{P}_\bullet$ of X on $(\mathcal{X},\mathscr{X})$ induced by $\mathbb{P}^{(\boldsymbol{\vartheta},X)}$ is given by

$$\mathbb{P}^X(B)=(\mathbb{P}^{\boldsymbol{\vartheta}}\mathbb{P}_\bullet)(B)=\mathbb{P}^{\boldsymbol{\vartheta}}\big(\mathbb{P}_\bullet(B)\big)=\int\mathbb{P}_\bullet(B)\,\mathrm{d}\mathbb{P}^{\boldsymbol{\vartheta}}=\int\mathbb{P}_\theta(B)\mathbb{P}^{\boldsymbol{\vartheta}}(\mathrm{d}\theta)$$

for all $B\in\mathscr{X}$ (compare Notation §06.07). □

§08.04 **Definition**. Let $(\mathcal{X},\mathscr{X},\mathbb{P}_\Theta)$ be a statistical experiment, $(\Theta,\mathscr{T})$ a measurable space, $(\boldsymbol{\vartheta},X)$ a random variable with values in $(\Theta\times\mathcal{X},\mathscr{T}\otimes\mathscr{X})$, $\mathbb{P}^{\boldsymbol{\vartheta}}\in\mathfrak{M}_{\text{pr}}(\mathscr{T})$ an a-priori distribution of $\boldsymbol{\vartheta}$ on $(\Theta,\mathscr{T})$, and $\mathbb{P}_\bullet$ a Markov kernel from $(\Theta,\mathscr{T})$ to $(\mathcal{X},\mathscr{X})$. The joint distribution $\mathbb{P}^{\boldsymbol{\vartheta},X}=\mathbb{P}^{\boldsymbol{\vartheta}}\odot\mathbb{P}_\bullet$ of $(\boldsymbol{\vartheta},X)$ is determined by the regular conditional distribution $\mathbb{P}^{X|\boldsymbol{\vartheta}}:=\mathbb{P}_\bullet$ of $\mathbb{P}^X$ given $\boldsymbol{\vartheta}$ and the a-priori distribution $\mathbb{P}^{\boldsymbol{\vartheta}}$. A regular version of the conditional distribution $\boxed{\mathbb{P}^{\boldsymbol{\vartheta}|X}}$ of $\mathbb{P}^{\boldsymbol{\vartheta}}$ given X, if it exists, is called *a-posteriori* distribution of $\boldsymbol{\vartheta}$ given X. If for every $x\in\mathcal{X}$ the a-posteriori distribution $\mathbb{P}^{\boldsymbol{\vartheta}|X=x}$ has a finite first absolute moment, then the random variable $\boxed{\mathbb{E}^{|X}(\boldsymbol{\vartheta})}=\mathbb{P}^{|X}(\boldsymbol{\vartheta})$ with $x\mapsto\mathbb{E}^{|X=x}(\boldsymbol{\vartheta}):=\mathbb{P}^{|X=x}(\boldsymbol{\vartheta}):=\mathbb{P}^{\boldsymbol{\vartheta}|X=x}(\mathrm{id}_\Theta)=\int_\Theta\theta\mathbb{P}^{\boldsymbol{\vartheta}|X=x}(\mathrm{d}\theta)$ is called *a-posteriori* expectation of $\boldsymbol{\vartheta}$ given X. □

§08.05 **Notation**. Let $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ be a statistical experiment, $(\Theta, \mathscr{T})$ a measurable space, and $\mathbb{P}_\Theta$ a family of probability measures dominated by a σ-finite measure $\mu \in \mathfrak{M}_\sigma(\mathscr{X})$ ($\mathbb{P}_\theta \ll \mu$ for all $\theta \in \Theta$) with corresponding μ-densities $(\mathbb{f}_\theta)_{\theta \in \Theta}$ in $\mathcal{M}_{\geqslant 0}(\mathscr{X})$. Let in addition $\mathbb{f}_\bullet : \Theta \times \mathcal{X} \to \mathbb{R}_{\geqslant 0}$ with $(\theta, x) \mapsto \mathbb{f}_\theta(x)$ be a $(\mathscr{T} \otimes \mathscr{X})$-$\mathscr{B}_{\geqslant 0}$-measurable function, i.e. $\mathbb{f}_\bullet \in \mathcal{M}_{\geqslant 0}(\mathscr{T} \otimes \mathscr{X})$. A θ-slice and x-slice of $\mathbb{f}_\bullet$ we denote by $\mathbb{f}_\theta \in \mathcal{M}_{\geqslant 0}(\mathscr{X})$ and $\mathbb{f}_\bullet(x) \in \mathcal{M}_{\geqslant 0}(\mathscr{T})$, respectively. Then, by Fubini's theorem

$$\mathbb{P}_\bullet = \mathbb{f}_\bullet \mu : \Theta \times \mathscr{X} \to [0,1] \text{ with } (\theta, B) \mapsto \mathbb{P}_\theta(B) = \mathbb{f}_\theta \mu(B) = \mu(\mathbb{1}_B \mathbb{f}_\theta) = \int_B \mathbb{f}_\theta \, \mathrm{d}\mu$$

is a Markov kernel from $(\Theta, \mathscr{T})$ to $(\mathcal{X}, \mathscr{X})$. Evidently, $\mathbb{f}_\bullet$ is the *μ-density* of the Markov kernel $\mathbb{P}_\bullet$ (see Definition §04.52). Consider a random variable $(\boldsymbol{\vartheta}, X)$ with values in $(\Theta \times \mathcal{X}, \mathscr{T} \otimes \mathscr{X})$. Let the a-priori distribution $\mathbb{P}^{\boldsymbol{\vartheta}}$ admit a ν-density $\mathbb{f}^{\boldsymbol{\vartheta}}$ with respect to a σ-finite measure $\nu \in \mathfrak{M}_\sigma(\mathscr{T})$. The joint distribution $\mathbb{P}^{\boldsymbol{\vartheta}, X} = \mathbb{P}^{\boldsymbol{\vartheta}} \odot \mathbb{P}^{X|\boldsymbol{\vartheta}}$ (determined by the regular conditional distribution $\mathbb{P}^{X|\boldsymbol{\vartheta}} := \mathbb{P}_\bullet$ and the a-priori distribution $\mathbb{P}^{\boldsymbol{\vartheta}}$) is dominated by the product measure $\nu \otimes \mu \in \mathfrak{M}_\sigma(\mathscr{T} \otimes \mathscr{X})$ with *$(\nu \otimes \mu)$-density*

$$\mathbb{f}^{\boldsymbol{\vartheta}} \mathbb{f}_\bullet \in \mathcal{M}_{\geqslant 0}(\mathscr{T} \otimes \mathscr{X}) \text{ with } (\theta, x) \mapsto \mathbb{f}^{\boldsymbol{\vartheta}}(\theta) \mathbb{f}_\theta(x).$$

The marginal distribution $\mathbb{P}^X = \mathbb{P}^{\boldsymbol{\vartheta}} \mathbb{P}_\bullet \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{X})$ of X is dominated by $\mu \in \mathfrak{M}_\sigma(\mathscr{X})$ with *μ-density*

$$\mathbb{f}^X \in \mathcal{M}_{\geqslant 0}(\mathscr{X}) \text{ with } x \mapsto \mathbb{f}^X(x) = \nu(\mathbb{f}_\bullet(x) \mathbb{f}^{\boldsymbol{\vartheta}}) = \int \mathbb{f}_\bullet(x) \mathbb{f}^{\boldsymbol{\vartheta}} \, \mathrm{d}\nu = \int_\Theta \mathbb{f}_\theta(x) \mathbb{f}^{\boldsymbol{\vartheta}}(\theta) \nu(\mathrm{d}\theta).$$

The $(\mathscr{X} \otimes \mathscr{T})$-$\mathscr{B}_{\geqslant 0}$-measurable function given by

$$\boxed{\mathbb{f}^{\boldsymbol{\vartheta}|X}} : \mathcal{X} \times \Theta \to \mathbb{R}_{\geqslant 0} \text{ with}$$

$$(x, \theta) \mapsto \mathbb{f}^{\boldsymbol{\vartheta}|X}(x, \theta) := \frac{\mathbb{f}_\theta(x) \mathbb{f}^{\boldsymbol{\vartheta}}(\theta)}{\mathbb{f}^X(x)} \mathbb{1}_{\{\mathbb{f}^X(x) \in \mathbb{R}_{>0}\}} + \mathbb{f}^{\boldsymbol{\vartheta}}(\theta) \mathbb{1}_{\{\mathbb{f}^X(x) = 0\}} \qquad (08.01)$$

is a *ν-density* of the Markov kernel $\mathbb{P}^{\boldsymbol{\vartheta}|X} := \mathbb{f}^{\boldsymbol{\vartheta}|X} \nu$ from $(\mathcal{X}, \mathscr{X})$ to $(\Theta, \mathscr{T})$, that is,

$$\mathbb{P}^{\boldsymbol{\vartheta}|X} = \mathbb{f}^{\boldsymbol{\vartheta}|X} \nu : \mathcal{X} \times \mathscr{T} \to [0,1] \text{ with } (x, F) \mapsto \mathbb{P}^{\boldsymbol{\vartheta}|X=x}(F) = \mathbb{f}^{\boldsymbol{\vartheta}|X=x} \nu(F) = \nu(\mathbb{1}_F \mathbb{f}^{\boldsymbol{\vartheta}|X=x}),$$

where $\mathbb{f}^{\boldsymbol{\vartheta}|X=x} \in \mathcal{M}_{\geqslant 0}(\mathscr{T})$ denotes a x-slice of $\mathbb{f}^{\boldsymbol{\vartheta}|X} \in \mathcal{M}_{\geqslant 0}(\mathscr{X} \otimes \mathscr{T})$ (see Theorem §07.17 (i)). □

§08.06 **Corollary**. *Let $(\mathcal{X}, \mathscr{X}, \mathbb{P}_\Theta)$ be a statistical experiment, $(\Theta, \mathscr{T}, \nu)$ a σ-finite measure space, and $\mathbb{P}_\bullet$ a Markov kernel from $(\Theta, \mathscr{T})$ to $(\mathcal{X}, \mathscr{X})$ with μ-density $\mathbb{f}_\bullet$ with respect to a σ-finite measure $\mu \in \mathfrak{M}_\sigma(\mathscr{X})$. Consider a random variable $(X, \boldsymbol{\vartheta})$ with values in $(\mathcal{X} \times \Theta, \mathscr{X} \otimes \mathscr{T})$. Let the a-priori distribution $\mathbb{P}^{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}$ admit a ν-density $\mathbb{f}^{\boldsymbol{\vartheta}}$ and let $\mathbb{P}^{X, \boldsymbol{\vartheta}} := \mathbb{P}^{\boldsymbol{\vartheta}} \odot \mathbb{P}_\bullet$ be the joint distribution of $(X, \boldsymbol{\vartheta})$. The Markov kernel $\mathbb{P}^{\boldsymbol{\vartheta}|X}$ from $(\mathcal{X}, \mathscr{X})$ to $(\Theta, \mathscr{T})$ with ν-density $\mathbb{f}^{\boldsymbol{\vartheta}|X}$ given in* (08.01) *is then a regular version of the conditional distribution of $\mathbb{P}^{\boldsymbol{\vartheta}}$ given X, i.e an a-posteriori distribution of $\boldsymbol{\vartheta}$.*

§08.07 **Proof** of Corollary §08.06. Follows directly from Theorem §07.17 (i). □

End LE12 2026/05/27

Chapter 3

Stochastic processes and stopping times

§09 Stochastic processes

In the sequel let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space, $(\mathcal{X}, \mathscr{X})$ a measurable space and $\mathcal{T}$ be a nonempty index set.

§09.01 **Definition**. A family $(X_t)_{t\in\mathcal{T}}$ of random variables on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in $(\mathcal{X}, \mathscr{X})$ is called *stochastic process* with time domain $\mathcal{T}$ and state space $\mathcal{X}$. A stochastic process $(X_t)_{t\in\mathcal{T}}$ is called *time-discrete* (or *time series*) if $\mathcal{T} \subseteq \mathbb{Z}$ and *time-continuous* if $(a, b) \subseteq \mathcal{T} \subseteq \overline{\mathbb{R}}$ for numerical numbers $a \in \overline{\mathbb{R}}$ and $b \in \overline{\mathbb{R}}_{>a}$. For each fixed $\omega \in \Omega$ the family $(X_t(\omega))_{t\in\mathcal{T}}$ is called *sample path* (Pfad), *trajectory* (Trajektorie) or *realisation* (Realisierung) of $(X_t)_{t\in\mathcal{T}}$. If $\mathcal{T} = \mathbb{Z}_{\geqslant 0}$ or $\mathcal{T} = \mathbb{R}_{\geqslant 0}$ the law of X_0 is called *initial distribution*. □

§09.02 **Example**. Let $(Y_j)_{j\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) Rademacher random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, that is Y_1 takes values in $\{-1, 1\}$ with $\mathbb{P}(Y_1 = -1) = 1/2 = \mathbb{P}(Y_1 = 1)$. Set $X_0 := 0$ and $X_n := \sum_{j\in[\![n]\!]} Y_j$ for $n \in \mathbb{N}$. The time-discrete stochastic process $(X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ with state space $(\mathbb{Z}, 2^{\mathbb{Z}})$ is called *symmetric simple random walk* on $\mathbb{Z}$. □

§09.03 **Remark**. Since we are interested in the behavior of the *family of random variables* $(X_t)_{t\in\mathcal{T}}$ and not a single random variable X_t, we consider $X_{\mathcal{T}} := (X_t)_{t\in\mathcal{T}}$ as a random variable taking values in the product space $(\mathcal{X}^{\mathcal{T}}, \mathscr{X}^{\mathcal{T}})$ (compare Definition §04.02). Then the map

$$X_{\mathcal{T}} : \Omega \ni \omega \mapsto X_{\mathcal{T}}(\omega) = (X_t(\omega))_{t\in\mathcal{T}} \in \mathcal{X}^{\mathcal{T}}$$

is $\mathscr{A}$-$\mathscr{X}^{\mathcal{T}}$-measurable, i.e. $X_{\mathcal{T}} \in \mathcal{M}(\mathscr{A}, \mathscr{X}^{\mathcal{T}})$. We denote by $\mathbb{P}^{X_{\mathcal{T}}} := \mathbb{P} \circ X_{\mathcal{T}}^{-1}$ the distribution of $X_{\mathcal{T}}$, i.e., the image measure of $\mathbb{P}$ under $X_{\mathcal{T}}$ on $(\mathcal{X}^{\mathcal{T}}, \mathscr{X}^{\mathcal{T}})$. The questions of interest to us in the following will depend only on this distribution, so it is not a limitation to assume that $(\mathcal{X}^{\mathcal{T}}, \mathscr{X}^{\mathcal{T}}, \mathbb{P}^{X_{\mathcal{T}}})$ is the underlying probability space and that each component X_t of the stochastic process correspond exactly to the coordinate map Π_t (see Reminder §04.01 and Notation §04.30). If $X_{\mathcal{T}} = (X_t)_{t\in\mathcal{T}}$ is a stochastic process on an abstract probability space $(\Omega, \mathscr{A}, \mathbb{P})$, then the *canonical proces* $\Pi_{\mathcal{T}} = (\Pi_t)_{t\in\mathcal{T}}$ of coordinate maps $\Pi_t : \mathcal{X}^{\mathcal{T}} \to \mathcal{X}$ form a stochastic process defined on $(\mathcal{X}^{\mathcal{T}}, \mathscr{X}^{\mathcal{T}}, \mathbb{P}^{X_{\mathcal{T}}})$ that has the same distribution as $X_{\mathcal{T}}$. A characterisation of the distribution of a stochastic process is often complicated and it is generally not given by an explicit formula. Therefore, we are interested in characterizing it using the distribution of the individual random variables X_t. □

§09.04 **Definition**. Let $X_{\mathcal{T}} = (X_t)_{t\in\mathcal{T}}$ be a stochastic process with distribution $\mathbb{P}^{X_{\mathcal{T}}}$ on $(\mathcal{X}^{\mathcal{T}}, \mathscr{X}^{\mathcal{T}})$. For each finite $\mathcal{J} \subseteq \mathcal{T}$ let $\mathbb{P}^{X_{\mathcal{J}}} = \mathbb{P}^{X_{\mathcal{T}}}_{\mathcal{J}} := \mathbb{P}^{X_{\mathcal{T}}} \circ \Pi_{\mathcal{J}}^{-1}$ be the joint distribution of the random variables $X_{\mathcal{J}} = (X_j)_{j\in\mathcal{J}} = \Pi_{\mathcal{J}} \circ X_{\mathcal{T}}$. The family of probability measures $\{\mathbb{P}^{X_{\mathcal{T}}}_{\mathcal{J}} = \mathbb{P}^{X_{\mathcal{J}}} \in \mathfrak{M}_{\mathrm{pr}}(\mathcal{X}^{\mathcal{J}}) \colon \mathcal{J} \subseteq \mathcal{T} \text{ finite}\}$ is called the family of the *finite-dimensional distributions* of $X_{\mathcal{T}}$ or $\mathbb{P}^{X_{\mathcal{T}}}$. □

§09.05 **Remark**. Let $\mathbb{P}^{X_{\mathcal{T}}} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{X}^{\mathcal{T}})$ be the distribution of a stochastic process $X_{\mathcal{T}}$ with values in $(\mathcal{X}^{\mathcal{T}}, \mathscr{X}^{\mathcal{T}})$ then the family of finite-dimensional distributions $\{\mathbb{P}^{X_{\mathcal{T}}}_{\mathcal{J}} = \mathbb{P}^{X_{\mathcal{J}}} \in \mathfrak{M}_{\mathrm{pr}}(\mathcal{X}^{\mathcal{J}}) \colon \mathcal{J} \subseteq \mathcal{T} \text{ finite}\}$

is a projective family (see Definition §04.31). Indeed, we have $\mathbb{P}^{X_\mathcal{T}}_\mathcal{J} = \mathbb{P}^{X_\mathcal{T}} \circ \Pi_\mathcal{J}^{-1} = \mathbb{P}^{X_\mathcal{T}} \circ (\Pi^\mathcal{K}_\mathcal{J} \circ \Pi_\mathcal{K})^{-1} = \mathbb{P}^{X_\mathcal{T}} \circ \Pi_\mathcal{K}^{-1} \circ (\Pi^\mathcal{K}_\mathcal{J})^{-1} = \mathbb{P}^{X_\mathcal{T}}_\mathcal{K} \circ (\Pi^\mathcal{K}_\mathcal{J})^{-1}$ for all finite $\mathcal{J} \subseteq \mathcal{K} \subseteq \mathcal{T}$. On the other hand, using *Kolmogorov's Extension Theorem* §04.33, the family of finite-dimensional distributions uniquely determines the probability measure $\mathbb{P}^{X_\mathcal{T}}$ when $\mathcal{X}$ s a Polish space endowed with the Borel σ-algebra $\mathscr{X} = \mathscr{B}_\mathcal{X}$. □

§09.06 **Definition**. Two stochastic processes $X_\mathcal{T} = (X_t)_{t\in\mathcal{T}}$ and $Y_\mathcal{T} = (Y_t)_{t\in\mathcal{T}}$ on $(\Omega, \mathscr{A}, \mathbb{P})$ with time domain $\mathcal{T}$ are called

indistinguishable (ununterscheidbar), if there is a $\mathbb{P}$-null set $N \in \mathscr{A}$ such that $\{X_t \neq Y_t\} \subseteq N$ for all $t \in \mathcal{T}$,

modifications (Versionen, Modifikationen) of each other, if $X_t = Y_t$ $\mathbb{P}$-a.e. for all $t \in \mathcal{T}$. □

§09.07 **Remark**.

(a) Obviously, indistinguishable processes are modifications of each other. The converse is in general false.

(b) If $X_\mathcal{T}$ and $Y_\mathcal{T}$ are modifications of each other, then $X_\mathcal{T}$ and $Y_\mathcal{T}$ share the same finite-dimensional distributions. Processes with the same finite-dimensional distributions need not even be defined on the same probability space and will in general not be modifications of each other.

(c) Let $X_\mathcal{T}$ and $Y_\mathcal{T}$ be modifications of each other. If the time domain $\mathcal{T}$ is countable, then $X_\mathcal{T}$ and $Y_\mathcal{T}$ are indistinguishable since the countable union of the $\mathbb{P}$-null sets $\{X_t \neq Y_t\}$, $t \in \mathcal{T}$, is still a $\mathbb{P}$-null set. □

§09.08 **Definition**. A stochastic process $X_\mathcal{T} = (X_t)_{t\in\mathcal{T}}$ with $\mathcal{T} \subseteq \mathbb{R}$ and state space $\mathcal{X}$ is called

real (*numerical*, *positive real*, etc.), if each X_t is real (numerical, positive real, etc.),

process with *independent increments*, if $X_\mathcal{T}$ is real and $\perp\!\!\!\perp_{j\in[\![n]\!]}(X_{t_j} - X_{t_{j-1}})$ (independent) for every $t_0 \in \mathcal{T}$, $t_j \in \mathcal{T}_{>t_{j-1}}$, $j \in [\![n]\!]$, and $n \in \mathbb{N}$,

Gaussian process, if $X_\mathcal{T}$ is real ist and $X_\mathcal{J} = (X_j)_{j\in\mathcal{J}}$ is normally distributed for every finite $\mathcal{J} \subseteq \mathcal{T}$,

($\mathcal{L}^s$-)integrable, if each X_t is ($\mathcal{L}^s$-)integrable.

Let in addition $\mathcal{T} \subseteq \mathbb{R}$ be closed under addition (e.g. $\mathcal{T} = \mathbb{Z}_{\geqslant 0}$). Define for $s \in \mathcal{T}$ the *shift* (-operator) $\vartheta_s : \mathcal{X}^\mathcal{T} \ni x_\mathcal{T} = (x_t)_{t\in\mathcal{T}} \mapsto \vartheta_s(x_\mathcal{T}) := x_{\mathcal{T}+s} := (x_t)_{t\in\mathcal{T}+s} = (x_{t+s})_{t\in\mathcal{T}} \in \mathcal{X}^\mathcal{T}$. Then $X_\mathcal{T}$ is called

stationary, if $X_\mathcal{T}$ and $X_{\mathcal{T}+s} := \vartheta_s(X_\mathcal{T})$ are identically distributed, symbolically $X_\mathcal{T} \stackrel{d}{=} X_{\mathcal{T}+s}$, for every $s \in \mathcal{T}$,

process with *stationary increments*, if $X_\mathcal{T}$ is real and $(\vartheta_t(X_\mathcal{T}))_{s+r} - (\vartheta_t(X_\mathcal{T}))_r \stackrel{d}{=} X_{s+r} - X_r$ for all $t, s, r \in \mathcal{T}$. (If $0 \in \mathcal{T}$, then it suffices to consider $r = 0$.) □

§09.09 **Example**.

(a) If $\{X_t : t \in \mathcal{T}\}$ are i.i.d. random variables, then $X_\mathcal{T} = (X_t)_{t\in\mathcal{T}}$ is stationary.

(b) Let $X_\mathbb{Z} = (X_z)_{z\in\mathbb{Z}}$ be real and stationary, and let $k \in \mathbb{N}$ and $a_i \in \mathbb{R}$, $i \in [\![0,k]\!]$. Setting $Y_z := \sum_{i\in[\![0,k]\!]} a_i X_{z-i}$ for $z \in \mathbb{Z}$ the real process $Y_\mathbb{Z} = (Y_z)_{z\in\mathbb{Z}}$ is stationary. If in addition $a_i \in \mathbb{R}_{\geqslant 0}$, $i \in [\![0,k]\!]$ and $\sum_{i\in[\![0,k]\!]} a_i = 1$ then $Y_\mathbb{Z}$ is called moving average of $X_\mathcal{T}$ (with weights $(a_i)_{i\in[\![0,k]\!]}$). □

§10 Stopping times

In the sequel let $\emptyset \neq \mathcal{T} \subseteq \mathbb{R}$ be a nonempty real time domain and $X_{\mathcal{T}} = (X_t)_{t\in\mathcal{T}}$ be a stochastic process on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with state space $(\mathcal{X}, \mathscr{X})$ and image probability measure $\mathbb{P}^{X_{\mathcal{T}}} \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{X}^{\mathcal{T}})$ on $(\mathcal{X}^{\mathcal{T}}, \mathscr{X}^{\mathcal{T}})$.

§10.01 **Definition**. A family $\mathscr{F}_{\mathcal{T}} = (\mathscr{F}_t)_{t\in\mathcal{T}}$ of σ-algebras with $\mathscr{F}_t \subseteq \mathscr{A}$ for each $t \in \mathcal{T}$ is called *filtration* if $\mathscr{F}_s \subseteq \mathscr{F}_t$ for all $s \in \mathcal{T}$ and $t \in \mathcal{T}_{\geqslant s}$. In this case $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ is called *filtered probability space*. □

Here and subsequently, we denote by $\sup\{\mathcal{T}\} \in \overline{\mathbb{R}}$ an arbitrary element $\sup\{\mathcal{T}\} \in \overline{\mathbb{R}}\backslash\mathcal{T}$ satisfying $t \leqslant \sup\{\mathcal{T}\}$ for all $t \in \mathcal{T}$. For example, $\sup\{\mathcal{T}\} = \infty$ is always possible.

§10.02 **Definition**. A stochastic process $X_{\mathcal{T}} = (X_t)_{t\in\mathcal{T}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ with state space $(\mathcal{X}, \mathscr{X})$ is called *adapted* to the filtration $\mathscr{F}_{\mathcal{T}} = (\mathscr{F}_t)_{t\in\mathcal{T}}$, or *$\mathscr{F}_{\mathcal{T}}$-adapted* for short, if $X_t \in \mathcal{M}(\mathscr{F}_t, \mathscr{X})$ is $\mathscr{F}_t$-$\mathscr{X}$-measurable, i.e. $\sigma(X_t) \subseteq \mathscr{F}_t$, for every $t \in \mathcal{T}$. If $\sigma_t(X_{\mathcal{T}}) := \bigvee_{s\in\mathcal{T}_{\leqslant t}} \sigma(X_s) = \sigma(\{X_s : s \in \mathcal{T}_{\leqslant t}\})$ for all $t \in \mathcal{T}$, then we call $\sigma_{\mathcal{T}}(X_{\mathcal{T}}) = (\sigma_t(X_{\mathcal{T}}))_{t\in\mathcal{T}}$ *natural* filtration or filtration *generated by* $X_{\mathcal{T}}$. Set $\mathscr{F}_{\sup\{\mathcal{T}\}} := \bigvee_{t\in\mathcal{T}} \mathscr{F}_t \subseteq \mathscr{A}$ and $\sigma_{\sup\{\mathcal{T}\}}(X_{\mathcal{T}}) := \bigvee_{t\in\mathcal{T}} \sigma(X_t) = \sigma(\{X_t : t \in \mathcal{T}\}) \subseteq \mathscr{A}$. □

§10.03 **Remark**. Clearly, a stochastic process is always adapted to the natural filtration it generates. The natural filtration is the smallest filtration to which the process is adapted. We write shortly, $\sigma_{\mathcal{T}}(X_{\mathcal{T}}) \subseteq \mathscr{F}_{\mathcal{T}}$, if $\sigma_t(X_{\mathcal{T}}) \subseteq \mathscr{F}_t$ for all $t \in \mathcal{T}$. □

§10.04 **Example**. Let $Y_{\mathbb{N}} = (Y_n)_{n\in\mathbb{N}}$ be a time series of real random variables generating the filtration $\sigma_{\mathbb{N}}(Y_{\mathcal{T}}) = (\sigma_n(Y_{\mathcal{T}}) = \sigma(\{Y_j : j \in [\![n]\!]\}))_{n\in\mathbb{N}}$, and let $X_{\mathbb{N}} = (X_n := \sum_{j\in[\![n]\!]} Y_j)_{n\in\mathbb{N}}$ be the time series of partial sums. Then $X_{\mathbb{N}}$ is adapted to $\sigma_{\mathbb{N}}(Y_{\mathbb{N}})$, that is, $\sigma_{\mathbb{N}}(X_{\mathbb{N}}) \subseteq \sigma_{\mathbb{N}}(Y_{\mathbb{N}})$. Clearly, $(Y_j)_{j\in[\![n]\!]}$ is also measurable with respect to $\sigma_n(X_{\mathcal{T}}) = \sigma(\{X_j : j \in [\![n]\!]\})$, and hence $\sigma_{\mathbb{N}}(Y_{\mathbb{N}}) \subseteq \sigma_{\mathbb{N}}(X_{\mathbb{N}})$. Consequently, $\sigma_{\mathbb{N}}(Y_{\mathbb{N}}) = \sigma_{\mathbb{N}}(X_{\mathbb{N}})$. Let $Z_{\mathbb{N}} = (Z_n := \sum_{j\in[\![n]\!]} \mathbb{1}_{(0,1)}(Y_j))_{n\in\mathbb{N}}$. Then $Z_{\mathbb{N}}$ is adapted to $\sigma_{\mathbb{N}}(Y_{\mathbb{N}})$ where in general $\sigma_{\mathbb{N}}(Z_{\mathbb{N}}) \subsetneqq \sigma_{\mathbb{N}}(Y_{\mathbb{N}})$. □

§10.05 **Definition**. Let $\mathcal{T} \in \{\mathbb{Z}_{\geqslant 0}, \mathbb{N}\}$. A stochastic process $X_{\mathcal{T}} = (X_n)_{n\in\mathcal{T}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ with state space $(\mathcal{X}, \mathscr{X})$ is called *predictable* (vorhersagbar) with respect to the filtration $\mathscr{F}_{\mathcal{T}}$, or *$\mathscr{F}_{\mathcal{T}}$-predictable* for short, if X_0 is constant (in case $0 \in \mathcal{T}$) and $\vartheta_1(X_{\mathcal{T}})$ is $\mathscr{F}_{\mathcal{T}}$-adapted, that is, $X_{n+1} \in \mathcal{M}(\mathscr{F}_n, \mathscr{X})$ is $\mathscr{F}_n$-$\mathscr{X}$-measurable for every $n \in \mathcal{T}$. A numerical, $\mathscr{F}_{\mathcal{T}}$-predictable process $X_{\mathcal{T}}$ on $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ is called *increasing* if $X_n \leqslant X_{n+1}$ $\mathbb{P}$-a.e. for all $n \in \mathcal{T}$. $X_{\mathcal{T}}$ is called *decreasing* if $-X_{\mathcal{T}}$ is increasing. □

§10.06 **Definition**. A numerical random variable $\tau \in \overline{\mathcal{M}}(\mathscr{A})$ with values in

$$\mathcal{T} \cup \{\sup\{\mathcal{T}\}\} =: \overline{\mathcal{T}} \subseteq \overline{\mathbb{R}}$$

is called *stopping time* with respect to the filtration $\mathscr{F}_{\mathcal{T}}$, or *$\mathscr{F}_{\mathcal{T}}$-stopping time* for short, if the process $(\mathbb{1}_{\{\tau\leqslant t\}})_{t\in\mathcal{T}}$ is $\mathscr{F}_{\mathcal{T}}$-adapted, that is, $\{\tau \leqslant t\} \in \mathscr{F}_t$ for every $t \in \mathcal{T}$. A stopping time is called *finite* if it takes values in $\mathcal{T}$ only. □

§10.07 **Remark**. Clearly, $\{\tau > t\} = \{\tau \leqslant t\}^c \in \mathscr{F}_t$ for all $t \in \mathcal{T}$. On the event $\{\tau \leqslant t\}$ the stopping time τ has already occurred by the time t, and in contrast, on the event $\{\tau > t\}$, it will occur later. Since there exists $(t_n)_{n\in\mathbb{N}}$ in $\mathcal{T}$ with $t_n \uparrow \sup\{\mathcal{T}\}$, it follows that

$$\{\tau \in \mathcal{T}\} = \bigcup_{n\in\mathbb{N}} \{\tau \leqslant t_n\} \in \mathscr{F}_{\sup\{\mathcal{T}\}}$$

and also $\{\tau = \sup\{\mathcal{T}\}\} = \{\tau \in \mathcal{T}\}^c \in \mathscr{F}_{\sup\{\mathcal{T}\}}$. Therefore, the stopping time τ is also $\mathscr{F}_{\sup\{\mathcal{T}\}}$-$\overline{\mathscr{B}}$-measurable, i.e. $\tau \in \mathcal{M}(\mathscr{F}_{\sup\{\mathcal{T}\}}, \overline{\mathscr{B}})$. □

§10.08 **Lemma**. *Let $\mathcal{T}$ be countable and $\tau \in \overline{\mathcal{M}}(\mathscr{A})$ with values in $\overline{\mathcal{T}}$. Then τ is a $\mathscr{F}_{\mathcal{T}}$-stopping time if and only if the process $(\mathbb{1}_{\{\tau=t\}})_{t\in\mathcal{T}}$ is $\mathscr{F}_{\mathcal{T}}$-adapted, that is, $\{\tau = t\} \in \mathscr{F}_t$ for all $t \in \mathcal{T}$.*

§10.09 **Proof** of Lemma §10.08. Exercise. □

§10.10 **Example**.

(a) Let $t_\circ \in \overline{\mathcal{T}}$, then $\tau = t_\circ$ (constant) is a $\mathscr{F}_{\mathcal{T}}$-stopping time. Indeed, for all $t \in \mathcal{T}$ either $\{\tau \leqslant t\} = \emptyset \in \mathscr{F}_t$ if $t \in \mathcal{T}_{<t_\circ}$, or $\{\tau \leqslant t\} = \Omega \in \mathscr{F}_t$ if $t \in \mathcal{T}_{\geqslant t_\circ}$.

(b) Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a stochastic process with state space $(\mathcal{X}, \mathscr{X})$ which is $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-adapted. For $B \in \mathscr{X}$ consider the hitting time (Treffzeit), the first time $X_{\mathbb{Z}_{\geqslant 0}}$ is in B, that is, for $\omega \in \Omega$

$$\tau_B(\omega) := \begin{cases} \min\{n \in \mathbb{Z}_{\geqslant 0} : X_n(\omega) \in B\}, & \text{if } \omega \in \bigcup_{n\in\mathbb{Z}_{\geqslant 0}} X_n^{-1}(B), \\ \infty = \sup\{\mathbb{Z}_{\geqslant 0}\}, & \text{otherwise,} \end{cases}$$

where $\tau_\emptyset = \infty$ and $\tau_{\mathcal{X}} = 0$ are constant. Then τ_B is a $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-stopping time. Indeed, since $X_n^{-1}(B) \in \mathscr{F}_n \subseteq \mathscr{F}_t$ for all $n \in [\![0, t]\!]$ we have $\{\tau_B \leqslant t\} = \bigcup_{n\in[\![0,t]\!]} X_n^{-1}(B) \in \mathscr{F}_t$ for all $t \in \mathbb{Z}_{\geqslant 0} = \mathcal{T}$. □

§10.11 **Lemma**. *Let τ and σ be stopping times. Then:*

(i) *$\tau \vee \sigma$ and $\tau \wedge \sigma$ are stopping times.*

Let $\mathcal{T} \subseteq \mathbb{R}_{\geqslant 0}$ be closed under addition. Then:

(ii) *$\tau + \sigma$ is a stopping time.*

(iii) *$\tau + s$ is a stopping time for $s \in \mathcal{T}$, but in general $\tau - s$ is not.*

§10.12 **Proof** of Lemma §10.11. Exercise. □

§10.13 **Remark**. We expect the properties (i) and (ii) in Lemma §10.11 from stopping times. For (iii), it should be noted that $\tau - s$ peeks s time units into the future (since $\{\tau - s \leqslant t\} \in \mathscr{F}_{t+s}$), while $\tau + s$ looks s time units into the past. However, stopping times are only allowed to look into the past. □

§10.14 **Example** *(§10.10 (b) continued)*. For $A, B \in \mathscr{X}$ consider the hitting times τ_A and τ_B. Then $\tau_B \leqslant \tau_A$ for $A \subseteq B$. In particular, this implies $\tau_{A\cup B} \leqslant \tau_A \wedge \tau_B$ and $\tau_A \vee \tau_B \leqslant \tau_{A\cap B}$. □

§10.15 **Definition**. Let τ be a $\mathscr{F}_{\mathcal{T}}$-stopping time on $(\Omega, \mathscr{A})$. Then we call

$$\mathscr{F}_\tau := \{A \in \mathscr{F}_{\sup\{\mathcal{T}\}} : A \cap \{\tau \leqslant t\} \in \mathscr{F}_t \text{ for all } t \in \mathcal{T}\}$$

the *σ-algebra of τ-past*. □

§10.16 **Remark**. Clearly, $\mathscr{F}_\tau \subseteq \mathscr{F}_{\sup\{\mathcal{T}\}}$ is closed under complements and σ-$\cup$-closed with $\Omega \in \mathscr{F}_\tau$. Consequently, $\mathscr{F}_\tau$ is a σ-algebra. Since $\{\tau \leqslant \sup\{\mathcal{T}\}\} = \Omega$, for all $A \in \mathscr{F}_{\sup\{\mathcal{T}\}}$ follows immediately $A \cap \{\tau \leqslant \sup\{\mathcal{T}\}\} = A \in \mathscr{F}_{\sup\{\mathcal{T}\}}$, and hence also the seemingly stronger identity

$$\mathscr{F}_\tau = \{A \in \mathscr{F}_{\sup\{\mathcal{T}\}} : A \cap \{\tau \leqslant t\} \in \mathscr{F}_t \text{ for all } t \in \overline{\mathcal{T}}\}.$$

Furthermore, since $\{\tau \in \mathcal{T}\} \in \mathscr{F}_{\sup\{\mathcal{T}\}}$ (compare Remark §10.07) and $\{\tau \in \mathcal{T}\} \cap \{\tau \leqslant t\} = \{\tau \leqslant t\} \in \mathscr{F}_t$ for all $t \in \mathcal{T}$, it follows $\{\tau \in \mathcal{T}\} \in \mathscr{F}_\tau$, and hence

$$\{\tau = \sup\{\mathcal{T}\}\} = \{\tau \in \mathcal{T}\}^c \in \mathscr{F}_\tau.$$

Finally, the σ-algebra of τ-past satisfies

$$\mathscr{F}_\tau = \{A \in \mathscr{F}_{\sup\{\mathcal{T}\}} : A \cap \{\tau = t\} \in \mathscr{F}_t \text{ for all } t \in \mathcal{T}\}$$

if $\mathcal{T}$ is countable. □

§10.17 **Example**. For a constant stopping time $\tau = t_\circ$ with $t_\circ \in \overline{\mathcal{T}}$ follows $\mathscr{F}_\tau = \mathscr{F}_{t_\circ}$. Indeed, it holds $\mathscr{F}_\tau = \mathscr{F}_\tau \cap \{\tau \leqslant t_\circ\} \subseteq \mathscr{F}_{t_\circ}$. On the other hand, for each $A \in \mathscr{F}_{t_\circ}$ and $t \in \mathcal{T}$ follows $A \cap \{\tau \leqslant t\} = A \in \mathscr{F}_{t_\circ} \subseteq \mathscr{F}_t$ for $t_\circ \leqslant t$ and $A \cap \{\tau \leqslant t\} = \varnothing \in \mathscr{F}_t$ for $t < t_\circ$. Consequently, $A \in \mathscr{F}_\tau$ and hence also $\mathscr{F}_{t_\circ} \subseteq \mathscr{F}_\tau$. □

§10.18 **Lemma**. *Let τ and σ be $\mathscr{F}_\mathcal{T}$-stopping times. Then:*

(i) $\mathscr{F}_\sigma \cap \{\sigma \leqslant \tau\} \subseteq \mathscr{F}_{\tau\wedge\sigma} = \mathscr{F}_\tau \cap \mathscr{F}_\sigma$,

(ii) *τ is $\mathscr{F}_\tau$-measurable,*

(iii) *$\mathscr{F}_\tau = \mathscr{F}_t$ on $\{\tau = t\}$, i.e., $\mathscr{F}_\tau \cap \{\tau = t\} = \mathscr{F}_t \cap \{\tau = t\}$, for all $t \in \overline{\mathcal{T}}$, and*

(iv) $\mathscr{F}_{\tau\vee\sigma} = \mathscr{F}_\tau \vee \mathscr{F}_\sigma$.

From (i) follows $\{\sigma \leqslant \tau\} \in \mathscr{F}_\sigma \cap \mathscr{F}_\tau$, $\mathscr{F}_\sigma = \mathscr{F}_\tau$ on $\{\sigma = \tau\}$, and $\mathscr{F}_\sigma \subseteq \mathscr{F}_\tau$ for $\sigma \leqslant \tau$.

§10.19 **Proof** of Lemma §10.18. is given in the lecture. □

§10.20 **Definition**. For a stopping time τ and a process $X_\mathcal{T} = (X_t)_{t\in\mathcal{T}}$ on $(\Omega, \mathscr{A})$ define

$$X_\tau : \Omega \supseteq \{\tau \in \mathcal{T}\} \to \mathfrak{X} \quad \text{with} \quad \omega \mapsto X_\tau(\omega) := X_{\tau(\omega)}(\omega)$$

or in equal $X_\tau := X_t$ on $\{\tau = t\}$ for all $t \in \mathcal{T}$. □

§10.21 **Remark**. Keep in mind that a function $h : \Omega \supseteq B \to \mathfrak{X}$ with $B \in \mathscr{A}$ is $\mathscr{A}$-$\mathscr{X}$-measurable, if $h^{-1}(\mathscr{X}) \subseteq \mathscr{A}$ (compare Definition §04.42). If $\mathcal{T}$ is countable and τ is a $\mathscr{F}_\mathcal{T}$-stopping time, then $f : \Omega \supseteq \{\tau \in \mathcal{T}\} \to \mathfrak{X}$ is $\mathscr{F}_\tau$-$\mathscr{X}$-measurable, if for all $t \in \mathcal{T}$ the restriction of f on $\{\tau = t\}$ is $\mathscr{F}_t$-$\mathscr{X}$-measurable, that is $\{f \in B\} \cap \{\tau = t\} = f^{-1}(B) \cap \{\tau = t\} \in \mathscr{F}_t$ for all $B \in \mathscr{X}$. Indeed, then $f^{-1}(B) \cap \{\tau \leqslant t\} = \bigcup_{s\in\mathcal{T}:s\leqslant t} f^{-1}(B) \cap \{\tau = s\} \in \mathscr{F}_t$ for all $B \in \mathscr{X}$ and $t \in \mathcal{T}$. From Definition §10.15 follows thus $f^{-1}(\mathscr{X}) \subseteq \mathscr{F}_\tau$ or in other words f is $\mathscr{F}_\tau$-$\mathscr{X}$-measurable. □

§10.22 **Lemma**. *Let $\mathcal{T}$ be countable, let $X_\mathcal{T}$ be a $\mathscr{F}_\mathcal{T}$-adapted process and let τ be a $\mathscr{F}_\mathcal{T}$-stopping time. Then X_τ is $\mathscr{F}_\tau$-$\mathscr{X}$-measurable.*

§10.23 **Proof** of Lemma §10.22. is given in the lecture. □

§10.24 **Remark**. For uncountable $\mathcal{T}$ and fixed ω, the mapping $X_\mathcal{T}(\omega) : \mathcal{T} \to \mathfrak{X}$ with $t \mapsto X_t(\omega)$ is generally not measurable, thus the composition X_τ is also not always measurable. Under additional assumptions regarding the regularity of the paths $X_\mathcal{T}(\omega) : t \mapsto X_t(\omega)$, such as right-continuity, the measurability of the composition X_τ can be ensured (c.f. Kallenberg (2002, Lemma 7.5, S.122)). □

§10.25 **Corollary**. *Let $\mathcal{T}$ be countable, $X_\mathcal{T}$ a $\mathscr{F}_\mathcal{T}$-adapted process, and $(\tau_t)_{t\in\mathcal{T}}$ an isotone family of finite $\mathscr{F}_\mathcal{T}$-stopping times, that is, τ_t takes values in $\mathcal{T}$ and $\tau_t \leqslant \tau_s$ for all $t \in \mathcal{T}$ and $s \in \mathcal{T}_{\geqslant t}$. The stochastic process $(X_{\tau_t})_{t\in\mathcal{T}}$ is adapted to the filtration $(\mathscr{F}_{\tau_t})_{t\in\mathcal{T}}$. In particular, for any $\mathscr{F}_\mathcal{T}$-stopping time τ the process $(X_{\tau\wedge t})_{t\in\mathcal{T}}$ is adapted to both filtration $(\mathscr{F}_{\tau\wedge t})_{t\in\mathcal{T}}$ and $\mathscr{F}_\mathcal{T} = (\mathscr{F}_t)_{t\in\mathcal{T}}$.*

§10.26 **Proof** of Corollary §10.25. is given in the lecture. □

§10.27 **Definition**. Let $\mathcal{T}$ be countable time domain, $X_\mathcal{T} = (X_t)_{t\in\mathcal{T}}$ a $\mathscr{F}_\mathcal{T}$-adapted process, and τ a $\mathscr{F}_\mathcal{T}$-stopping time. We define the *stopped process* $X^\tau = (X^\tau_t := X_{\tau\wedge t})_{t\in\mathcal{T}}$, which is adapted to both filtrations $\mathscr{F}_\mathcal{T} = (\mathscr{F}_t)_{t\in\mathcal{T}}$ and $\mathscr{F}^\tau_\mathcal{T} = (\mathscr{F}^\tau_t := \mathscr{F}_{\tau\wedge t})_{t\in\mathcal{T}}$. □

End LE13
2026/05/29

Chapter 4

Martingale theory

In the following, let $\mathcal{T} \subseteq \mathbb{R}$ be a non-empty time domain and $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}} = (\mathscr{F}_t)_{t\in\mathcal{T}})$ a filtered probability space. For $Y \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ or $Y \in \bar{\mathcal{L}}^1(\mathbb{P})$ and a sub-σ-algebra $\mathscr{S} \subseteq \mathscr{A}$, $\mathbb{E}(Y|\mathscr{S})$ denotes, as before, a version of the conditional expectation of Y given $\mathscr{S}$.

§11 Positive (super-) martingale

§11.01 **Definition**. A positive (numerical) $\mathscr{F}_{\mathcal{T}}$-adapted stochastic process $X_{\mathcal{T}} = (X_t)_{t\in\mathcal{T}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ is called

positive ($\mathscr{F}_{\mathcal{T}}$-) supermartingale, if $X_s \geqslant \mathbb{E}(X_t|\mathscr{F}_s)$ $\mathbb{P}$-a.e. for all $s \in \mathcal{T}$ and $t \in \mathcal{T}_{>s}$,

positive ($\mathscr{F}_{\mathcal{T}}$-) martingale, if $X_s = \mathbb{E}(X_t|\mathscr{F}_s)$ $\mathbb{P}$-a.e. for all $s \in \mathcal{T}$ and $t \in \mathcal{T}_{>s}$.

A d-dimensional stochastic process $X_{\mathcal{T}} = ((X_t^1, \dots, X_t^d))_{t\in\mathcal{T}}$ on $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ is called *positive ($\mathscr{F}_{\mathcal{T}}$-) (super-) martingale* if each coordinate process $X_{\mathcal{T}}^k = (X_t^k)_{t\in\mathcal{T}}$ is a positive ($\mathscr{F}_{\mathcal{T}}$-) (super-) martingale. □

§11.02 **Remark**.

(a) Let $X_{\mathcal{T}} = (X_t)_{t\in\mathcal{T}}$ be a $\mathscr{F}_{\mathcal{T}}$-supermartingale, then

$$\mathbb{E}(X_r|\mathscr{F}_s) \geqslant \mathbb{E}(\mathbb{E}(X_t|\mathscr{F}_r)|\mathscr{F}_s) \geqslant \mathbb{E}(X_t|\mathscr{F}_s) \quad \mathbb{P}\text{-a.e.} \quad \text{for all } s \in \mathcal{T}, r \in \mathcal{T}_{>s}, t \in \mathcal{T}_{>r},$$

and hence $\mathbb{E}(X_r) \geqslant \mathbb{E}(X_t)$ and the map $\mathbb{E}(X_{\mathcal{T}}) := (\mathbb{E}(X_t))_{t\in\mathcal{T}}$ is monotonically decreasing, and constant in case of a $\mathscr{F}_{\mathcal{T}}$-martingale.

(b) In case $\mathcal{T} \in \{\mathbb{N}, \mathbb{Z}_{\geqslant 0}, \mathbb{Z}\}$, it is sufficient to consider at each instant s only $t = s + 1$, since

$$X_s \geqslant \mathbb{E}(X_{s+1}|\mathscr{F}_s) \geqslant \mathbb{E}(\mathbb{E}(X_{s+2}|\mathscr{F}_{s+1})|\mathscr{F}_s) = \mathbb{E}(X_{s+2}|\mathscr{F}_s) \quad \mathbb{P}\text{-a.e.} \quad \text{for all } s \in \mathcal{T}$$

by employing the tower property, and by induction $X_s \geqslant \mathbb{E}(X_{s+m}|\mathscr{F}_s)$ $\mathbb{P}$-a.e. for all $s \in \mathcal{T}$ and $m \in \mathbb{N}$.

(c) If we do not explicitly specify the filtration $\mathscr{F}_{\mathcal{T}}$, we implicitly assume that $\mathscr{F}_{\mathcal{T}} = \sigma_{\mathcal{T}}(X_{\mathcal{T}})$ is the filtration generated by $X_{\mathcal{T}}$.

(d) Let $\mathscr{F}_{\mathcal{T}}$ and $\mathscr{F}_{\mathcal{T}}^\circ$ be filtrations on $(\Omega, \mathscr{A}, \mathbb{P})$ with $\mathscr{F}_{\mathcal{T}} \subseteq \mathscr{F}_{\mathcal{T}}^\circ$, and let $X_{\mathcal{T}}$ be $\mathscr{F}_{\mathcal{T}}$-adapted (and hence also $\mathscr{F}_{\mathcal{T}}^\circ$-adapted). If $X_{\mathcal{T}}$ is a positive $\mathscr{F}_{\mathcal{T}}^\circ$- (super-) martingale, then it is also a positive $\mathscr{F}_{\mathcal{T}}$-(super-)martingale, since

$$X_s = \mathbb{E}(X_s|\mathscr{F}_s) \geqslant \mathbb{E}(\mathbb{E}(X_t|\mathscr{F}_s^\circ)|\mathscr{F}_s) = \mathbb{E}(X_t|\mathscr{F}_s) \quad \mathbb{P}\text{-a.e.} \quad \text{for all } s \in \mathcal{T}, t \in \mathcal{T}_{>s}.$$

Consequently, a positive $\mathscr{F}_{\mathcal{T}}$- (super-) martingale $X_{\mathcal{T}}$ is also a (super-) martingale with respect to its own natural filtration $\sigma_{\mathcal{T}}(X_{\mathcal{T}})$. □

§11.03 **Example.**

(a) Let $Y_{\mathbb{N}} = (Y_n)_{n\in\mathbb{N}}$ be a sequence of independent positive numerical random variables in $\overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ with $\mathbb{E}(Y_n) = 1$ for all $n \in \mathbb{N}$. Moreover let $\mathscr{F}_{\mathbb{N}} = \sigma_{\mathbb{N}}(Y_{\mathbb{N}})$ be the filtration generated by $Y_{\mathbb{N}}$ such that $\sigma_n(Y_{\mathbb{N}})$ and Y_{n+1} are independent. Then $(X_n := \prod_{i\in[\![n]\!]} Y_i)_{n\in\mathbb{N}}$ is a positive $\sigma_{\mathbb{N}}(Y_{\mathbb{N}})$-martingale, since

$$\mathbb{E}\big(X_{n+1}\big|\sigma_n(Y_{\mathbb{N}})\big) = \mathbb{E}\big(X_nY_{n+1}\big|\sigma_n(Y_{\mathbb{N}})\big) = X_n\mathbb{E}\big(Y_{n+1}\big|\sigma_n(Y_{\mathbb{N}})\big) = X_n\mathbb{E}(Y_{n+1}) = X_n \quad \mathbb{P}\text{-a.e.}$$

(b) Let $Z \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ be a positive numerical random variable and $\mathscr{F}_{\mathbb{N}} = (\mathscr{F}_n)_{n\in\mathbb{N}}$ a filtration. Then $(X_n := \mathbb{E}\big(Z\big|\mathscr{F}_n\big))_{n\in\mathbb{N}}$ is a positive numerical $\mathscr{F}_{\mathbb{N}}$-martingale since

$$\mathbb{E}\big(X_{n+1}\big|\mathscr{F}_n\big) = \mathbb{E}\big(\mathbb{E}\big(Z\big|\mathscr{F}_{n+1}\big)\big|\mathscr{F}_n\big) = \mathbb{E}\big(Z\big|\mathscr{F}_n\big) = X_n \quad \mathbb{P}\text{-a.e.}$$ □

§11.04 **Lemma.**

(i) *$aX_{\mathcal{T}} + bY_{\mathcal{T}} = (aX_t + bY_t)_{t\in\mathcal{T}}$ is a positive $\mathscr{F}_{\mathcal{T}}$-(super-)martingale, if $X_{\mathcal{T}}$ and $Y_{\mathcal{T}}$ are positive $\mathscr{F}_{\mathcal{T}}$-(super-)martingales, and $a, b \in \mathbb{R}_{\geqslant 0}$.*

(ii) *$X_{\mathcal{T}} \wedge Y_{\mathcal{T}} = (X_t \wedge Y_t)_{t\in\mathcal{T}}$ is a positive $\mathscr{F}_{\mathcal{T}}$-supermartingale, if $X_{\mathcal{T}}$ and $Y_{\mathcal{T}}$ are positive $\mathscr{F}_{\mathcal{T}}$-supermartingales.*

(iii) *If $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ is a positive $\mathscr{F}_{\mathbb{N}}$-supermartingale with $\mathbb{E}(X_k) \geqslant \mathbb{E}(X_1)$ for some $k \in \mathbb{N}$, then $X_{[\![k]\!]} = (X_n)_{n\in[\![k]\!]}$ is a positive $\mathscr{F}_{[\![k]\!]} = (\mathscr{F}_n)_{n\in[\![k]\!]}$-martingale. If there exists a sub-sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ with $k_n \uparrow \infty$ and $\mathbb{E}(X_{k_n}) \geqslant \mathbb{E}(X_1)$ for all $n \in \mathbb{N}$, then $X_{\mathbb{N}}$ is a positive $\mathscr{F}_{\mathbb{N}}$-martingales.*

(iv) *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ and $Y_{\mathbb{N}} = (Y_n)_{n\in\mathbb{N}}$ be positive $\mathscr{F}_{\mathbb{N}}$-supermartingale and τ a $\mathscr{F}_{\mathbb{N}}$-stopping time with $X_\tau(\omega) \geqslant Y_\tau(\omega)$ for all $\omega \in \{\tau \in \mathbb{N}\}$. Then $Z_{\mathbb{N}} = (Z_n := X_n\mathbb{1}_{\{n<\tau\}} + Y_n\mathbb{1}_{\{\tau\leqslant n\}})_{n\in\mathbb{N}}$ is a positive $\mathscr{F}_{\mathbb{N}}$-supermartingale.*

§11.05 **Proof** of Lemma §11.04. is given in the lecture. □

§11.06 **Proposition** *(Maximal inequality). Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a positive $\mathscr{F}_{\mathbb{N}}$-supermartingale. Then $\sup_{n\in\mathbb{N}} X_n$ is $\mathbb{P}$-a.e. finite on the event $\{X_1 \in \mathbb{R}_{\geqslant 0}\}$ and for every $a \in \mathbb{R}_{>0}$ holds*

$$\mathbb{P}\big(\sup_{n\in\mathbb{N}} X_n \geqslant a\big|\mathscr{F}_1\big) = \mathbb{E}\big(\mathbb{1}_{\{\sup_{n\in\mathbb{N}} X_n\geqslant a\}}\big|\mathscr{F}_1\big) \leqslant \min(X_1/a, 1) \quad \mathbb{P}\text{-a.e.}$$

§11.07 **Proof** of Proposition §11.06. is given in the lecture. □

§11.08 **Remark.** The last statement also holds if the constant $a \in \mathbb{R}_{>0}$ is replaced by a positive $\mathscr{F}_1$-measurable random variable $A \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}_1)$, i.e., it also holds

$$\mathbb{P}\big(\sup_{n\in\mathbb{N}} X_n \geqslant A\big|\mathscr{F}_1\big) = \mathbb{E}\big(\mathbb{1}_{\{\sup_{n\in\mathbb{N}} X_n\geqslant A\}}\big|\mathscr{F}_1\big) \leqslant \min(X_1/A, 1) \quad \mathbb{P}\text{-a.e. on } \{A \in \overline{\mathbb{R}}_{>0}\}.$$

In particular, it follows that:

(a) $1 = \mathbb{P}\big(\sup_{n\in\mathbb{N}} X_n \geqslant A\big|\mathscr{F}_1\big) \leqslant \min\big(X_1/A, 1\big)$ $\mathbb{P}$-a.e. on $\{A \in \overline{\mathbb{R}}_{>0}\}$ for every positive $\mathscr{F}_1$-measurable random variable $A \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}_1)$ with $A \leqslant \sup_{n\in\mathbb{N}} X_n$, and hence $A \leqslant X_1$ $\mathbb{P}$-a.e. In other words, X_1 is the largest $\mathscr{F}_1$-measurable lower bound of $\sup_{n\in\mathbb{N}} X_n$.

(b) More general, for every $k \in \mathbb{N}$, $\max_{n\in[\![k]\!]} X_n$ is the largest $\mathscr{F}_k$-measurable lower bound of $\sup_{n\in\mathbb{N}} X_n$. Indeed, $(\max_{n\in[\![k]\!]} X_n, X_{k+1}, X_{k+2}, \dots)$ is a supermartingale adapted to the filtration $(\mathscr{F}_{k+n})_{n\in\mathbb{Z}_{\geqslant 0}}$, and, from Proposition §11.06 follows $A \leqslant \max_{n\in[\![k]\!]} X_n$ $\mathbb{P}$-a.e. for every positive $\mathscr{F}_k$-measurable random variable $A \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}_k)$ with $A \leqslant \sup_{n\in\mathbb{N}} X_n$. □

§11.09 **Definition**. Let $x_{\mathbb{N}} = (x_n)_{n\in\mathbb{N}}$ be a sequence in $\overline{\mathbb{R}}$. For $a \in \mathbb{R}$ and $b \in \mathbb{R}_{>a}$ define by induction

$$\tau_0 := \tau_0(x_{\mathbb{N}}) := 1, \quad \sigma_k := \sigma_k(x_{\mathbb{N}}) := \inf\{n \in \mathbb{N}_{\geqslant \tau_{k-1}} : x_n \in \overline{\mathbb{R}}_{\leqslant a}\} \quad \text{and}$$
$$\tau_k := \tau_k(x_{\mathbb{N}}) := \inf\{n \in \mathbb{N}_{\geqslant \sigma_k} : x_n \in \overline{\mathbb{R}}_{\geqslant b}\}, \quad \forall\, k \in \mathbb{N}$$

(where $\inf \varnothing = \infty$). The number of upward crossings of the interval $[a, b]$ by the sequence $x_{\mathbb{N}}$ is denoted by $\beta_{a,b} := \beta_{a,b}(x_{\mathbb{N}}) := \sup\{k \in \mathbb{Z}_{\geqslant 0} : \tau_k(x_{\mathbb{N}}) \in \mathbb{N}\}$. □

§11.10 **Remark**. Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a numerical $\mathscr{F}_{\mathbb{N}}$-adapted stochastic process. Then $\tau_k := \tau_k(X_{\mathbb{N}})$ (and $\sigma_k := \sigma_k(X_{\mathbb{N}})$) defined in Definition §11.09 is a $\mathscr{F}_{\mathbb{N}}$-stopping time for every $k \in \mathbb{N}$. Indeed, for each $n \in \mathbb{N}$ $\{\tau_k = n\}$ (and also $\{\sigma_k = n\}$) is a $\mathscr{F}_n$-measurable event. In particular, $\tau_k \leqslant \tau_{k+1}$ holds for all $k \in \mathbb{N}$. □

§11.11 **Lemma**. *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a sequence of numerical random variables defined on $(\Omega, \mathscr{A})$, $\sigma_\infty(X_{\mathbb{N}}) = \sigma(\{X_t : t \in \mathbb{N}\}) \subseteq \mathscr{A}$, $a \in \mathbb{R}$ and $b \in \mathbb{R}_{>a}$. The random number $\beta_{a,b}(X_{\mathbb{N}}) : \Omega \to \overline{\mathbb{Z}}_{\geqslant 0} \subseteq \overline{\mathbb{R}}_{\geqslant 0}$ with $\omega \mapsto \beta_{a,b}(X_{\mathbb{N}}(\omega))$ of upward crossings of the interval $[a, b]$ is a $\sigma_\infty(X_{\mathbb{N}})$-$\overline{\mathscr{B}}_{\geqslant 0}$-measurable random variable.*

§11.12 **Proof** of Lemma §11.11. Exercise. □

§11.13 **Remark**. Evidently, $\beta_{a,b}(x_{\mathbb{N}}) = \infty$ if $\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n$. Conversely, from $\beta_{a,b}(x_{\mathbb{N}}) = \infty$ follows $\liminf_{n\to\infty} x_n \leqslant a < b \leqslant \limsup_{n\to\infty} x_n$. In other words, a sequence $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ in $\overline{\mathbb{R}}$ is convergent if and only if $\beta_{a,b}(x_{\mathbb{N}}) \in \mathbb{Z}_{\geqslant 0}$ for all $a \in \mathbb{Q}$ and $b \in \mathbb{Q}_{>a}$. □

§11.14 **Lemma**. *A sequence $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ of numerical random variables is convergent* $\mathbb{P}$-a.e. *if and only if the associated random number $\beta_{a,b}(X_{\mathbb{N}})$ of upward crossings of the interval $[a, b]$ is finite* $\mathbb{P}$-a.e., *i.e.,* $\mathbb{P}(\beta_{a,b}(X_{\mathbb{N}}) \in \mathbb{Z}_{\geqslant 0}) = 1$, *for all $a \in \mathbb{Q}$ and $b \in \mathbb{Q}_{>a}$.*

§11.15 **Proof** of Lemma §11.14. Exercise. □

§11.16 **Lemma** *(Dubin's inequality)*. *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a positive $\mathscr{F}_{\mathbb{N}}$-Supermartingale, $k \in \mathbb{N}$ and $a \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}_{>a}$. The associated random number $\beta_{a,b}(X_{\mathbb{N}})$ of upward crossings satisfies*

$$\mathbb{P}\left(\beta_{a,b}(X_{\mathbb{N}}) \geqslant k \middle| \mathscr{F}_1\right) \leqslant (a/b)^k \min\left(X_1/a, 1\right) \quad \mathbb{P}\text{-a.e.}.$$

In particular, $\beta_{a,b}(X_{\mathbb{N}})$ is finite $\mathbb{P}$-a.e., *i.e.,* $\mathbb{P}(\beta_{a,b}(X_{\mathbb{N}}) \in \mathbb{Z}_{\geqslant 0}) = 1$.

§11.17 **Proof** of Lemma §11.16. is given in the lecture. □

§11.18 **Remark**. Let $X_{\mathbb{Z}} = (X_z)_{z\in\mathbb{Z}}$ be a positive $\mathscr{F}_{\mathbb{Z}}$-supermartingale, $k \in \mathbb{N}$ und $a \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}_{>a}$. Then

$$\mathbb{P}\left(\beta_{a,b}(X_{\mathbb{N}}) \geqslant k \middle| \mathscr{F}_1\right) \leqslant (a/b)^k \min\left(\left(\sup_{z\in\mathbb{Z}_{\leqslant 1}} X_z\right)/a, 1\right) \quad \mathbb{P}\text{-a.e.}$$ □

§11.19 **Theorem**. *A positive $\mathscr{F}_{\mathbb{N}}$-supermartingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ converges* $\mathbb{P}$-a.e., *i.e.* $X_n \xrightarrow{\mathbb{P}\text{-a.e.}} X_\infty$. *The* $\mathbb{P}$-a.e. *limit X_∞ is $\mathscr{F}_\infty$-$\overline{\mathscr{B}}_{\geqslant 0}$-measurable and satisfies $X_n \geqslant \mathbb{E}\left(X_\infty \middle| \mathscr{F}_n\right)$* $\mathbb{P}$-a.e. *for all $n \in \mathbb{N}$.*

§11.20 **Proof** of Theorem §11.19. is given in the lecture. □

§11.21 **Remark**.

(a) Since $X_n \geqslant \mathbb{E}\left(X_\infty \middle| \mathscr{F}_n\right)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$, X_∞ is finite $\mathbb{P}$-a.e. on the event $\bigcup_{n\in\mathbb{N}}\{X_n \in \mathbb{R}_{\geqslant 0}\}$. Indeed, for every $n \in \mathbb{N}$, $\mathbb{E}\left(X_\infty \mathbb{1}_{\{X_n \in \mathbb{R}_{\geqslant 0}\}} \middle| \mathscr{F}_n\right)$ is finite $\mathbb{P}$-a.e. and hence by Theorem §07.04 (iv) also X_∞ on the event $\{X_n \in \mathbb{R}_{\geqslant 0}\}$.

(b) If $(X_n)_{n\in\mathbb{N}}$ is a $\mathbb{P}$-integrable positive $\mathscr{F}_{\mathbb{N}}$-supermartingale, i.e., $X_n \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mathbb{P}) = \bar{\mathcal{L}}^1(\mathbb{P}) \cap \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ for all $n \in \mathbb{N}$, then $X_n \geqslant \mathbb{E}(X_\infty|\mathscr{F}_n)$ implies also $X_\infty \in \bar{\mathcal{L}}^1_{\geqslant 0}(\mathbb{P})$. However, a $\mathbb{P}$-integrable positive $\mathscr{F}_{\mathbb{N}}$-supermartingale does generally not converge in $\mathcal{L}^1(\mathbb{P})$ to X_∞.

(c) If $(X_n)_{n\in\mathbb{N}}$ is a positive $\mathscr{F}_{\mathbb{N}}$-martingale, i.e., $X_n = \mathbb{E}(X_{n+1}|\mathscr{F}_n)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$, then from Theorem §11.19 follows $X_n \xrightarrow{\mathbb{P}\text{-a.e.}} X_\infty$ and $X_n \geqslant \mathbb{E}(X_\infty|\mathscr{F}_n)$ for all $n \in \mathbb{N}$, where, in general, the inequalities do not become equalities. The next statement characterises a situation in which this phenomenon does not occur. □

End LE14 2026/06/03

§11.22 **Proposition**. *Let* $s \in \mathbb{R}_{\geqslant 1}$. *For every* $Z \in \bar{\mathcal{L}}^s_{\geqslant 0}(\mathbb{P}) = \bar{\mathcal{L}}^s(\mathbb{P}) \cap \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, *the stochastic process*

$$Z_{\mathbb{N}} = (Z_n := \mathbb{E}(Z|\mathscr{F}_n))_{n\in\mathbb{N}}$$

is a positive $\mathscr{F}_{\mathbb{N}}$*-martingale, which converges* $\mathbb{P}$*-a.e. and in* $\mathcal{L}^s(\mathbb{P})$ *to* $Z_\infty := \mathbb{E}(Z|\mathscr{F}_\infty)$.

§11.23 **Proof** of Proposition §11.22. is given in the lecture. □

§11.24 **Remark**.

(a) A positive $\mathscr{F}_{\mathbb{N}}$-martingale $Z_{\mathbb{N}} = (Z_n)_{n\in\mathbb{N}}$ as in Proposition §11.22 and its $\mathbb{P}$-a.e. limit Z_∞ satisfy the equation $Z_n = \mathbb{E}(Z_\infty|\mathscr{F}_n)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$ since

$$Z_n = \mathbb{E}(Z|\mathscr{F}_n) = \mathbb{E}(\mathbb{E}(Z|\mathscr{F}_\infty)|\mathscr{F}_n) = \mathbb{E}(Z_\infty|\mathscr{F}_n) \quad \mathbb{P}\text{-a.e.}$$

by the tower property (Theorem §07.06 (ii)).

(b) Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a positive $\mathscr{F}_{\mathbb{N}}$-martingale, which converges in $\mathcal{L}^s(\mathbb{P})$ to $X_\infty \in \mathcal{L}^s_{\geqslant 0}(\mathbb{P})$. The martingale equations $X_n = \mathbb{E}(X_m|\mathscr{F}_n)$ $\mathbb{P}$-a.e. for $m \in \mathbb{N}_{\geqslant n}$ and the contraction of the conditional expectation on $\mathcal{L}^s(\mathbb{P})$, that is,

$$\|X_n - \mathbb{E}(X_\infty|\mathscr{F}_n)\|_{\mathcal{L}^p(\mathbb{P})} = \|\mathbb{E}(X_m - X_\infty|\mathscr{F}_n)\|_{\mathcal{L}^p(\mathbb{P})} \leqslant \|X_m - X_\infty\|_{\mathcal{L}^p(\mathbb{P})} \to 0,$$

imply together $X_n = \mathbb{E}(X_\infty|\mathscr{F}_n)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$. Consequently, Proposition §11.22 implies, that $\mathscr{F}_{\mathbb{N}}$-martingales of the form $(\mathbb{E}(Z|\mathscr{F}_n))_{n\in\mathbb{N}}$ with $Z \in \bar{\mathcal{L}}^s_{\geqslant 0}(\mathbb{P})$ are the positive, $\mathcal{L}^s(\mathbb{P})$-convergent $\mathscr{F}_{\mathbb{N}}$-martingales in $\bar{\mathcal{L}}^s_{\geqslant 0}(\mathbb{P})$. A positive $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ is called *closable* (abschließbar) in $\bar{\mathcal{L}}^s_{\geqslant 0}(\mathbb{P})$ if there exists $X \in \bar{\mathcal{L}}^s_{\geqslant 0}(\mathbb{P})$ such that $X_n = \mathbb{E}(X|\mathscr{F}_n)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$.

(c) For $Z \in \mathcal{L}^s_{\geqslant 0}(\mathbb{P})$ a statement holds similar to Proposition §11.22 by applying the usual decomposition $Z = Z^+ - Z^-$. □

§11.25 **Corollary**. *For every* $Z \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$, $\mathbb{E}(Z|\mathscr{F}_n)$ *converges* $\mathbb{P}$*-a.e. to* $\mathbb{E}(Z|\mathscr{F}_\infty)$ *on the complement of the event* $\bigcap_{n\in\mathbb{N}}\{\mathbb{E}(Z|\mathscr{F}_n) = \infty\}$.

§11.26 **Proof** of Corollary §11.25. Exercise. □

§11.27 **Remark**. In Corollary §11.25, integrability is not assumed. In Neveu (1975, S. 31), an $\mathscr{F}_\infty$-measurable random variable Z is constructed, which is finite $\mathbb{P}$-a.e. and satisfies $\mathbb{E}(Z|\mathscr{F}_n) = \infty$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$. In this case, $\mathbb{E}(Z|\mathscr{F}_n) \xrightarrow{\mathbb{P}\text{-a.e.}} \mathbb{E}(Z|\mathscr{F}_\infty) = Z$ only holds on a $\mathbb{P}$-null set. □

§11.28 **Reminder**. Let $\mathbb{P}_0, \mathbb{P}_1 \in \mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$ be probability measures on the measurable space $(\Omega, \mathscr{A})$, where $\mathbb{P}_0 \ll \mathbb{P}_1$ does not necessarily hold. Then any positive, numerical function $\mathrm{L} \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ with $\mathbb{P}_0 = \mathrm{L}\mathbb{P}_1 + \mathbb{1}_{\{\mathrm{L}=\infty\}}\mathbb{P}_0$ and $\mathbb{P}_1(\mathrm{L} \in \mathbb{R}_{\geqslant 0}) = 1$ is called likelihood ratio of $\mathbb{P}_0$ with respect to $\mathbb{P}_1$ (vgl. Definition §03.14). In the particular case $\mathbb{P}_0 \ll \mathbb{P}_1$ the $\mathbb{P}_1$-likelihood ratio L of $\mathbb{P}_0$ coincides with the $\mathbb{P}_1$-density $\mathrm{d}\mathbb{P}_0/\mathrm{d}\mathbb{P}_1$ of $\mathbb{P}_0$ and L is unique up to $\mathbb{P}_1$-a.e. equivalence. Moreover, given a sub-σ-algebra $\mathscr{F} \subseteq \mathscr{A}$ for each $i \in \{0, 1\}$ we denote by $\mathbb{E}_i(\bullet|\mathscr{F}) : \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A}) \ni X \mapsto \mathbb{E}_i(X|\mathscr{F}) \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F})$ a version of the conditional expectation with respect to $\mathbb{P}_i$ given $\mathscr{F}$ (see Definition §06.04). □

§11.29 **Property**. *Let $\mathbb{P}_0, \mathbb{P}_1 \in \mathfrak{M}_{pr}(\mathscr{A})$ be probability measures and let $L \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ be a $\mathbb{P}_1$-likelihood ratio of $\mathbb{P}_0$. Then by definition $\mathbb{P}_1(L) = \mathbb{P}_0(L \in \mathbb{R}_{\geqslant 0}) \in [0,1]$ and $\mathbb{P}_1(L = \infty) = 0$, and hence $\mathbb{P}_0(L = 0) = L\mathbb{P}_1(L = 0) + \mathbb{P}_0(\{L = 0\} \cap \{L = \infty\}) = 0$.*

(i) *$\mathbb{P}_0 \perp \mathbb{P}_1 \Leftrightarrow \exists A \in \mathscr{A} : \mathbb{P}_1(A) = 0$ (hence $L\mathbb{P}_1(A) = 0$) and $\mathbb{P}_0(A) = 1$ (hence $\mathbb{P}_0(A \cap \{L = \infty\}) = 1$) $\Leftrightarrow \mathbb{P}_0(L = \infty) = 1 \Leftrightarrow \mathbb{P}_1(L) = 0$;*

(ii) *$\mathbb{P}_0 \not\perp \mathbb{P}_1 \Leftrightarrow \forall A \in \mathscr{A} : \mathbb{P}_1(A) = 0$ implies $\mathbb{P}_0(A) \in [0,1)$ (in particular for $A = \{L = \infty\}$) $\Leftrightarrow \mathbb{P}_0(L = \infty) \in [0,1) \Leftrightarrow \mathbb{P}_1(L) \in (0,1]$;*

(iii) *$\mathbb{P}_0 \ll \mathbb{P}_1 \Leftrightarrow \forall A \in \mathscr{A} : \mathbb{P}_1(A) = 0$ implies $\mathbb{P}_0(A) = 0$ (in particular for $A = \{L = \infty\}$) $\Leftrightarrow \mathbb{P}_0(L = \infty) = 0 \Leftrightarrow \mathbb{P}_1(L) = 1$.* □

§11.30 **Lemma**. *Let $\mathbb{P}_0, \mathbb{P}_1 \in \mathfrak{M}_{pr}(\mathscr{A})$ be probability measures and let $\mathscr{F}_{\mathbb{N}} = (\mathscr{F}_n)_{n\in\mathbb{N}}$ be a filtration in $\mathscr{A}$. For $n \in \mathbb{N}$ denote by $\mathbb{P}_{0|n}, \mathbb{P}_{1|n} \in \mathfrak{M}_{pr}(\mathscr{F}_n)$ the restrictions of $\mathbb{P}_0$ and $\mathbb{P}_1$ on $\mathscr{F}_n$, and let $L_n \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}_n)$ be a $\mathbb{P}_{1|n}$-likelihood ratio of $\mathbb{P}_{0|n}$.*

(i) *$(L_n)_{n\in\mathbb{N}}$ is a positive $\mathscr{F}_{\mathbb{N}}$-supermartingale on the filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}_1, \mathscr{F}_{\mathbb{N}})$.*

(ii) *$(L_n)_{n\in\mathbb{N}}$ is a positive $\mathscr{F}_{\mathbb{N}}$-martingale on $(\Omega, \mathscr{A}, \mathbb{P}_1, \mathscr{F}_{\mathbb{N}})$ if $\mathbb{P}_{0|n} \ll \mathbb{P}_{1|n}$ and $d\mathbb{P}_{0|n}/d\mathbb{P}_{1|n} \in \mathcal{M}_{\geqslant 0}(\mathscr{F}_n)$ is a $\mathbb{P}_{1|n}$-density of $\mathbb{P}_{0|n}$, and hence $L_n = d\mathbb{P}_{0|n}/d\mathbb{P}_{1|n}$ $\mathbb{P}_1$-a.e. for all $n \in \mathbb{N}$.*

(iii) *$(L_n)_{n\in\mathbb{N}}$ is a closable positive $\mathscr{F}_{\mathbb{N}}$-martingale on $(\Omega, \mathscr{A}, \mathbb{P}_1, \mathscr{F}_{\mathbb{N}})$ with $L_n = \mathbb{P}_1(L|\mathscr{F}_n)$ $\mathbb{P}_1$-a.e. for all $n \in \mathbb{N}$ if $\mathbb{P}_0 \ll \mathbb{P}_1$ and $L = d\mathbb{P}_0/d\mathbb{P}_1 \in \mathcal{M}_{\geqslant 0}(\mathscr{A})$ is a $\mathbb{P}_1$-density of $\mathbb{P}_0$.*

§11.31 **Proof** of Lemma §11.30. is given in the lecture. □

§11.32 **Remark**. The positive $\mathscr{F}_{\mathbb{N}}$-supermartingale $(L_n)_{n\in\mathbb{N}}$ of likelihood ratios given in Lemma §11.30 converges $\mathbb{P}_1$-a.e. by applying Theorem §11.19. Its $\mathbb{P}_1$-a.e.-limit L_∞ is $\mathscr{F}_\infty$-measurable and satisfies $L_n \geqslant \mathbb{P}_1(L_\infty|\mathscr{F}_n)$ $\mathbb{P}_1$-a.e. for all $n \in \mathbb{N}$. Making use of Remark §11.24 (b) the closable positive $\mathscr{F}_{\mathbb{N}}$-martingale $(L_n)_{n\in\mathbb{N}}$ of likelihood ratios in Lemma §11.30 (iii) satisfies also $L_n = \mathbb{P}_1(L_\infty|\mathscr{F}_n)$ $\mathbb{P}_1$-a.e. for all $n \in \mathbb{N}$. If we assume in addition $\mathscr{A} = \mathscr{F}_\infty$ then $L_\infty = d\mathbb{P}_0/d\mathbb{P}_1$ $\mathbb{P}_1$-a.e., and hence $(d\mathbb{P}_{0|n}/d\mathbb{P}_{1|n})_{n\in\mathbb{N}}$ converges $\mathbb{P}_1$-a.e. to $d\mathbb{P}_0/d\mathbb{P}_1$. □

§11.33 **Lemma**. *Let $\mathbb{P}_0, \mathbb{P}_1 \in \mathfrak{M}_{pr}(\mathscr{A})$ be probability measures and $\mathscr{F}_{\mathbb{N}} = (\mathscr{F}_n)_{n\in\mathbb{N}}$ a filtration in $\mathscr{A} = \mathscr{F}_\infty$. For $n \in \mathbb{N}$, we denote by $\mathbb{P}_{0|n}, \mathbb{P}_{1|n} \in \mathfrak{M}_{pr}(\mathscr{F}_n)$ the restrictions of $\mathbb{P}_0, \mathbb{P}_1$ to $\mathscr{F}_n$, by $L_n \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}_n)$ a $\mathbb{P}_{1|n}$-likelihood ratio of $\mathbb{P}_{0|n}$, and by $L \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ a $\mathbb{P}_1$-likelihood ratio of $\mathbb{P}_0$. Then $L_{\mathbb{N}} = (L_n)_{n\in\mathbb{N}}$ converges $\frac{1}{2}(\mathbb{P}_0 + \mathbb{P}_1)$-a.e. to L.*

§11.34 **Proof** of Lemma §11.33. is given in the lecture. □

§11.35 **Theorem** *(Kakutani dichotomy)*. *Let $(\mathcal{S}, \mathscr{B})$ and $(\mathcal{S}_n, \mathscr{B}_n)$, $n \in \mathbb{N}$, be Polish spaces endowed with corresponding Borel σ-algebras. Then:*

(i) *For $n \in \mathbb{N}$ let $\mathbb{P}_{0|n}, \mathbb{P}_{1|n} \in \mathfrak{M}_{pr}(\mathscr{B}_n)$ be probability measures with $\mathbb{P}_{0|n} \ll \mathbb{P}_{1|n}$. The corresponding product measures $\mathbb{P}_{0|\mathbb{N}} := \bigotimes_{n\in\mathbb{N}} \mathbb{P}_{0|n}$ and $\mathbb{P}_{1|\mathbb{N}} := \bigotimes_{n\in\mathbb{N}} \mathbb{P}_{1|n}$ on the product space $(\mathcal{S}_{\mathbb{N}}, \mathscr{B}_{\mathbb{N}})$ (Definition §04.35) satisfy either $\mathbb{P}_{0|\mathbb{N}} \ll \mathbb{P}_{1|\mathbb{N}}$ or $\mathbb{P}_{0|\mathbb{N}} \perp \mathbb{P}_{1|\mathbb{N}}$.*

(ii) *Let $\mathbb{P}_0, \mathbb{P}_1 \in \mathfrak{M}_{pr}(\mathscr{B})$ be probability measures with $\mathbb{P}_0 \ll\gg \mathbb{P}_1$ (equivalent). The corresponding product measures $\mathbb{P}_0^{\mathbb{N}} := \bigotimes_{n\in\mathbb{N}} \mathbb{P}_0$ and $\mathbb{P}_1^{\mathbb{N}} := \bigotimes_{n\in\mathbb{N}} \mathbb{P}_1$ on the product space $(\mathcal{X}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}})$ (Definition §04.35) satisfy $\mathbb{P}_0^{\mathbb{N}} \perp \mathbb{P}_1^{\mathbb{N}}$ if $\mathbb{P}_0 \neq \mathbb{P}_1$.*

§11.36 **Proof** of Theorem §11.35. is given in the lecture. □

§11.37 **Lemma** *(Optional Stopping)*. *For every positive $\mathscr{F}_{\mathbb{N}}$-(super-)martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ and for every $\mathscr{F}_{\mathbb{N}}$-stopping time τ the stopped process $X_{\mathbb{N}}^\tau = (X_n^\tau := X_{\tau\wedge n})_{n\in\mathbb{N}}$ is a positive (super-) martingale with respect to $\mathscr{F}_{\mathbb{N}}^\tau = (\mathscr{F}_n^\tau := \mathscr{F}_{\tau\wedge n})_{n\in\mathbb{N}}$ and hence $\mathscr{F}_{\mathbb{N}} = (\mathscr{F}_n)_{n\in\mathbb{N}}$.*

§11.38 **Proof** of Lemma §11.37. Exercise □

§11.39 **Theorem** *(Optional Sampling)*. *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a positive $\mathscr{F}_{\mathbb{N}}$-supermartingale with* $\mathbb{P}$-a.e.-*limit X_∞. For $\mathscr{F}_{\mathbb{N}}$-stopping times τ, σ define $X_\tau \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}_\tau)$ and $X_\sigma \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{F}_\sigma)$ with $X_\tau := X_n$ on $\{\tau = n\}$ and $X_\sigma := X_n$ on $\{\sigma = n\}$ for all $n \in \overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Then $X_\tau \geqslant \mathbb{E}(X_\sigma | \mathscr{F}_\tau)$* $\mathbb{P}$-a.e. *on the event $\{\tau \leqslant \sigma\}$.*

§11.40 **Proof** of Theorem §11.39. is given in the lecture. □

§11.41 **Remark**. In the case of a positive martingale $(X_n)_{n\in\mathbb{N}}$, the inequality $X_\tau \geqslant \mathbb{E}(X_\sigma | \mathscr{F}_\tau)$ stated in Theorem §11.39 does not generally become an equality. □

End LE15 2026/06/05

§12 Integrable (sub-, super-)martingale

§12.01 **Definition**. A ($\mathbb{P}$-, $\mathcal{L}^r$-)integrable $\mathscr{F}_{\mathcal{T}}$-adapted stochastic prozess $X_{\mathcal{T}} = (X_t)_{t\in\mathcal{T}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ is called

(($\mathbb{P}$-, $\mathcal{L}^r$-)integrable) ($\mathscr{F}_{\mathcal{T}}$-)supermartingale, if $X_s \geqslant \mathbb{E}(X_t | \mathscr{F}_s)$ $\mathbb{P}$-a.e. for all $s \in \mathcal{T}$ and $t \in \mathcal{T}_{>s}$,

(($\mathbb{P}$-, $\mathcal{L}^r$-)integriable) ($\mathscr{F}_{\mathcal{T}}$-)submartingale, if $X_s \leqslant \mathbb{E}(X_t | \mathscr{F}_s)$ $\mathbb{P}$-a.e. for all $s \in \mathcal{T}$ and $t \in \mathcal{T}_{>s}$,

(($\mathbb{P}$-, $\mathcal{L}^r$-)integrable) ($\mathscr{F}_{\mathcal{T}}$-)martingale, if $X_s = \mathbb{E}(X_t | \mathscr{F}_s)$ $\mathbb{P}$-a.e. for all $s \in \mathcal{T}$ and $t \in \mathcal{T}_{>s}$.

A d-dimensional stochastic process $X_{\mathcal{T}} = ((X_t^1, \dots, X_t^d))_{t\in\mathcal{T}}$ on $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ is called *(($\mathbb{P}$-, $\mathcal{L}^r$-)integrable) ($\mathscr{F}_{\mathcal{T}}$-)(sub-,super-)martingale* if each coordinate process $X_{\mathcal{T}}^k = (X_t^k)_{t\in\mathcal{T}}$ is a (($\mathbb{P}$-, $\mathcal{L}^r$-)integrable) ($\mathscr{F}_{\mathcal{T}}$-)(sub-,super-)martingale. □

§12.02 **Lemma**.

(i) *Let $X_{\mathcal{T}}$ and $Y_{\mathcal{T}}$ be integrable $\mathscr{F}_{\mathcal{T}}$-submartingales. Then $X \vee Y := (X_t \vee Y_t)_{t\in\mathcal{T}}$ is also an integrable $\mathscr{F}_{\mathcal{T}}$-submartingales, and hence $X^+ = (X_t^+)_{t\in\mathcal{T}}$.*

(ii) *Let $X_{\mathcal{T}}$ be a $\mathbb{P}$-integrable real $\mathscr{F}_{\mathcal{T}}$-martingale and $f : \mathbb{R} \to \mathbb{R}$ a convex function. Then $(f(X_t))_{t\in\mathcal{T}}$ is a $\mathbb{P}$-integrable $\mathscr{F}_{\mathcal{T}}$-submartingale, if in addition* (A) $\mathbb{P}(f(X_t)^+) \in \mathbb{R}_{\geqslant 0}$ *for all $t \in \mathcal{T}$.*

(iii) $\mathbb{P}(f(X_{t_\infty})^+) \in \mathbb{R}_{\geqslant 0}$ *implies* (A), *if $t_\infty := \max\{\mathcal{T}\} \in \mathcal{T}$ (i.e. $\mathcal{T}_{>t_\infty} = \emptyset$).*

§12.03 **Proof** of Lemma §12.02. is given in the lecture. □

§12.04 **Remark**. Let $X_{\mathcal{T}}$ be a $\mathcal{L}^s$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale for $s \in \mathbb{R}_{\geqslant 1}$. Then $(|X_t|^s)_{t\in\mathcal{T}}$ is an integrable $\mathscr{F}_{\mathcal{T}}$-submartingale since $x \mapsto |x|^s$ is convex. □

§12.05 **Theorem**. *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be an $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-submartingale which satisfies in addition* (A) $\sup\{\mathbb{P}(X_n^+) : n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$. *Then $X_{\mathbb{N}}$ converges $\mathbb{P}$-a.e. to a $\mathbb{P}$-integrable $X_\infty \in \mathcal{M}(\mathscr{F}_\infty)$, i.e. $\mathbb{P}(|X_\infty|) \in \mathbb{R}_{\geqslant 0}$. If $X_{\mathbb{N}}$ is an $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale, then* (B) $\sup\{\mathbb{P}(|X_n|) : n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$ *and* (A) *are equivalent.*

§12.06 **Proof** of Theorem §12.05. is given in the lecture. □

§12.07 **Remark**. In the Proof §12.06 we show the *Krickeberg decomposition* $X_{\mathbb{N}} = (X_n = M_n - A_n)_{n\in\mathbb{N}}$ of an integrable $\mathscr{F}_{\mathcal{T}}$-submartingale into a positive integrable $\mathscr{F}_{\mathbb{N}}$-martingale $M_{\mathbb{N}} = (M_n)_{n\in\mathbb{N}}$ and a positive integrable $\mathscr{F}_{\mathbb{N}}$-supermartingale $A_{\mathbb{N}} = (A_n)_{n\in\mathbb{N}}$. □

§12.08 **Lemma**.

(i) *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be an $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale and let τ be a bounded $\mathscr{F}_{\mathbb{N}}$-stopping time, i.e. there is $K \in \mathbb{N}$ with $\tau \leqslant K$. Then $X_\tau = \mathbb{E}\big(X_K\big|\mathscr{F}_\tau\big)$ $\mathbb{P}$-a.e. and hence $\mathbb{E}(X_\tau) = \mathbb{E}(X_1)$.*

(ii) *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be an integrable $\mathscr{F}_{\mathbb{N}}$-adapted stochastic process. $X_{\mathbb{N}}$ is an integrable $\mathscr{F}_{\mathbb{N}}$-martingale if and only if $\mathbb{E}(X_\tau) = \mathbb{E}(X_1)$ for every bounded $\mathscr{F}_{\mathbb{N}}$-stopping time τ.*

§12.09 **Proof** of Lemma §12.08. is given in the lecture. □

§12.10 **Definition**. Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a real, $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-adapted process and $H_{\mathbb{Z}_{\geqslant 0}} = (H_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ a real, $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-predictable process (Definition §10.05). The *discrete stochastic integral* of $H_{\mathbb{Z}_{\geqslant 0}}$ with respect to $X_{\mathbb{Z}_{\geqslant 0}}$ is the $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-adapted stochastic process $[H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_{\mathbb{Z}_{\geqslant 0}} = ([H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ defined by

$$[H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_0 := 0 \quad \text{and} \quad [H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_n := \sum_{k\in[\![n]\!]} H_k(X_k - X_{k-1}), \quad \forall\, n \in \mathbb{N}.$$

If $X_{\mathbb{Z}_{\geqslant 0}}$ is a martingale, then $[H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_{\mathbb{Z}_{\geqslant 0}}$ is called *martingale transform* of $X_{\mathbb{Z}_{\geqslant 0}}$. □

§12.11 **Example**. Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a (possibly unfair) game, where $X_n - X_{n-1}$ represents the winnings per game in the n-th round. We interpret H_n as the number of game tickets bet in the n-th game and understand $H_{\mathbb{Z}_{\geqslant 0}}$ as a betting strategy. Clearly, the value of H_n must be determined at the time $n-1$, that is, before the outcome of X_n is known. In other words, $H_{\mathbb{Z}_{\geqslant 0}}$ must be predictable. Now let $X_{\mathbb{Z}_{\geqslant 0}}$ be a fair game, that is, a martingale, and let $H_{\mathbb{Z}_{\geqslant 0}}$ be $\mathcal{L}^\infty$-integrable (*locally bounded*), i.e., each $H_n \in \mathcal{L}^\infty(\mathbb{P})$ is bounded. Because $\mathbb{E}\big(X_{n+1} - X_n\big|\mathscr{F}_n\big) = 0$ $\mathbb{P}$-a.e., it follows that

$$\begin{aligned}\mathbb{E}\big([H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_{n+1}\big|\mathscr{F}_n\big) &= \mathbb{E}\big([H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_n + H_{n+1}(X_{n+1} - X_n)\big|\mathscr{F}_n\big)\\ &= [H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_n + H_{n+1}\mathbb{E}\big(X_{n+1} - X_n\big|\mathscr{F}_n\big) = [H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_n \quad \mathbb{P}\text{-a.e.},\end{aligned}$$

and thus the martingale transform $[H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_{\mathbb{Z}_{\geqslant 0}}$ is also a martingale. In the following theorem, we show that the converse also holds, that is, $X_{\mathbb{Z}_{\geqslant 0}}$ is a martingale if for sufficiently many predictable processes the discrete stochastic integral is a martingale. □

§12.12 **Proposition**. *Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a real, $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-adapted process with $X_0 \in \mathcal{L}^1(\mathbb{P})$.*

(i) *$X_{\mathbb{Z}_{\geqslant 0}}$ is an integrable martingale if and only if for every locally bounded, predictable process $H_{\mathbb{Z}_{\geqslant 0}}$, the discrete stochastic integral $[H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_{\mathbb{Z}_{\geqslant 0}}$ is an integrable martingale.*

(ii) *$X_{\mathbb{Z}_{\geqslant 0}}$ is an integrable submartingale (supermartingale) if and only if $[H_{\mathbb{Z}_{\geqslant 0}} \boxdot X_{\mathbb{Z}_{\geqslant 0}}]_{\mathbb{Z}_{\geqslant 0}}$ is a submartingale (supermartingale) for every locally bounded, predictable, and positive process $H_\bullet$.*

§12.13 **Proof** of Proposition §12.12. is given in the lecture. □

§12.14 **Remark**. The preceding theorem states, in particular, that there is no (locally bounded) betting strategy that can turn a martingale (or a positive supermartingale in the case of a real-life betting strategy) into a submartingale. However, this is exactly what is suggested by advertisements for so-called „system lottery“ and similar schemes. □

§13 Regular integrable martingale

§13.01 **Reminder**.

(a) A $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ with (B) $\sup\{\mathbb{P}(|X_n|) : n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$ converges $\mathbb{P}$-a.e. to $X_\infty \in \mathcal{L}^1(\mathscr{F}_\infty, \mathbb{P})$ by Theorem §12.05.

(b) Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be an uniformly $\mathbb{P}$-integrable (Definition §02.39) process in $\mathcal{L}^1(\mathscr{A}, \mathbb{P})$, that is $\inf\{\sup_{n\in\mathbb{N}} \mathbb{P}\big(|X_n|\mathbb{1}_{\{|X_n|\geqslant a\}}\big) : a \in \mathbb{R}_{\geqslant 0}\} = 0$. According to Theorem §02.43 (uI1), then the condition (B) also holds. Moreover, $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ converges in $\mathcal{L}^1(\mathbb{P})$ due to Theorem §02.45 if and only if $X_{\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable and convergent in probability, where $\mathbb{P}$-a.e. convergence implies convergence in probability (Lemma §02.31 (v)). □

§13.02 **Theorem**. *For a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathbb{N}})$ the following four statements are equivalent:*

(rM1) *$X_{\mathbb{N}}$ converges in $\mathcal{L}^1(\mathbb{P})$;*

(rM2) (B) $\sup\{\mathbb{P}(|X_n|) : n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$ *is satisfied. The $\mathbb{P}$-a.e.-limit $X_\infty = \lim_{n\to\infty} X_n$, which by Theorem §12.05 in $\mathcal{L}^1(\mathbb{P})$ exists, satisfies $X_n = \mathbb{E}\big(X_\infty \big| \mathscr{F}_n\big)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$;*

(rM3) *The $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}}$ is closable, i.e., there is a random variable $X \in \mathcal{L}^1(\mathscr{A}, \mathbb{P})$ with $X_n = \mathbb{E}\big(X \big| \mathscr{F}_n\big)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$;*

(rM4) *$X_{\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable.*

The integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ is called regular if one of the equivalent statements (rM1)-(rM4) is satisfied.

§13.03 **Proof** of Theorem §13.02. is given in the lecture. □

§13.04 **Remark**. By Theorem §13.02, for a regular integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ there exists also $X_\infty \in \mathcal{L}^1(\mathscr{F}_\infty, \mathbb{P})$ with

$$X_n \xrightarrow{\mathbb{P}\text{-a.e.}} X_\infty \quad \text{and} \quad X_n \xrightarrow{\mathcal{L}^1(\mathbb{P})} X_\infty.$$

If there is in addition $X \in \mathcal{L}^1(\mathscr{A}, \mathbb{P})$ with $X_n = \mathbb{E}\big(X \big| \mathscr{F}_n\big)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$, then $X_\infty = \mathbb{E}\big(X \big| \mathscr{F}_\infty\big)$ $\mathbb{P}$-a.e. In general, only the random variable X_∞ is $\mathbb{P}$-a.e. unique, but not the random variable X. However, $X_\infty = X$ $\mathbb{P}$-a.e. holds for $\mathscr{F}_\infty = \mathscr{A}$. □

§13.05 **Corollary** *(Optional sampling)*. *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a regular $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale with $\mathbb{P}$-a.e. limit X_∞.*

(i) *For every $\mathscr{F}_{\mathbb{N}}$-stopping time τ, $X_\tau \in \overline{\mathcal{M}}(\mathscr{F}_\tau)$ with $X_\tau := X_n$ on $\{\tau = n\}$ for all $n \in \overline{\mathbb{N}}$ is a $\mathbb{P}$-integrable random variable.*

(ii) *For every pair of $\mathscr{F}_{\mathbb{N}}$-stopping times τ, σ with $\tau \leqslant \sigma$ $\mathbb{P}$-a.e. holds also $X_\tau = \mathbb{E}\big(X_\sigma \big| \mathscr{F}_\tau\big)$ $\mathbb{P}$-a.e.*

(iii) *The family $\{X_\tau : \tau \text{ is a finite } \mathscr{F}_{\mathbb{T}}\text{-stopping time}\}$ is uniformly $\mathbb{P}$-integrable.*

§13.06 **Proof** of Corollary §13.05. is given in the lecture. □

§13.07 **Remark**. Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a regular $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale with $\mathbb{P}$-a.e. limit X_∞ and τ, σ $\mathscr{F}_{\mathbb{N}}$-stopping times. By definition the random variable X_τ equals X_∞ on $\{\tau = \infty\}$ (respectively X_σ on $\{\sigma = \infty\}$). Since $\tau \wedge \sigma \leqslant \sigma$ Corollary §13.05 implies $X_{\tau\wedge\sigma} = \mathbb{E}\big(X_\sigma \big| \mathscr{F}_{\tau\wedge\sigma}\big)$ $\mathbb{P}$-a.e. and $\mathbb{E}\big(X_\sigma \big| \mathscr{F}_\tau\big) = \mathbb{E}\big(X_\sigma \big| \mathscr{F}_{\tau\wedge\sigma}\big)$ $\mathbb{P}$-a.e. By using Lemma §10.18 (ii) and $\{\tau \leqslant \sigma\} \in \mathscr{F}_{\tau\wedge\sigma} \subseteq \mathscr{F}_\tau$ the random variable $Y := \mathbb{E}\big(X_\sigma \big| \mathscr{F}_{\tau\wedge\sigma}\big)\mathbb{1}_{\{\tau\leqslant\sigma\}} + X_{\tau\wedge\sigma}\mathbb{1}_{\{\tau>\sigma\}}$ is $\mathscr{F}_{\tau\wedge\sigma}$-measurable, and hence also $\mathscr{F}_\tau$-measurable. Moreover, for all $F \in \mathscr{F}_\tau$ (and hence also for all $F \in \mathscr{F}_{\tau\wedge\sigma} \subseteq \mathscr{F}_\tau$) it satisfies the

Radon-Nikodym equation (applying Lemma §10.18 (i), i.e. $F \cap \{\tau \leqslant \sigma\} \in \mathscr{F}_{\tau\wedge\sigma}$)

$$\begin{aligned}\mathbb{E}(\mathbb{1}_F \mathbb{E}(X_\sigma|\mathscr{F}_\tau)) &= \mathbb{E}(\mathbb{1}_F X_\sigma) = \mathbb{E}(\mathbb{1}_{F\cap\{\tau\leqslant\sigma\}} X_\sigma) + \mathbb{E}(\mathbb{1}_{F\cap\{\tau>\sigma\}} X_\sigma)\\ &= \mathbb{E}(\mathbb{1}_{F\cap\{\tau\leqslant\sigma\}} \mathbb{E}(X_\sigma|\mathscr{F}_{\tau\wedge\sigma})) + \mathbb{E}(\mathbb{1}_{F\cap\{\tau>\sigma\}} X_{\tau\wedge\sigma})\\ &= \mathbb{E}\Big(\mathbb{1}_F\big(\mathbb{E}(X_\sigma|\mathscr{F}_{\tau\wedge\sigma})\mathbb{1}_{\{\tau\leqslant\sigma\}} + X_{\tau\wedge\sigma}\mathbb{1}_{\{\tau>\sigma\}}\big)\Big).\end{aligned}$$

Therewith, we have $\mathbb{E}(X_\sigma|\mathscr{F}_\tau) = Y$ $\mathbb{P}$-a.e. and also $\mathbb{E}(X_\sigma|\mathscr{F}_{\tau\wedge\sigma}) = Y$ $\mathbb{P}$-a.e. implying together $\mathbb{E}(X_\sigma|\mathscr{F}_\tau) = \mathbb{E}(X_\sigma|\mathscr{F}_{\tau\wedge\sigma})$ $\mathbb{P}$-a.e.. Consequently, for arbitrary $\mathscr{F}_{\mathcal{T}}$-stopping times τ, σ follows $X_{\tau\wedge\sigma} = \mathbb{E}(X_\sigma|\mathscr{F}_\tau)$ $\mathbb{P}$-a.e. □

§13.08 **Reminder**. A sequence $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable, if it is bounded in $\mathcal{L}^s(\mathbb{P})$ with $s \in \mathbb{R}_{>1}$, i.e. $\sup\{\|X_n\|_{\mathcal{L}^s} : n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$ (Lemma §02.41 (iv)). □

§13.09 **Proposition**. *Every $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}}$ is regular, if it is bounded in $\mathcal{L}^s(\mathbb{P})$ with $s \in \mathbb{R}_{>1}$. In this case the martingale converges in $\mathcal{L}^s(\mathbb{P})$ to its $\mathbb{P}$-a.e.-limit $X_\infty \in \overline{\mathcal{M}}(\mathscr{F}_\infty)$.*

§13.10 **Proof** of Proposition §13.09. is given in the lecture. □

§13.11 **Remark**. The last statement does in general not hold for $s = 1$. □

End LE16 2026/06/10

§13.12 **Lemma**. *For every positive, $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-submartingale $(X_n)_{n\in\mathbb{N}}$ and for all $a \in \mathbb{R}_{>0}$ holds*

$$a\mathbb{P}(\sup_{m\in[\![n]\!]} X_m > a) \leqslant \mathbb{E}(X_n \mathbb{1}_{\{\sup_{m\in[\![n]\!]} X_m > a\}}) \quad \textit{for all } n \in \mathbb{N}.$$

§13.13 **Proof** of Lemma §13.12. is given in the lecture. □

§13.14 **Reminder**. A sequence $(X_n)_{n\in\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable, if $\sup_{n\in\mathbb{N}} |X_n|$ is $\mathbb{P}$-integrable (Remark §02.40 (d)). □

§13.15 **Lemma**. *Let $(X_n)_{n\in\mathbb{N}}$ be a $\mathscr{F}_{\mathbb{N}}$-martingale bounded in $\mathcal{L}^s(\mathscr{A}, \mathbb{P})$ with $s \in \mathbb{R}_{>1}$. Then the random variable $\sup_{n\in\mathbb{N}} |X_n| \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ is also $\mathcal{L}^s$-integrable with $\|\sup_{n\in\mathbb{N}} |X_n|\|_{\mathcal{L}^s} \leqslant \frac{s}{s-1} \sup_{n\in\mathbb{N}} \|X_n\|_{\mathcal{L}^s}$.*

§13.16 **Proof** of Lemma §13.15. is given in the lecture. □

§13.17 **Remark**. The last statement does in general not hold for $s = 1$. However, if the martingale $(X_n)_{n\in\mathbb{N}}$ satisfies the condition $\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|(\log|X_n|)^+) \in \mathbb{R}_{\geqslant 0}$, then the random variable $\sup_{n\in\mathbb{N}} |X_n| \in \overline{\mathcal{M}}_{\geqslant 0}(\mathscr{A})$ is also integrable (Neveu (1975, Proposition IV-2-10, S.70)) and hence the martingale $(X_n)_{n\in\mathbb{N}}$ is regular. □

The concepts of filtration and (sub-, super-)martingale have never required that the index set $\mathcal{T}$ (interpreted as time) be a subset of $\mathbb{R}_{\geqslant 0}$. We now consider the case $\mathcal{T} = \mathbb{Z}_{\leqslant 0}$. For a filtration $\mathscr{F}_{\mathbb{Z}_{\leqslant 0}} = (\mathscr{F}_z)_{z\in\mathbb{Z}_{\leqslant 0}}$ and a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\leqslant 0}}$-(sub-, super-) martingale $X_{\mathbb{Z}_{\leqslant 0}} = (X_z)_{z\in\mathbb{Z}_{\leqslant 0}}$, it holds then $\mathscr{F}_{z-1} \subseteq \mathscr{F}_z$, $X_z \in \mathcal{L}^1(\mathscr{F}_z, \mathbb{P})$ and $\mathbb{E}(X_z|\mathscr{F}_{z-1}) = X_{z-1}$ $\mathbb{P}$-a.e. for all $z \in \mathbb{Z}_{\leqslant 0}$.

§13.18 **Definition**. A stochastic process $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ is called a *$\mathbb{P}$-integrable backward $\mathscr{F}_{\mathbb{Z}_{\leqslant 0}}$- (sub-, super-) martingale*, if $(X_{-z})_{z\in\mathbb{Z}_{\leqslant 0}}$ is a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\leqslant 0}}$- (sub-, super-) martingale. □

§13.19 **Remark**. A $\mathbb{P}$-integrable backward $\mathscr{F}_{\mathbb{Z}_{\leqslant 0}}$-martingale $(X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ is always uniformly $\mathbb{P}$-integrable by Lemma §07.32 using that $X_{-z} = \mathbb{E}(X_0|\mathscr{F}_z)$ $\mathbb{P}$-a.e. for all $z \in \mathbb{Z}_{\leqslant 0}$. □

§13.20 **Theorem**. *A $\mathbb{P}$-integrable backward $\mathscr{F}_{\mathbb{Z}_{\leqslant 0}}$-martingale $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ converges $\mathbb{P}$-a.e. The $\mathbb{P}$-a.e. limit X_∞ is measurable with respect to $\mathscr{F}_{-\infty} := \bigcap_{z\in\mathbb{Z}_{\leqslant 0}} \mathscr{F}_z$ and $\mathbb{P}$-integrable. Moreover, it holds $X_\infty = \mathbb{E}\big(X_0 \big| \mathscr{F}_{-\infty}\big)$ $\mathbb{P}$-a.e. and $X_{\mathbb{Z}_{\geqslant 0}}$ converges to X_∞ also in $\mathcal{L}^1(\mathbb{P})$.*

§13.21 **Proof** of Theorem §13.20. is given in the lecture. □

§13.22 **Corollary** *(Kolmogorov's strong law of large numbers). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) real random variables in $\mathcal{L}^1(\Omega, \mathscr{A}, \mathbb{P})$, then:*

$$n^{-1}\sum_{k\in[\![n]\!]} X_k \xrightarrow{\mathbb{P}\text{-a.e.}} \mathbb{E}(X_1) \quad \textit{and} \quad n^{-1}\sum_{k\in[\![n]\!]} X_k \xrightarrow{\mathcal{L}^1(\mathbb{P})} \mathbb{E}(X_1).$$

§13.23 **Proof** of Corollary §13.22. is given in the lecture. □

§14 Regular stopping time for an integrable martingale

§14.01 **Lemma**. *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be an integrable $\mathscr{F}_{\mathbb{N}}$-(sub-, super-)martingale. For every $\mathscr{F}_{\mathbb{N}}$-stopping time τ the stopped process $X^\tau_{\mathbb{N}} = (X^\tau_n := X_{\tau\wedge n})_{n\in\mathbb{N}}$ is an integrable (sub-, super-)martingale with respect to both filtration $\mathscr{F}_{\mathbb{N}} = (\mathscr{F}_n)_{n\in\mathbb{N}}$ and $\mathscr{F}^\tau_{\mathbb{N}} = (\mathscr{F}^\tau_n := \mathscr{F}_{\tau\wedge n})_{n\in\mathbb{N}}$.*

§14.02 **Proof** of Lemma §14.01. Exercise. □

§14.03 **Definition**. A $\mathscr{F}_{\mathbb{N}}$-stopping time τ is called *regular* for an integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$, if the stopped process $X^\tau_{\mathbb{N}} = (X^\tau_n := X_{\tau\wedge n})_{n\in\mathbb{N}}$ is regular. □

§14.04 **Proposition**. *For a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathbb{N}})$ and a $\mathscr{F}_{\mathbb{N}}$-stopping time τ the following three statements are equivalent:*

(rS1) *τ is regular for $X_{\mathbb{N}}$;*

(rS2) *τ fulfils*

(rS2a) *$X_{\mathbb{N}}$ converges $\mathbb{P}$-a.e. to $X_\infty \in \overline{\mathcal{M}}(\mathscr{F}_\infty)$ on the event $\{\tau = \infty\}$,*

(rS2b) *$X_\tau \in \overline{\mathcal{M}}(\mathscr{F}_\tau)$ is $\mathbb{P}$-integrable (with $X_\tau := X_n$ on $\{\tau = n\}$ for all $n \in \overline{\mathbb{N}}$),*

(rS2c) *$X_{\tau\wedge n} = \mathbb{E}\big(X_\tau \big| \mathscr{F}_n\big)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$.*

(rS3) *τ fulfils*

(rS3a) *$(X_n \mathbb{1}_{\{\tau>n\}})_{n\in\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable,*

(rS3b) *$\mathbb{E}(|X_\tau|\mathbb{1}_{\{\tau\in\mathbb{N}\}}) \in \mathbb{R}_{\geqslant 0}$.*

§14.05 **Proof** of Proposition §14.04. is given in the lecture. □

§14.06 **Remark**. If $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ is an $\mathbb{P}$-integrable martingale satisfying $\sup\{\mathbb{P}(|X_n|) : n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$, The conditions (rS2a)-(rS2b) and (rS3b) are satisfied, and hence in particular if $X_{\mathbb{N}}$ is a positive $\mathbb{P}$-integrable martingale (since $\mathbb{P}(|X_n|) = \mathbb{P}(X_n) = \mathbb{P}(X_1)$). □

§14.07 **Theorem** *(Optional sampling). Let τ be a regular $\mathscr{F}_{\mathbb{N}}$-stopping time for a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ with $\mathbb{P}$-a.e.-limit $X_\infty \in \overline{\mathcal{M}}(\mathscr{F}_\infty)$ on $\{\tau = \infty\}$, and let σ_1, σ_2 be $\mathscr{F}_{\mathbb{N}}$-stopping times with $\sigma_1 \leqslant \sigma_2 \leqslant \tau$. For $i \in [\![2]\!]$ define $X_{\sigma_i} \in \overline{\mathcal{M}}(\mathscr{F}_{\sigma_i})$ with $X_{\sigma_i} := X_n$ on $\{\sigma_i = n\}$ for all $n \in \overline{\mathbb{N}}$. Then, X_{σ_1} and X_{σ_2} are $\mathbb{P}$-integrable and $X_{\sigma_1} = \mathbb{E}\big(X_{\sigma_2} \big| \mathscr{F}_{\sigma_1}\big)$ $\mathbb{P}$-a.e..*

§14.08 **Proof** of Theorem §14.07. is given in the lecture. □

§14.09 **Corollary**. *Let τ and σ be two $\mathscr{F}_{\mathbb{N}}$-stopping times with $\tau \leqslant \sigma$ $\mathbb{P}$-a.e. Then τ is regular for a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$, if σ is regular for $X_{\mathbb{N}}$.*

§14.10 **Proof** of Corollary §14.09. is given in the lecture. □

§14.11 **Remark**. From Corollary §14.09 follows immediately, that every stopping time is regular for a regular integrable martingale (choose the constant stopping time $\sigma = \infty$). On the other hand, for an integrable martingale is every finite constant stopping time regular, and hence by Corollary §14.09 also every bounded stopping time. □

§14.12 **Corollary**. *For a $\mathbb{P}$-integrable martingale $(X_n)_{n\in\mathbb{N}}$ with $\sup\{\mathbb{P}(|X_n|): n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$ (hence for a positive, $\mathbb{P}$-integrable martingale), the hitting time $\tau_a := \inf\{n \in \mathbb{N}: |X_n| > a\}$ with $a \in \mathbb{R}_{\geqslant 0}$ is regular (with $\inf \varnothing = \infty$, see Example §10.10 (b)).*

§14.13 **Proof** of Corollary §14.12. is given in the lecture. □

§14.14 **Lemma**. *Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-martingale. A $\mathscr{F}_{\mathbb{N}}$-stopping time τ is regular for $X_{\mathbb{N}}$ and $X_{\mathbb{N}}$ converges $\mathbb{P}$-a.e. to 0 on $\{\tau = \infty\}$ if and only if the following two conditions* (c1) $\mathbb{P}\big(|X_\tau|\mathbb{1}_{\{\tau\in\mathbb{N}\}}\big) \in \mathbb{R}_{\geqslant 0}$ *and* (c2) $\lim_{n\to\infty} \mathbb{P}\big(|X_n|\mathbb{1}_{\{\tau>n\}}\big) = 0$ *are satisfied.*

§14.15 **Proof** of Lemma §14.14. is given in the lecture. □ End LE17 2026/06/17

§14.16 **Example** *(Wald's equation)*. Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. real random variables defined on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \sigma_{\mathbb{N}}(X_{\mathbb{N}}))$, where $\sigma_{\mathbb{N}}(X_{\mathbb{N}})$ is the filtration generated by $X_{\mathbb{N}}$ (Definition §10.02), and set $S_{\mathbb{N}} = (S_n := \sum_{i\in[\![n]\!]} X_i)_{n\in\mathbb{N}}$. If in addition $X_1 \in \mathcal{L}^2(\mathbb{P})$, then $S^\circ_{\mathbb{N}} := (S^\circ_n := S_n - n\mathbb{P}(X_1))_{n\in\mathbb{N}}$ and $V_{\mathbb{N}} := (V_n := |S_n - n\mathbb{P}(X_1)|^2 - n\mathbb{V}\mathrm{ar}(X_1))_{n\in\mathbb{N}}$ are $\mathbb{P}$-integrable $\sigma_{\mathbb{N}}(X_{\mathbb{N}})$-martingales, which are in general not regular. On the other hand, every $\mathbb{P}$-integrable $\sigma_{\mathbb{N}}(X_{\mathbb{N}})$-stopping time τ, i.e. $\mathbb{P}(\tau) \in \mathbb{R}_{\geqslant 0}$, is regular for both $S^\circ_{\mathbb{N}}$ and $V_{\mathbb{N}}$, and *Wald's equations* (a) $\mathbb{E}(S_\tau) = \mathbb{E}(\tau)\mathbb{E}(X_1)$ and (b) $\mathbb{E}((S_\tau - \tau\mathbb{E}(X_1))^2) = \mathbb{E}(\tau)\mathbb{V}\mathrm{ar}(X_1)$ are satisfied. If in addition $\mathbb{E}(\tau^2) \in \mathbb{R}_{\geqslant 0}$ and $\tau, X_{\mathbb{N}}$ are independent, then it holds also (c) $\mathbb{V}\mathrm{ar}(S_\tau) = \mathbb{V}\mathrm{ar}(\tau)(\mathbb{E}(X_1))^2 + \mathbb{E}(\tau)\mathbb{V}\mathrm{ar}(X_1)$. □

§15 Regular integrable submartingale

§15.01 **Preliminary**. The *Krickeberg decomposition* for submartingales (see Remark §12.07) allows the theory for integrable martingales to be extended to integrable submartingales without further difficulties. □

§15.02 **Reminder**.

(a) A $\mathbb{P}$-integrable $\mathscr{F}_{\mathcal{T}}$-submartingale $(X_n)_{n\in\mathbb{N}}$ with (A) $\sup\{\mathbb{P}(X_n^+): n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$ converges $\mathbb{P}$-a.e. to $X_\infty \in \mathcal{L}^1(\mathscr{F}_\infty, \mathbb{P})$ by Theorem §12.05.

(b) A sequence $(X_n^+)_{n\in\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable (Definition §02.39) in $\mathcal{L}^1(\mathscr{A}, \mathbb{P})$, i.e. $\inf\{\sup_{n\in\mathbb{N}} \mathbb{P}\big(X_n^+\mathbb{1}_{\{X_n^+\geqslant a\}}\big): a \in \mathbb{R}_{\geqslant 0}\} = 0$, if $\sup_{n\in\mathbb{N}} X_n^+ \in \mathcal{L}^1(\mathbb{P})$ (Remark §02.40 (d)). □

§15.03 **Theorem**. *For a $\mathbb{P}$-integrable $\mathscr{F}_{\mathcal{T}}$-submartingale $(X_n)_{n\in\mathbb{N}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathcal{T}})$ the following four statements are equivalent:*

(rSM1) *The sequence $(X_n^+)_{n\in\mathbb{N}}$ converges in $\mathcal{L}^1(\mathbb{P})$;*

(rSM2) *It holds* (A) $\sup\{\mathbb{P}(X_n^+): n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$. *The $\mathbb{P}$-a.e.-limit $X_\infty = \lim_{n\to\infty} X_n$, which exists in $\mathcal{L}^1(\mathbb{P})$ by Theorem §12.05, satisfies $X_n \leqslant \mathbb{E}\big(X_\infty \big| \mathscr{F}_n\big)$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$;*

(rSM3) *There is a random variable* $Y \in \mathcal{L}^1(\mathscr{A}, \mathbb{P})$ *with* $X_n \leqslant \mathbb{E}(Y|\mathscr{F}_n)$ *for all* $n \in \mathbb{N}$;

(rSM4) $(X_n^+)_{n\in\mathbb{N}}$ *is uniformly* $\mathbb{P}$*-integrable in* $\mathcal{L}^1(\mathscr{A}, \mathbb{P})$.

The $\mathbb{P}$*-integrable* $\mathscr{F}_{\mathcal{T}}$*-submartingale* $(X_n)_{n\in\mathbb{N}}$ *is called* regular *if one of the equivalent statements* (rSM1)-(rSM4) *is satisfied.*

§15.04 **Proof** of Theorem §15.03. is given in the lecture. □

§15.05 **Remark.** For a negative integrable submartingale (that is, for a positive integrable supermartingale with its sign changed), the statements in §15.03 are obviously fulfilled. In general, however, the submartingale does not converge in the mean, although it always converges almost surely. In particular, the statement (rSM1) of the preceding proposition is strictly less restrictive than the convergence of the submartingale in $\mathcal{L}^1$. On the other hand, for a positive submartingale, (rSM1) naturally implies $\mathcal{L}^1$-convergence. □

§15.06 **Corollary** *(Optional sampling). Let* $(X_n)_{n\in\mathbb{N}}$ *be a regular* $\mathbb{P}$*-integrable* $\mathscr{F}_{\mathcal{T}}$*-submartingale with* $\mathbb{P}$-a.e.*-limit* X_∞.

(i) *For every* $\mathscr{F}_{\mathcal{T}}$*-stopping time* τ*, the random variable* $X_\tau \in \overline{\mathcal{M}}(\mathscr{F}_\tau)$ *is* $\mathbb{P}$*-integrable (with* $X_\tau := X_n$ *on* $\{\tau = n\}$ *for all* $n \in \overline{\mathbb{N}}$*).*

(ii) *Every pair of* $\mathscr{F}_{\mathcal{T}}$*-stopping times* τ, σ *with* $\tau \leqslant \sigma$ $\mathbb{P}$-a.e. *satisfy* $X_\tau \leqslant \mathbb{E}(X_\sigma|\mathscr{F}_\tau)$ $\mathbb{P}$-a.e..

§15.07 **Proof** of Corollary §15.06. is given in the lecture. □

§15.08 **Remark.** In conclusion, it is possible without difficulties to introduce also for integrable submartingales the regularity of a stopping time (see Proposition §14.04 and Theorem §14.07). The only differences in the statements are the replacement of the word «martingale» with «submartingale» and the retention of the inequalities $X_{\tau\wedge n} \leqslant \mathbb{E}(X_\tau|\mathscr{F}_n)$ and $X_{\sigma_1} \leqslant \mathbb{E}(X_{\sigma_2}|\mathscr{F}_{\sigma_1})$ instead of the corresponding equalities. □

§16 Doob decomposition and square variation

§16.01 **Preliminary..** Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-adapted, $\mathbb{P}$-integrable process on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathbb{Z}_{\geqslant 0}})$. We decompose $X_{\mathbb{Z}_{\geqslant 0}}$ into a sum consisting of a $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-predictable process $A_{\mathbb{Z}_{\geqslant 0}} = (A_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ (Definition §10.05), the trend component or systematic component, and a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale $M_{\mathbb{Z}_{\geqslant 0}} = (M_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ (Definition §12.01), the error component. To this end, define $M_0 := X_0$, $A_0 := 0$ and for $n \in \mathbb{N}$

$$M_n := X_0 + \sum_{k\in[\![n]\!]} \big(X_k - \mathbb{E}(X_k|\mathscr{F}_{k-1})\big) \quad \text{and} \quad A_n := \sum_{k\in[\![n]\!]} \big(\mathbb{E}(X_k|\mathscr{F}_{k-1}) - X_{k-1}\big).$$

Evidently, $X_{\mathbb{Z}_{\geqslant 0}} = (X_n = M_n + A_n)_{n\in\mathbb{Z}_{\geqslant 0}} = M_{\mathbb{Z}_{\geqslant 0}} + A_{\mathbb{Z}_{\geqslant 0}}$. By construction, for every $n \in \mathbb{N}$

$$M_n - M_{n-1} = X_n - \mathbb{E}(X_n|\mathscr{F}_{n-1}) \quad \text{and} \quad A_n - A_{n-1} = \mathbb{E}(X_n|\mathscr{F}_{n-1}) - X_{n-1}.$$

Consequently, $A_{\mathbb{Z}_{\geqslant 0}}$ is $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-predictable (with $A_0 = 0$), and $M_{\mathbb{Z}_{\geqslant 0}}$ is a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale, since $\mathbb{E}(M_n - M_{n-1}|\mathscr{F}_{n-1}) = \mathbb{E}\big(X_n - \mathbb{E}(X_n|\mathscr{F}_{n-1})|\mathscr{F}_{n-1}\big) = 0$ $\mathbb{P}$-a.e. □

§16.02 **Theorem** *(Doob decomposition). Let* $X_{\mathbb{Z}_{\geqslant 0}}$ *be a* $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$*-adapted,* $\mathbb{P}$*-integrable process. Then there exists a unique decomposition* $X_{\mathbb{Z}_{\geqslant 0}} = M_{\mathbb{Z}_{\geqslant 0}} + A_{\mathbb{Z}_{\geqslant 0}}$*, where* $A_{\mathbb{Z}_{\geqslant 0}}$ *is* $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$*-predictable with* $A_0 = 0$ *and* $M_{\mathbb{Z}_{\geqslant 0}}$ *is a* $\mathbb{P}$*-integrable* $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$*-martingale. This representation of* $X_{\mathbb{Z}_{\geqslant 0}}$ *is called the* Doob decomposition. $X_{\mathbb{Z}_{\geqslant 0}}$ *is a submartingale (respectively supermartingale) if and only if* $A_{\mathbb{Z}_{\geqslant 0}}$ *is an increasing (respectively decreasing) process (Definition §10.05).*

§16.03 **Proof** of Theorem §16.02. is given in the lecture. □

§16.04 **Reminder**.

(a) A $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{N}}$-submartingale $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ with (A) $\sup\{\mathbb{P}(X_n^+) : n \in \mathbb{N}\} \in \mathbb{R}_{\geqslant 0}$ converges $\mathbb{P}$-a.e. to $X_\infty \in \mathcal{L}^1(\mathscr{F}_\infty, \mathbb{P})$ by Theorem §12.05.

(b) A sequence $(X_n^+)_{n\in\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable in $\mathcal{L}^1(\mathscr{A}, \mathbb{P})$ (Definition §02.39), i.e. $\inf\{\sup_{n\in\mathbb{N}} \mathbb{P}\big(X_n^+ \mathbb{1}_{\{X_n^+ \geqslant a\}}\big) : a \in \mathbb{R}_{\geqslant 0}\} = 0$, if $\sup_{n\in\mathbb{N}} X_n^+ \in \mathcal{L}^1(\mathbb{P})$ (Remark §02.40 (d)). □

§16.05 **Proposition**. *Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-submartingale on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}_{\mathbb{Z}_{\geqslant 0}})$ and let $X_{\mathbb{Z}_{\geqslant 0}} = M_{\mathbb{Z}_{\geqslant 0}} + A_{\mathbb{Z}_{\geqslant 0}}$ be its Doob decomposition.*

(i) *The condition* (A) $\sup\{\mathbb{P}(X_n^+) : n \in \mathbb{Z}_{\geqslant 0}\} \in \mathbb{R}_{\geqslant 0}$ *(which suffices to ensure $\mathbb{P}$-a.e. convergence of the submartingale) is equivalent to the conjunction of the two conditions:* (i.c1) $A_\infty \in \mathcal{L}^1(\mathbb{P})$ *and* (i.c2) $\sup\{\mathbb{P}(|M_n|) : n \in \mathbb{Z}_{\geqslant 0}\} \in \mathbb{R}_{\geqslant 0}$.

(ii) *The convergence in $\mathcal{L}^1(\mathbb{P})$ of the submartingale $X_{\mathbb{Z}_{\geqslant 0}}$ is equivalent to the conjunction of the two conditions* (ii.c1) *$M_{\mathbb{Z}_{\geqslant 0}}$ is a regular $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale and* (i.c1).

(iii) *For every $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-stopping time τ regular for the martingale $M_{\mathbb{Z}_{\geqslant 0}}$, the random variable X_τ is $\mathbb{P}$-integrable if and only if A_τ is $\mathbb{P}$-integrable, and then $\mathbb{E}(X_\tau) = \mathbb{E}(M_0) + \mathbb{E}(A_\tau)$.*

§16.06 **Proof** of Proposition §16.05. is given in the lecture. □

§16.07 **Example**. Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a square $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale, i.e. $X_n \in \mathcal{L}^2(\mathscr{A}, \mathbb{P})$, $n \in \mathbb{Z}_{\geqslant 0}$. By Remark §12.04, $X^2_{\mathbb{Z}_{\geqslant 0}} = (X_n^2)_{n\in\mathbb{Z}_{\geqslant 0}}$ is a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-submartingale. Furthermore, $\mathbb{E}\big(X_{i-1}X_i \big| \mathscr{F}_{i-1}\big) = X_{i-1}\mathbb{E}\big(X_i \big| \mathscr{F}_{i-1}\big) = X_{i-1}^2$ $\mathbb{P}$-a.e. for all $i \in \mathbb{N}$, hence considering the Doob decomposition $X^2_{\mathbb{Z}_{\geqslant 0}} = M_{\mathbb{Z}_{\geqslant 0}} + A_{\mathbb{Z}_{\geqslant 0}}$ of $X^2_{\mathbb{Z}_{\geqslant 0}}$, where $M_{\mathbb{Z}_{\geqslant 0}} = (X_n^2 - A_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ is a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale, for all $n \in \mathbb{N}$ we have

$$A_n = \sum_{i\in[\![n]\!]} \big(\mathbb{E}\big(X_i^2 \big| \mathscr{F}_{i-1}\big) - X_{i-1}^2\big) = \sum_{i\in[\![n]\!]} \big(\mathbb{E}\big(|X_i - X_{i-1}|^2 \big| \mathscr{F}_{i-1}\big) - 2X_{i-1}^2 + 2\mathbb{E}\big(X_{i-1}X_i \big| \mathscr{F}_{i-1}\big)\big)$$

$$= \sum_{i\in[\![n]\!]} \mathbb{E}\big(|X_i - X_{i-1}|^2 \big| \mathscr{F}_{i-1}\big) \quad \mathbb{P}\text{-a.e.}$$ □

§16.08 **Definition**. Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a square $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale. The unique increasing process $A_{\mathbb{Z}_{\geqslant 0}}$ for which $(X_n^2 - A_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ becomes a $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale is called *square variation process* of $X_{\mathbb{Z}_{\geqslant 0}}$ and is denoted by $\langle X_{\mathbb{Z}_{\geqslant 0}}\rangle_{\mathbb{Z}_{\geqslant 0}} := (\langle X_{\mathbb{Z}_{\geqslant 0}}\rangle_n)_{n\in\mathbb{Z}_{\geqslant 0}} := A_{\mathbb{Z}_{\geqslant 0}}$. □

§16.09 **Proposition**. *Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ be a square $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale with square variation process $\langle X_{\mathbb{Z}_{\geqslant 0}}\rangle_{\mathbb{Z}_{\geqslant 0}} = (\langle X_{\mathbb{Z}_{\geqslant 0}}\rangle)_{n\in\mathbb{Z}_{\geqslant 0}}$ as in Definition §16.08. Then for each $n \in \mathbb{N}$ we have $\langle X_{\mathbb{Z}_{\geqslant 0}}\rangle_n = \sum_{i\in[\![n]\!]} \mathbb{E}\big(|X_i - X_{i-1}|^2 \big| \mathscr{F}_{i-1}\big)$ $\mathbb{P}$-a.e. and $\mathbb{E}\big(\langle X_{\mathbb{Z}_{\geqslant 0}}\rangle_n\big) = \mathbb{V}\mathrm{ar}(X_n - X_0)$.*

§16.10 **Proof** of Proposition §16.09. The statement follows immediately from Example §16.07. □

§16.11 **Example**. Let $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ be sequence of independent and square $\mathbb{P}$-integrable random variables defined on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{N}})))$, where $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{N}}) = (\sigma_n(X_{\mathbb{N}}))_{n\in\mathbb{Z}_{\geqslant 0}}$ with $\sigma_0(X_{\mathbb{N}}) = \{\emptyset, \Omega\}$ and $\sigma_n(X_{\mathbb{N}}) = \sigma(X_j : j \in [\![n]\!])$, $n \in \mathbb{N}$, is the filtration generated by $X_{\mathbb{N}}$ (Example §10.04).

(a) If $\mathbb{E}(X_n) = 0$ for all $n \in \mathbb{N}$, then $S_{\mathbb{Z}_{\geqslant 0}} := (S_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ with $S_0 := 0$ and $S_n := \sum_{i\in[\![n]\!]} X_i$, $n \in \mathbb{N}$, is a square $\mathbb{P}$-integrable $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{N}})$-martingale with square variation process $\langle S_{\mathbb{Z}_{\geqslant 0}}\rangle_{\mathbb{Z}_{\geqslant 0}} = (\langle S_{\mathbb{Z}_{\geqslant 0}}\rangle_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ given by $\langle S_{\mathbb{Z}_{\geqslant 0}}\rangle_0 = 0$ and $\langle S_{\mathbb{Z}_{\geqslant 0}}\rangle_n = \sum_{i\in[\![n]\!]} \mathbb{E}\big(X_i^2 \big| \sigma_{i-1}(X_{\mathbb{N}})\big) = \sum_{i\in[\![n]\!]} \mathbb{E}(X_i^2)$ $\mathbb{P}$-a.e. for $n \in \mathbb{N}$. Note that in order for $\langle S_{\mathbb{Z}_{\geqslant 0}}\rangle_{\mathbb{Z}_{\geqslant 0}}$ to have this simple form, it is not enough for the random variables $X_{\mathbb{N}} = (X_n)_{n\in\mathbb{N}}$ to be uncorrelated.

(b) If $\mathbb{E}(X_n) = 1$ for all $n \in \mathbb{N}$, then $P_{\mathbb{Z}_{\geqslant 0}} := (P_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ with $P_0 := 1$ and $P_n := \prod_{i \in [\![n]\!]} X_i$, $n \in \mathbb{N}$, is a square $\mathbb{P}$-integrable $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{N}})$-martingale with square variation process $\langle P_{\mathbb{Z}_{\geqslant 0}} \rangle_{\mathbb{Z}_{\geqslant 0}} = (\langle P_{\mathbb{Z}_{\geqslant 0}} \rangle_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ given by $\langle P_{\mathbb{Z}_{\geqslant 0}} \rangle_0 = 0$ and $\langle P_{\mathbb{Z}_{\geqslant 0}} \rangle_n = \sum_{i \in [\![n]\!]} \mathbb{V}\mathrm{ar}(X_i) P_{i-1}^2$ $\mathbb{P}$-a.e. for $n \in \mathbb{N}$, since $\mathbb{E}\big(|P_n - P_{n-1}|^2 \big| \mathscr{F}_{n-1}\big) = \mathbb{E}\big(|X_n - 1|^2 P_{n-1}^2 \big| \mathscr{F}_{n-1}\big) = \mathbb{V}\mathrm{ar}(X_n) P_{n-1}^2$ $\mathbb{P}$-a.e. for all $n \in \mathbb{N}$. Hence, the square variation process is a truly random process. □

§16.12 **Lemma**. *Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ be a square $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale with square variation process $\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\mathbb{Z}_{\geqslant 0}}$ and let τ be a $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-stopping time. Then the stopped process $X_{\mathbb{Z}_{\geqslant 0}}^{\tau}$ has square variation process $\langle X_{\mathbb{Z}_{\geqslant 0}}^{\tau} \rangle_{\mathbb{Z}_{\geqslant 0}} = \langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\mathbb{Z}_{\geqslant 0}}^{\tau} = (\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\tau \wedge n})_{n \in \mathbb{Z}_{\geqslant 0}}$.*

§16.13 **Proof** of Lemma §16.12. is given in the lecture. □

§16.14 **Proposition**. *Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ be a square $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale with $X_0 = 0$.*

End LE18 2026/06/19

(i) *If $\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\infty} := \sup\{\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_n : n \in \mathbb{Z}_{\geqslant 0}\}$ is $\mathbb{P}$-integrable, then the martingale $X_{\mathbb{Z}_{\geqslant 0}}$ converges in $\mathcal{L}^2(\mathbb{P})$ and, hence $X_{\mathbb{Z}_{\geqslant 0}}$ is regular; further $\mathbb{E}(\sup\{X_n^2 : n \in \mathbb{Z}_{\geqslant 0}\}) \leqslant 4\mathbb{E}(\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\infty}) \in \mathbb{R}_{\geqslant 0}$.*

(ii) *If $\sqrt{\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\infty}}$ is $\mathbb{P}$-integrable, then the martingale $X_{\mathbb{Z}_{\geqslant 0}}$ is regular and hence converges in $\mathcal{L}^1(\mathbb{P})$; further $\mathbb{E}(\sup\{|X_n| : n \in \mathbb{Z}_{\geqslant 0}\}) \leqslant 3\mathbb{E}(\sqrt{\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\infty}}) \in \mathbb{R}_{\geqslant 0}$.*

(iii) *A $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-stopping time τ with $\mathbb{E}\big(\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\tau}^{1/2}\big) \in \mathbb{R}_{\geqslant 0}$ is regular for the martingale $X_{\mathbb{Z}_{\geqslant 0}}$; further $\mathbb{E}(\sup\{|X_n| : n \in [\![\tau]\!]\}) \leqslant 3\mathbb{E}(\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\tau}^{1/2}) \in \mathbb{R}_{\geqslant 0}$.*

(iv) *In every case the martingale $X_{\mathbb{Z}_{\geqslant 0}}$ converges $\mathbb{P}$-a.e. to a finite limit on the event $\{\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\infty} \in \mathbb{R}_{\geqslant 0}\}$.*

§16.15 **Proof** of Proposition §16.14. is given in the lecture. □

§16.16 **Corollary**. *For a square $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ with square variation process $\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\mathbb{Z}_{\geqslant 0}}$ the following four statements are equivalent:* (C1) $\sup\{\mathbb{E}(X_n^2) : n \in \mathbb{Z}_{\geqslant 0}\} \in \mathbb{R}_{\geqslant 0}$, (C2) $\mathbb{E}(\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\infty}) \in \mathbb{R}_{\geqslant 0}$, (C3) *$X_{\mathbb{Z}_{\geqslant 0}}$ converges in $\mathcal{L}^2(\mathbb{P})$, and* (C4) *$X_{\mathbb{Z}_{\geqslant 0}}$ converges $\mathbb{P}$-a.e. and in $\mathcal{L}^2(\mathbb{P})$.*

§16.17 **Proof** of Corollary §16.16. is given in the lecture. □

§16.18 **Proposition**. *If $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ is a square $\mathbb{P}$-integrable $\mathscr{F}_{\mathbb{Z}_{\geqslant 0}}$-martingale with square variation process $\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\mathbb{Z}_{\geqslant 0}}$ then for any $\alpha \in \mathbb{R}_{>1/2}$,*

$$\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_n^{-\alpha} (X_n - X_0) \xrightarrow{\mathbb{P}\text{-a.e.}} 0 \quad \textit{on } \{\langle X_{\mathbb{Z}_{\geqslant 0}} \rangle_{\infty} = \infty\}.$$

§16.19 **Proof** of Proposition §16.18. is given in the lecture. □

§16.20 **Example** *(Example §16.11 (a) continued).* Let $X_{\mathbb{N}} = (X_n)_{n \in \mathbb{N}}$ be sequence of independent and square $\mathbb{P}$-integrable random variables defined on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{N}}))$, where $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{N}})$ is the filtration generated by $X_{\mathbb{N}}$ (Example §16.11). Consider again the square $\mathbb{P}$-integrable $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{N}})$-martingale $S^{\circ}_{\mathbb{Z}_{\geqslant 0}} := (S^{\circ}_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ with $S^{\circ}_0 := 0$ and $S^{\circ}_n := \sum_{i \in [\![n]\!]} (X_i - \mathbb{E}(X_i))$, $n \in \mathbb{N}$, with square variation process $\langle S^{\circ}_{\mathbb{Z}_{\geqslant 0}} \rangle_{\mathbb{Z}_{\geqslant 0}} = (\langle S^{\circ}_{\mathbb{Z}_{\geqslant 0}} \rangle_n)_{n \in \mathbb{Z}_{\geqslant 0}}$ given by $\langle S^{\circ}_{\mathbb{Z}_{\geqslant 0}} \rangle_0 = 0$ and $\langle S^{\circ}_{\mathbb{Z}_{\geqslant 0}} \rangle_n = \sum_{i \in [\![n]\!]} \mathbb{V}\mathrm{ar}(X_i^2)$ $\mathbb{P}$-a.e. for $n \in \mathbb{N}$. By Proposition §16.18 for any $\alpha \in \mathbb{R}_{>1/2}$ we have

$$\Big(\sum_{i \in [\![n]\!]} \mathbb{V}\mathrm{ar}(X_i)\Big)^{-\alpha} S_n \xrightarrow{\mathbb{P}\text{-a.e.}} 0 \quad \text{whenever } \sum_{i \in \mathbb{N}} \mathbb{V}\mathrm{ar}(X_i) = \infty.$$

In particular, $n^{-\alpha} S^{\circ}_n \xrightarrow{\mathbb{P}\text{-a.e.}} 0$ if $X_{\mathbb{N}}$ is a sequence of i.i.d. square $\mathbb{P}$-integrable random variables. On the other hand, if $(a_n)_{n \in \mathbb{N}}$ is an increasing and diverging sequence of real numbers, then by Kronecker's Lemma, $a_n^{-1} \sum_{i \in [\![n]\!]} y_i = \mathrm{o}(1)$ for any sequence $(y_n)_{n \in \mathbb{N}}$ of real numbers with $\sum_{i\in\mathbb{N}} a_i^{-1} y_i \in \mathbb{R}_{\geqslant 0}$. Thereby, if $\sum_{i\in\mathbb{N}} a_i^{-2}\mathbb{V}\mathrm{ar}(X_i) \in \mathbb{R}_{\geqslant 0}$, then by Corollary §16.16 the $\mathbb{P}$-integrable $\sigma_{\mathbb{N}}(X_{\mathbb{N}})$-martingale $(\sum_{i\in[\![n]\!]} a_i^{-1}(X_i - \mathbb{E}(X_i)))_{n\in\mathbb{N}}$ converges $\mathbb{P}$-a.e. to a finite limit and by Kronecker's Lemma, $a_n^{-1}\sum_{i\in[\![n]\!]}(X_i - \mathbb{E}(X_i)) \xrightarrow{\mathbb{P}\text{-a.e.}} 0$. In case of independent and identically distributed (i.i.d.) random variables $(X_n)_{n\in\mathbb{N}}$ we find $n^{-1}\sum_{i\in[\![n]\!]}(X_i - \mathbb{E}(X_i)) \xrightarrow{\mathbb{P}\text{-a.e.}} 0$. □

Chapter 5

Markov chains

§17 Markov chains

§17.01 **Definition**. Let $\mathcal{T} = \mathbb{Z}_{\geqslant 0}$ (discrete time) or $\mathcal{T} = \mathbb{R}_{\geqslant 0}$ (continuous time), $(\Omega, \mathscr{A}, \mathbb{P})$ a probability space and $(\mathcal{X}, 2^{\mathcal{X}})$ a countable nonempty state space. A stochastic process $(X_t)_{t\in\mathcal{T}}$ on $(\Omega, \mathscr{A}, \mathbb{P})$ with state space $(\mathcal{X}, 2^{\mathcal{X}})$ is called *Markov chain* if for all $n \in \mathbb{N}$, for all $t_1 < t_2 < \ldots < t_n < t$ in $\mathcal{T}$ and for all $x_1, \ldots, x_n, x$ in $\mathcal{X}$ with $\mathbb{P}(X_{t_1} = x_1, \ldots, X_{t_n} = x_n) \in \mathbb{R}_{>0}$ the *Markov property* is satisfied:

$$\mathbb{P}\big(X_t = x \big| X_{t_1} = x_1, \ldots, X_{t_n} = x_n\big) := \mathbb{P}\big(X_t^{-1}(\{x\}) \big| \bigcap_{i\in[\![n]\!]} X_{t_i}^{-1}(\{x_i\})\big)$$
$$= \mathbb{P}\big(X_t^{-1}(\{x\}) \big| X_{t_n}^{-1}(\{x_n\})\big) =: \mathbb{P}\big(X_t = x \big| X_{t_n} = x_n\big).$$

For a Markov chain $(X_t)_{t\in\mathcal{T}}$ and $t_1 \in \mathcal{T}$, $t_2 \in \mathcal{T}_{\geqslant t_1}$, $x_1, x_2 \in \mathcal{X}$ the *transition probability* to reach state x_2 at time t_2 from state x_1 at time t_1 is for $\mathbb{P}(X_{t_1} = x_1) \in \mathbb{R}_{>0}$ defined by

$$\mathrm{P}_{x_1,x_2}(t_1, t_2) := \mathbb{P}\big(X_{t_2} = x_2 \big| X_{t_1} = x_1\big)$$

(and $\mathrm{P}_{x_1,x_2}(t_1, t_2) := \mathbb{P}(X_{t_2} = x_2)$ if $\mathbb{P}(X_{t_1} = x_1) = 0$). The *transition matrix* is given by

$$\mathrm{P}_{\bullet,\bullet}(t_1 t_2) := \big[\mathrm{P}_{x_1,x_2}(t_1, t_2)\big] = \big(\mathrm{P}_{x_1,x_2}(t_1, t_2)\big)_{x_1,x_2\in\mathcal{X}}.$$

The transition matrix and the Markov chain are called *time-homogeneous* if

$$\mathrm{P}_{\bullet,\bullet}(t_1, t_2) = \mathrm{P}_{\bullet,\bullet}(0, t_2 - t_1) =: \mathrm{P}_{\bullet,\bullet}(t_2 - t_1)$$

holds for all $t_1 \in \mathcal{T}$, $t_2 \in \mathcal{T}_{\geqslant t_1}$. □

§17.02 **Lemma**. *The transition matrices given in Definition §17.01 satisfy the Chapman-Kolmogorov equation, that is, (matrix multiplication)*

$$\mathrm{P}_{\bullet,\bullet}(t_1, t_3) = \mathrm{P}_{\bullet,\bullet}(t_1, t_2)\mathrm{P}_{\bullet,\bullet}(t_2, t_3) \quad \textit{for any } t_1 \leqslant t_2 \leqslant t_3 \textit{ in } \mathcal{T}.$$

In the time-homogeneous case this gives the semigroup property

$$\mathrm{P}_{\bullet,\bullet}(t_1 + t_2) = \mathrm{P}_{\bullet,\bullet}(t_1)\mathrm{P}_{\bullet,\bullet}(t_2) \quad \textit{for all } t_1, t_2 \in \mathcal{T},$$

and in particular with one step transition matrix $\mathrm{P}_{\bullet,\bullet} := \mathrm{P}_{\bullet,\bullet}(1)$ *also* $\mathrm{P}_{\bullet,\bullet}(n) = \mathrm{P}_{\bullet,\bullet}^n$ *for all* $n \in \mathbb{N}$.

§17.03 **Proof** of Lemma §17.02. is given in the lecture. □

§17.04 **Remark**. Consider the one step transition matrix $\mathrm{P}_{\bullet,\bullet}$ of a time-homogeneous Markov chain as defined in Lemma §17.02. Identifying the map $\mathrm{P}_{\bullet,\bullet} : \mathcal{X} \times \mathcal{X} \to [0, 1]$ with $(x, y) \mapsto \mathrm{P}_{x,y}$ and the (infinite dimensional) matrix $\mathrm{P}_{\bullet,\bullet} := (\mathrm{P}_{x,y})_{x,y\in\mathcal{X}}$, then $\mathrm{P}_{\bullet,\bullet}$ is a regular version of the conditional $\zeta_{\mathcal{X}}$-mass function of $\mathbb{P}^{X_1}$ given X_0 (compare Remark §07.19 (a)). In particular, the Markov kernel $\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}$ from $(\mathcal{X}, 2^{\mathcal{X}})$ to $(\mathcal{X}, 2^{\mathcal{X}})$ with $\zeta_{\mathcal{X}}$-mass function $\mathrm{P}_{\bullet,\bullet}$, that is,

$$\mathcal{X} \times 2^{\mathcal{X}} \ni (x, B) \mapsto \mathrm{P}_{x,\bullet}\zeta_{\mathcal{X}}(B) = \zeta_{\mathcal{X}}(\mathbb{1}_B \mathrm{P}_{x,\bullet}) = \int_B \mathrm{P}_{x,\bullet}\mathrm{d}\zeta_{\mathcal{X}} = \sum_{y\in B} \mathrm{P}_{x,y} \in [0, 1]$$

is a regular version of the conditional distribution of $\mathbb{P}^{X_1}$ given X_0, i.e. $\mathbb{P}^{X_1|X_0} = \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}$. Evidently, the joint distribution of (X_0, X_1) is given by $\mathbb{P}^{(X_0,X_1)} = \mathbb{P}^{X_0} \odot \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}} \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}^2})$ where

$$\mathbb{P}^{(X_0,X_1)}(B_0 \times B_1) = \mathbb{P}^{X_0} \odot \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(B_0 \times B_1) = \sum_{x_0 \in B_0} \mathbb{P}^{X_0}(\{x_0\}) \sum_{x_1 \in B_1} \mathrm{P}_{x_0,x_1} \quad \text{for } B_0, B_1 \in 2^{\mathcal{X}}.$$

For $n \in \mathbb{N}$ the joint distribution of $X_{[\![0,n]\!]} = (X_i)_{i \in [\![0,n]\!]}$ is then given iteratively by

$$\mathbb{P}^{X_{[\![0,n]\!]}} = \mathbb{P}^{X_{[\![0,n-1]\!]}} \odot \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}} = \mathbb{P}^{X_0} \odot \bigodot_{i \in [\![n]\!]} \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}} \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}^{n+1}})$$

where for $B_i \in 2^{\mathcal{X}}$, $i \in [\![0,n]\!]$ we have

$$\mathbb{P}(X_0 \in B_0, \dots, X_n \in B_n) = \mathbb{P}^{X_{[\![0,n]\!]}}(\times_{i \in [\![0,n]\!]} B_i) = \sum_{x_0 \in B_0} \mathbb{P}^{X_0}(\{x_0\}) \sum_{x_1 \in B_1} \mathrm{P}_{x_0,x_1} \cdots \sum_{x_n \in B_n} \mathrm{P}_{x_{n-1},x_n}.$$

For each finite $\mathcal{J} \subseteq \mathbb{Z}_{\geqslant 0}$ and $n \in \mathbb{N}$ with $\mathcal{J} \subseteq [\![0,n]\!]$ we have $X_{\mathcal{J}} = (X_i)_{i \in \mathcal{J}} = \Pi_{\mathcal{J}}^{[\![0,n]\!]} \circ X_{[\![0,n]\!]}$, and hence $\mathbb{P}^{X_{\mathcal{J}}} = \mathbb{P}^{X_{[\![0,n]\!]}} \circ (\Pi_{\mathcal{J}}^{[\![0,n]\!]})^{-1}$. Evidently, the family $\{\mathbb{P}^{X_{\mathcal{J}}} : \mathcal{J} \subseteq \mathbb{Z}_{\geqslant 0} \text{ finite}\}$ of probability measures is consistent and determines by *Kolmogorov's consistency theorem* §04.33 the unique probability measure $\mathbb{P}^{X_{\mathbb{Z}_{\geqslant 0}}} \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$ on $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$. □

§17.05 **Preliminary**. Consider now an arbitrary one step transition matrix $\mathrm{P}_{\bullet,\bullet}$ and an initial (discrete) probability measure $\mu \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$ on $(\mathcal{X}, 2^{\mathcal{X}})$. We eventually define for each $n \in \mathbb{N}$ a probability measure

$$\mathbb{P}_{[\![0,n]\!]} := \mu \odot \bigodot_{i \in [\![n]\!]} \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}} \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}^{n+1}})$$

on $(\mathcal{X}^{n+1}, 2^{\mathcal{X}^{n+1}})$ given by

$$\mathbb{P}_{[\![0,n]\!]}(\times_{i \in [\![0,n]\!]} B_i) := \sum_{x_0 \in B_0} \mu(\{x_0\}) \sum_{x_1 \in B_1} \mathrm{P}_{x_0,x_1} \cdots \sum_{x_n \in B_n} \mathrm{P}_{x_{n-1},x_n} \quad \text{for } B_i \in 2^{\mathcal{X}}, i \in [\![0,n]\!].$$

For each finite $\mathcal{J} \subseteq \mathbb{Z}_{\geqslant 0}$ and $n \in \mathbb{N}$ with $\mathcal{J} \subseteq [\![0,n]\!]$ we set $\mathbb{P}_{\mathcal{J}} := \mathbb{P}_{[\![0,n]\!]} \circ (\Pi_{\mathcal{J}}^{[\![0,n]\!]})^{-1}$. Therewith, we obtain a consistent family $\{\mathbb{P}_{\mathcal{J}} : \mathcal{J} \subseteq \mathbb{Z}_{\geqslant 0} \text{ finite}\}$ of probability measures on the product space $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$ which determines by *Kolmogorov's consistency theorem* §04.33 a unique probability measure $\mathbb{P}_{\mu} \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$ on $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$. For $\mu = \delta_x$, a point mass at $x \in \mathcal{X}$ (see Example §01.18), we use $\mathbb{P}_x$ as an abbreviation for $\mathbb{P}_{\delta_x}$ where for every initial probability measure μ and for every $A \in 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}}$ evidently holds

$$\mathbb{P}_{\mu}(A) = \sum_{x \in \mathcal{X}} \mu(\{x\}) \mathbb{P}_x(A).$$

Consider the canonical process $\Pi_{\mathbb{Z}_{\geqslant 0}} = (\Pi_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ on $(\Omega = \mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{A} = 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$, where

$$\Pi_z : \mathcal{X}^{\mathbb{Z}_{\geqslant 0}} \ni x_{\mathbb{Z}_{\geqslant 0}} = (x_n)_{n \in \mathbb{Z}_{\geqslant 0}} \mapsto \Pi_z(x_{\mathbb{Z}_{\geqslant 0}}) := x_z \in \mathcal{X}$$

for any $z \in \mathbb{Z}_{\geqslant 0}$ (see Remark §09.03). For the family of probability measures $(\mathbb{P}_x)_{x \in \mathcal{X}}$ on $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$ the canonical process $\Pi_{\mathbb{Z}_{\geqslant 0}}$ on the probability space $(\Omega = \mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{A} = 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}}, \mathbb{P}_x)$ is then a time-homogeneous Markov chain with one step transition matrix $\mathrm{P}_{\bullet,\bullet}$ and marginal distribution $\mathbb{P}_x^{\Pi_0} = \delta_x$ of Π_0 for every $x \in \mathcal{X}$. □

§17.06 **Definition**. A stochastic process $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ on $(\Omega, \mathscr{A})$ with values in a countable state space $(\mathcal{X}, 2^{\mathcal{X}})$ is called a *time-homogeneous Markov chain* with family of probability measures $(\mathbb{P}_x)_{x\in\mathcal{X}}$ in $\mathfrak{M}_{\mathrm{pr}}(\mathscr{A})$, if

(a) For every $x \in \mathcal{X}$, $X_{\mathbb{Z}_{\geqslant 0}}$ is a stochastic process on the probability space $(\Omega, \mathscr{A}, \mathbb{P}_x)$ with marginal distribution $\mathbb{P}_x(X_0 = x) = 1$.

(b) The map

$$\kappa : \mathcal{X} \times 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}} \ni (x, A) \mapsto \kappa(x, A) := \mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}(A) = \mathbb{P}_x(X_{\mathbb{Z}_{\geqslant 0}} \in A) \in [0, 1]$$

is a Markov kernel. Then for every $z \in \mathbb{Z}_{\geqslant 0}$, the map

$$\begin{aligned}\kappa_z : \mathcal{X} \times 2^{\mathcal{X}} \ni (x, B) \mapsto \kappa_z(x, B) &:= \kappa(x, \Pi_z^{-1}(B))\\ &= \mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}(\Pi_z^{-1}(B)) = \mathbb{P}_x^{X_n}(B) = \mathbb{P}_x(X_n \in B) \in [0, 1]\end{aligned}$$

is also a Markov kernel

(c) For every $x \in \mathcal{X}$, $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ has on $(\Omega, \mathscr{A}, \mathbb{P}_x)$ with respect to its natural filtration $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{Z}_{\geqslant 0}}) = \big(\sigma_z(X_{\mathbb{Z}_{\geqslant 0}}) = \sigma(\{X_n : n \in [\![0, z]\!]\})\big)_{z\in\mathbb{Z}_{\geqslant 0}}$ the *time-homogeneous Markov property*: For every $x, y \in \mathcal{X}$ and all $n, z \in \mathbb{Z}_{\geqslant 0}$ we have

$$\mathbb{P}_x\big(X_{n+z} = y \big| \sigma_n(X_{\mathbb{Z}_{\geqslant 0}})\big) = \kappa_z(X_n, \{y\}) \quad \mathbb{P}_x\text{-a.e.}$$

with n-step transition matrix $\mathrm{P}^n_{\bullet,\bullet} = (\mathrm{P}^n_{x,y})_{x,y\in\mathcal{X}}$ of $X_{\mathbb{Z}_{\geqslant 0}}$ given by

$$\mathrm{P}^n_{x,y} = \kappa_n(x, \{y\}) = \mathbb{P}_x(X_n = y) \qquad \text{for all } x, y \in \mathcal{X}.$$

We write $\mathbb{E}_x$ for the expectation with respect to $\mathbb{P}_x$, $\mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}\big(\bullet \,\big|\mathscr{A}\big)$ for a regular version of the conditional distribution of $\mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}$ given $\mathscr{A}$ and $\mathbb{E}_x\big(f(X_{\mathbb{Z}_{\geqslant 0}})\big|\mathscr{A}\big) = \mathbb{E}_x^{X_{\mathbb{Z}_{\geqslant 0}}}\big(f\big|\mathscr{A}\big) = \mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}\big(f\big|\mathscr{A}\big)$ for a conditional expectation of $f(X_{\mathbb{Z}_{\geqslant 0}})$ given $\mathscr{A}$. In particular, we use the notation $\mathbb{P}_{X_m}^{X_{\mathbb{Z}_{\geqslant 0}}} = \kappa(X_m, \bullet)$, that is, we understand X_m as the initial value of a second Markov chain with the same family of probability measures $(\mathbb{P}_x)_{x\in\mathcal{X}}$. □

§17.07 **Remark**. The existence of the family $(\kappa_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ of Markov kernels implies the existence of the Markov kernel κ (cf. Klenke (2020, Theorem 17.8, p.391)). Thus, a time-homogeneous Markov chain is simply a stochastic process with the Markov property and for which the transition probabilities are time-homogeneous. □

End LE19 2026/06/19

§17.08 **Reminder**. For every $z \in \mathbb{Z}_{\geqslant 0}$,

$$\vartheta_z : \mathcal{X}^{\mathbb{Z}_{\geqslant 0}} \ni x_{\mathbb{Z}_{\geqslant 0}} = (x_m)_{m\in\mathbb{Z}_{\geqslant 0}} \mapsto \vartheta_z(x_{\mathbb{Z}_{\geqslant 0}}) := (x_{m+z})_{m\in\mathbb{Z}_{\geqslant 0}} \in \mathcal{X}^{\mathbb{Z}_{\geqslant 0}}$$

is called *shift*(-operator) (see Definition §09.08). □

§17.09 **Theorem**. *A stochastic process $X_{\mathbb{Z}_{\geqslant 0}} = (X_n)_{n\in\mathbb{Z}_{\geqslant 0}}$ is a time-homogeneous Markov chain if and only if for every $z \in \mathbb{Z}_{\geqslant 0}$ and $x \in \mathcal{X}$,*

$$\mathbb{P}_x^{\vartheta_z(X_{\mathbb{Z}_{\geqslant 0}})}\big(\bullet \,\big|\sigma_z(X_{\mathbb{Z}_{\geqslant 0}})\big) = \mathbb{P}_{X_z}^{X_{\mathbb{Z}_{\geqslant 0}}} \quad \mathbb{P}_x\text{-a.e.}$$

if and only if there exists a Markov kernel $\kappa : \mathcal{X} \times 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}} \to [0, 1]$ such that, for every bounded $2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}}$-measurable function $f \in \mathcal{M}(2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$ and for every $z \in \mathbb{Z}_{\geqslant 0}$ and $x \in \mathcal{X}$, we have

$$\mathbb{E}_x\big(f(\vartheta_z(X_{\mathbb{Z}_{\geqslant 0}}))\big|\sigma_z(X_{\mathbb{Z}_{\geqslant 0}})\big) = \mathbb{E}_{X_z}[f(X_{\mathbb{Z}_{\geqslant 0}})] := \int f(x_{\mathbb{Z}_{\geqslant 0}})\kappa(X_z, \mathrm{d}x_{\mathbb{Z}_{\geqslant 0}}) \quad \mathbb{P}_x\text{-a.e.}$$

§17.10 **Proof** of Theorem §17.09. Klenke (2020, Theorem 17.9, p.392 and Corollary 17.10, p.393) □

§17.11 **Definition**. A time-homogeneous Markov chain $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ with family of probability measures $(\mathbb{P}_x)_{x\in\mathcal{X}}$ has the *strong Markov property* if, for every $x \in \mathcal{X}$, and every $\mathbb{P}_x$-a.e. finite $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{Z}_{\geqslant 0}})$-stopping time τ we have

$$\mathbb{P}_x^{\vartheta_\tau(X_{\mathbb{Z}_{\geqslant 0}})}\big(\bullet\,\big|\sigma_\tau(X_{\mathbb{Z}_{\geqslant 0}})\big) = \mathbb{P}_{X_\tau}^{X_{\mathbb{Z}_{\geqslant 0}}} := \kappa(X_\tau, \bullet) \quad \mathbb{P}_x\text{-a.e.}$$

or equivalently for every bounded $2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}}$-measurable function $f \in \mathcal{M}(2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}})$ we have

$$\mathbb{E}_x\Big(f(\vartheta_\tau(X_{\mathbb{Z}_{\geqslant 0}}))\Big|\sigma_\tau(X_{\mathbb{Z}_{\geqslant 0}})\Big) = \mathbb{E}_{X_\tau}(f(X_{\mathbb{Z}_{\geqslant 0}})) = \int f(x_{\mathbb{Z}_{\geqslant 0}} m)\kappa(X_\tau, \mathrm{d}x_{\mathbb{Z}_{\geqslant 0}}) \quad \mathbb{P}_x\text{-a.e.}$$ □

§17.12 **Lemma**. *Every time-homogeneous Markov chain $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ has the strong Markov property.*

§17.13 **Proof** of Lemma §17.12. is given in the lecture. □

§18 Recurrence and transience

In the sequel, let $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ be a time-homogeneous Markov chain with values in a countable state space $(\mathcal{X}, 2^{\mathcal{X}})$, family of probability measures $(\mathbb{P}_x)_{x\in\mathcal{X}}$ on $(\Omega, \mathscr{A})$ and n-step transition matrix $\mathrm{P}^n_{\bullet,\bullet} := (\mathrm{P}^n_{x,y})_{x,y\in\mathcal{X}}$ for $n \in \mathbb{N}$.

§18.01 **Definition**. For $y \in \mathcal{X}$, $z \in \mathbb{Z}_{\geqslant 0}$ introduce for $x_{\mathbb{Z}_{\geqslant 0}} \in \mathcal{X}^{\mathbb{Z}_{\geqslant 0}}$ the *z-th time $\tau_y^z(x_{\mathbb{Z}_{\geqslant 0}}) \in \overline{\mathbb{Z}}_{\geqslant 0}$ of return to y* recursively by

$$\tau_y^{z+1}(x_{\mathbb{Z}_{\geqslant 0}}) := \inf\{n \in \mathbb{N}_{>\tau_y^z(x_{\mathbb{Z}_{\geqslant 0}})} : x_n = y\} \quad \text{and} \quad \tau_y^0(x_{\mathbb{Z}_{\geqslant 0}}) := 0$$

(where $\inf \varnothing = \infty$). We set $\tau_y := \tau_y^1$ and for $x \in \mathcal{X}$, $\rho_{x,y} := \mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}(\tau_y \in \mathbb{N}) = \mathbb{P}_x(\tau_y(X_{\mathbb{Z}_{\geqslant 0}}) \in \mathbb{N})$. □

§18.02 **Remark**. Evidently, $\rho_{x,y} = \mathbb{P}_x(\text{ there is an } k \in \mathbb{N} \text{ with } X_k = y)$ is the probability of ever going from x to y. In particular, if $\rho_{x,y} \in (0,1]$ then there is an $k \in \mathbb{N}$ with $\mathbb{P}_x(X_k = y) = \mathrm{P}^k_{x,y} \in (0,1]$. Moreover, $\rho_{x,x}$ is the return probability (after the first jump) from x to x where $\tau_x(x_{\mathbb{Z}_{\geqslant 0}}) \in \overline{\mathbb{N}}$ even if the chain $x_{\mathbb{Z}_{\geqslant 0}}$ starts at $x_0 = x$. □

§18.03 **Definition**. A state $x \in \mathcal{X}$ is called

recurrent if $\rho_{x,x} = 1$,

positive recurrent if $\mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}(\tau_x) \in \mathbb{R}_{\geqslant 0}$,

null recurrent if x is recurrent but not positive recurrent,

transient if $\rho_{x,x} \in [0,1)$, and

absorbing if $\mathrm{P}_{x,x} = 1$.

The Markov chain $X_{\mathbb{Z}_{\geqslant 0}}$ is called (positive, null) recurrent if every state $x \in \mathcal{X}$ is (positive, null) recurrent and is called transient if every recurrent state is absorbing. □

§18.04 **Remark**. Clearly, we have: "absorbing" ⇒ "positive recurrent" ⇒ "recurrent". □

§18.05 **Lemma**. *For $k \in \mathbb{N}$ and $x, y \in \mathcal{X}$ we have $\mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}(\tau_y^k \in \mathbb{N}) = \rho_{x,y}\rho_{y,y}^{k-1}$.*

§18.06 **Proof** of Lemma §18.05. is given in the lecture. □

§18.07 **Definition**. For $y \in \mathcal{X}$ and $x_{\mathbb{Z}_{\geq 0}} \in \mathcal{X}^{\mathbb{Z}_{\geq 0}}$ denote by $N_y(x_{\mathbb{Z}_{\geq 0}}) := \sum_{n \in \mathbb{Z}_{\geq 0}} \mathbb{1}_{\{x_n = y\}}$ the total *number of visits* of $x_{\mathbb{Z}_{\geq 0}}$ to state y and by $\mathrm{G}_{\bullet,\bullet} := (\mathrm{G}_{x,y})_{x,y \in \mathcal{X}}$ given by

$$\mathrm{G}_{x,y} = \mathbb{P}_x^{X_{\mathbb{Z}_{\geq 0}}}(N_y) = \mathbb{P}_x(N_y(X_{\mathbb{Z}_{\geq 0}})) = \sum_{n \in \mathbb{Z}_{\geq 0}} \mathbb{P}_x(X_n = y) = \sum_{n \in \mathbb{Z}_{\geq 0}} \mathrm{P}^n_{x,y} \in \overline{\mathbb{Z}}_{\geq 0}$$

the *Green function* of $X_{\mathbb{Z}_{\geq 0}}$. □

§18.08 **Lemma**.

(i) *A state $x \in \mathcal{X}$ is recurrent if and only if $\mathrm{G}_{x,x} = \infty$;*

(ii) *If a state $x \in \mathcal{X}$ is transient then for all $y \in \mathcal{X}$, $\mathrm{G}_{y,x} \in \mathbb{R}_{\geq 0}$ with*

$$\mathrm{G}_{y,x} = \left\{ \begin{array}{ll} \dfrac{\rho_{y,x}}{1-\rho_{x,x}}, & \text{if } x \neq y \\ \dfrac{1}{1-\rho_{x}}, & \text{otherwise} \end{array} \right\} = \frac{\rho_{y,x}}{1-\rho_{x,x}} + \mathbb{1}_{\{x=y\}}.$$

§18.09 **Proof** of Lemma §18.08. is given in the lecture. □

§18.10 **Proposition**. *If a state $x \in \mathcal{X}$ is recurrent and $\rho_{x,y} \in (0,1]$ for some $y \in \mathcal{X}$, then the state y is also recurrent, and $\rho_{x,y} = \rho_{y,x} = 1$.*

§18.11 **Proof** of Proposition §18.10. is given in the lecture. □

§18.12 **Definition**. A subset $B \subseteq \mathcal{X}$ of states is called

closed, if $\rho_{x,y} = 0$ for all $x \in B$ and $y \in B^c = \mathcal{X} \setminus B$,

irreducible, if $\rho_{x,y} \in (0,1]$ for all $x, y \in B$.

If the state space $\mathcal{X}$ is irreducible, then the Markov chain is called *irreducible*. □

§18.13 **Corollary**. *An irreducible Markov chain is either recurrent or transient. If in addition $|\mathcal{X}| \in \overline{\mathbb{N}}_{\geq 2}$, then there is no absorbing state.*

§18.14 **Proof** of Corollary §18.13. The result is an immediate consequence of Proposition §18.10. □

§18.15 **Proposition**. *For an irreducible Markov chain on a finite state space $\mathcal{X}$ all states are recurrent.*

§18.16 **Proof** of Proposition §18.15. is given in the lecture. □

§19 Invariant distribution

In the following, let $\mathrm{P}_{\bullet,\bullet} = (\mathrm{P}_{x,y})_{x,y \in \mathcal{X}}$ be a transition matrix on a countable state space $\mathcal{X}$, and let $X_{\mathbb{Z}_{\geq 0}} = (X_n)_{n \in \mathbb{Z}_{\geq 0}}$ be a Markov chain with with one step transition matrix $\mathrm{P}_{\bullet,\bullet}$. Then $\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}$ is a Markov kernel from $(\mathcal{X}, 2^{\mathcal{X}})$ to $(\mathcal{X}, 2^{\mathcal{X}})$ with

$$(\mathcal{X}, 2^{\mathcal{X}}) \ni (x, B) \mapsto \mathrm{P}_{x,\bullet}\zeta_{\mathcal{X}}(B) = \zeta_{\mathcal{X}}(\mathbb{1}_B \mathrm{P}_{x,\bullet}) = \int_B \mathrm{P}_{x,\bullet} \mathrm{d}\zeta_{\mathcal{X}} = \sum_{y \in B} \mathrm{P}_{x,y} \in [0,1]$$

(compare Notation §04.61 and Remark §17.04).

§19.01 **Definition**. Let $\mu \in \mathfrak{M}(2^{\mathcal{X}})$ be a measure on $(\mathcal{X}, 2^{\mathcal{X}})$ and let $f : \mathcal{X} \to \mathbb{R}$ be a map. We define the measure $\mu\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}} \in \mathfrak{M}(2^{\mathcal{X}})$ given by

$$2^{\mathcal{X}} \ni A \mapsto (\mu\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}})(A) := \int \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(A)\mathrm{d}\mu = \sum_{x\in\mathcal{X}} \mu(\{x\})\mathrm{P}_{x,\bullet}\zeta_{\mathcal{X}}(A) = \sum_{x\in\mathcal{X}} \mu(\{x\}) \sum_{y\in A} \mathrm{P}_{x,y} \in \mathbb{R}_{\geqslant 0}.$$

If f is $\mathrm{P}_{x,\bullet}\zeta_{\mathcal{X}}$-quasiintegrable for any $x \in \mathcal{X}$, or $\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}$-quasiintegrable for short, we define the function $\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(f) \in \overline{\mathcal{M}}(2^{\mathcal{X}})$ given by

$$\mathcal{X} \ni x \mapsto (\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(f))(x) := \mathrm{P}_{x,\bullet}\zeta_{\mathcal{X}}(f) = \int \mathrm{P}_{x,\bullet} f \mathrm{d}\zeta_{\mathcal{X}} = \sum_{y\in\mathcal{X}} \mathrm{P}_{x,y} f(y)$$

(compare Notation §04.61 and Remark §17.04). □

§19.02 **Definition**.

(a) A σ-finite measure $\mu \in \mathfrak{M}_\sigma(2^{\mathcal{X}})$ on $(\mathcal{X}, 2^{\mathcal{X}})$ is called *invariant measure*, if $\mu\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}} = \mu$. A probability measure that is an invariant measure is called an *invariant distribution*. Denote by $\mathcal{I}$ the set of invariant distributions.

(b) A $\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}$-quasiintegrable function $f : \mathcal{X} \to \mathbb{R}$ is called *subharmonic* if $f \leqslant \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(f)$ that is $f(x) \leqslant (\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(f))(x) = \mathrm{P}_{x,\bullet}\zeta_{\mathcal{X}}(f)$ for all $x \in \mathcal{X}$. f is called *superharmonic*, if $f \geqslant \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(f)$ and *harmonic* if $f = \mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}(f)$. □

§19.03 **Remark**. In the terminology of linear algebra, an invariant measure is a left eigenvector of $\mathrm{P}_{\bullet,\bullet}$ corresponding to the eigenvalue 1. A harmonic function is a right eigenvector corresponding to the eigenvalue 1. □

§19.04 **Comment**.

(a) If $\mathcal{X}$ is finite then any invariant measure can be normalised to an invariant distribution.

(b) Any invariant measure μ for $\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}$ is also an invariant measure for $\mathrm{P}^n_{\bullet,\bullet}\zeta_{\mathcal{X}}$ for every $n \in \mathbb{N}$.

(c) If $X_{\mathbb{Z}_{\geqslant 0}}$ is irreducible and $\mu \neq 0$ an invariant measure for $\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}}$ then $\mu(\{x\}) \in \mathbb{R}_{>0}$ for every $x \in \mathcal{X}$. Since $\mu \in \mathfrak{M}_\sigma(2^{\mathcal{X}})$ is σ-finite we have $\mu(\{x\}) \in \mathbb{R}_{\geqslant 0}$ for every $x \in \mathcal{X}$. Let $x \in \mathcal{X}$ with $\mu(\{x\}) \in \mathbb{R}_{>0}$ (since $\mu \neq 0$) and $y \in \mathcal{X}$ is arbitrary, then there exists $n \in \mathbb{N}$ with $\mathrm{P}^n_{x,y} \in (0,1]$ (since $X_{\mathbb{Z}_{\geqslant 0}}$ is irreducible and hence $\rho_{x,y} \in (0,1]$) and

$$\mu(\{y\}) = (\mu\mathrm{P}^n_{\bullet,\bullet}\zeta_{\mathcal{X}})(\{y\}) = \sum_{\widetilde{x}\in\mathcal{X}} \mu(\{\widetilde{x}\})\mathrm{P}^n_{\widetilde{x},y} \geqslant \mu(\{x\})\mathrm{P}^n_{x,y} \in \mathbb{R}_{>0}.$$

(d) If π is an invariant distribution, then for every $n \in \mathbb{Z}_{\geqslant 0}$ and $y \in \mathcal{X}$

$$\mathbb{P}_\pi(X_n = y) = (\pi\mathrm{P}^n_{\bullet,\bullet}\zeta_{\mathcal{X}})(\{y\}) = (\pi\mathrm{P}_{\bullet,\bullet}\zeta_{\mathcal{X}})(\{y\}) = \sum_{x\in\mathcal{X}} \pi(\{x\})\mathrm{P}_{x,y} = \pi(\{y\}).$$

The chain started with distribution π has the same marginal distribution π at every point n in time. □

End LE20
2026/06/24

§19.05 **Lemma**. *If $f : \mathcal{X} \to \mathbb{R}$ is bounded and (sub-, super-) harmonic, then for every $x \in \mathcal{X}$ on $(\Omega, \mathscr{A}, \mathbb{P}_x, \sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{Z}_{\geqslant 0}}))$, the stochastic process $(f(X_n))_{n\in\mathbb{Z}_{\geqslant 0}}$ is a (sub-, super-) martingale with respect to the natural filtration $\sigma_{\mathbb{Z}_{\geqslant 0}}(X_{\mathbb{Z}_{\geqslant 0}})$ generated by $X_{\mathbb{Z}_{\geqslant 0}}$.*

§19.06 **Proof** of Lemma §19.05. is given in the lecture. □

§19.07 **Theorem**. *Let $X_{\mathbb{Z}_{\geqslant 0}}$ be irreducible and recurrent, $x \in \mathcal{X}$, $\tau_x := \inf\{n \in \mathbb{N} : X_n = x\}$ and*

$$\mu_x(\{y\}) := \mathbb{P}_x\Big(\sum_{n \in [\![\tau_x]\!]} \mathbb{1}_{\{X_n = y\}}\Big) \quad \textit{for all } y \in \mathcal{X}.$$

Then the following three statements hold true

(i) $\mu_x = \mu_x \mathrm{P}_{\bullet,\bullet} \zeta_{\mathcal{X}}$

(ii) $\mu_x(\{y\}) \in \mathbb{R}_{>0}$ *for all* $y \in \mathcal{X}$,

(iii) μ_x *is the unique invariant measure with value* $\mu_x(\{x\}) = 1$.

§19.08 **Proof** of Theorem §19.07. is given in the lecture. □

§19.09 **Theorem**. *Let $X_{\mathbb{Z}_{\geqslant 0}}$ be irreducible, then the following three statements are equivalent:*

(iD1) *There exists an invariant distribution* $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$.

(iD2) *There exists a positive recurrent state in* $\mathcal{X}$.

(iD3) *All states in* $\mathcal{X}$ *are positive recurrent.*

If one of the equivalent statements is satisfied, then the invariant distribution π *is uniquely determined by* $\pi(\{x\}) = \frac{1}{\mathbb{E}_x(\tau_x)}$ *for all* $x \in \mathcal{X}$.

§19.10 **Proof** of Theorem §19.09. is given in the lecture. □

§19.11 **Remark**. An irreducible Markov chain belongs to exactly one of the following three cases:

transient: The chain is transient, i.e., for all $x \in \mathcal{X}$ it holds that $\mathbb{P}_x(\tau_x \in \mathbb{N}) \in [0,1)$, and there is no invariant measure.

null recurrent: The chain is null recurrent, i.e., for all $x \in \mathcal{X}$ it holds that $\mathbb{P}_x(\tau_x \in \mathbb{N}) = 1$ and $\mathbb{E}_x(\tau_x) = \infty$, and there is an invariant measure but not an invariant distribution.

positive recurrent: The chain is positive recurrent, i.e., for all $x \in \mathcal{X}$ it holds that $\mathbb{E}_x(\tau_x) \in \mathbb{R}_{\geqslant 0}$, and there is an invariant distribution.

If the state space is finite, only the positive recurrent case occurs. □

We investigate under which conditions a positive recurrent Markov chain $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ with values in a countable state space $\mathcal{X}$ and with family of probability measures $(\mathbb{P}_x)_{x \in \mathcal{X}}$ on $(\Omega, \mathscr{A})$, started at any point $x \in \mathcal{X}$, converges to its invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$ as the runtime diverges, i.e., $\mathbb{P}_x^{X_n}(\{y\}) = \mathbb{P}_x(X_n = y)$ converges to $\pi(\{y\})$ as $n \to \infty$ for every $y \in \mathcal{X}$.

§19.12 **Remark**. If $\mathbb{P}_x^{X_n}(\{y\}) \xrightarrow{n \to \infty} \pi(\{y\})$ for all $x, y \in \mathcal{X}$, then it follows for all $\mu \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$ by dominated convergence that $\mathbb{P}_\mu^{X_n}(\{y\}) \xrightarrow{n \to \infty} \pi(\{y\})$ for all $y \in \mathcal{X}$, as well as $\mathbb{P}_\mu^{X_n}(h) \xrightarrow{n \to \infty} \pi(h)$ for every bounded function $h : \mathcal{X} \to \mathbb{R}$. □

§19.13 **Comment**. In the lecture course EWS (*Beispiel* §30.24 *(b)*) we have shown (keep in mind that $\mathcal{X}$ is countable), that $\mathbb{P}_\mu^{X_n}(\{y\}) \xrightarrow{n \to \infty} \pi(\{y\})$ holds for all $y \in \mathcal{X}$ if and only if

$$2\|\mathbb{P}_\mu^{X_n} - \pi\|_{\mathrm{TV}} = \sum_{y \in \mathcal{X}} |\mathbb{P}_\mu^{X_n}(\{y\}) - \pi(\{y\})| \xrightarrow{n \to \infty} 0$$

($\|\mathbb{P}_\mu^{X_n} - \pi\|_{\mathrm{TV}} := \sup\{|\mathbb{P}_\mu^{X_n}(A) - \pi(A)| : A \subseteq \mathcal{X}\}$ is called total variation distance). Both statements hold if and only if $\mathbb{P}_\mu^{X_n}$ converges in distribution to π. □

§19.14 **Notation**. We denote by $\gcd(M)$ the greatest common divisor of all $n \in M \subseteq \mathbb{N}$. □

§19.15 **Definition**. For any $x \in \mathcal{X}$, $\mathrm{d}_x := \gcd(\{n \in \mathbb{N} : \mathrm{P}^n_{x,x} \in (0,1]\})$ is called the *period* of the state x. If $\mathrm{d}_x = \mathrm{d}_y$ for all $x, y \in \mathcal{X}$, then $\mathrm{d} = \mathrm{d}_x$ is called the *period* of $X_{\mathbb{Z}_{\geqslant 0}}$ and $X_{\mathbb{Z}_{\geqslant 0}}$ is called *aperiodic*, if $\mathrm{d} = 1$. □

§19.16 **Lemma**. *If $X_{\mathbb{Z}_{\geqslant 0}}$ is irreducible, then $\mathrm{d} := \mathrm{d}_x = \mathrm{d}_y$ for all $x, y \in \mathcal{X}$. If $X_{\mathbb{Z}_{\geqslant 0}}$ is in addition aperiodic, i.e., $\mathrm{d} = 1$, then for all $x, y \in \mathcal{X}$, there exists $\mathrm{n}_{x,y} \in \mathbb{N}$ such that $\mathrm{P}^n_{x,y} \in (0,1]$ for all $n \in \mathbb{N}_{\geqslant \mathrm{n}_{x,y}}$.*

§19.17 **Proof** of Lemma §19.16. Klenke (2020, Lemma 18.3, p.433) □

We apply a *coupling argument*. To this end, a bivariate Markov chain $((Y_z, X_z))_{z \in \mathbb{Z}_{\geqslant 0}}$ with values in a countable state space $\mathcal{X}^2 = \mathcal{X} \times \mathcal{X}$ and with family of probability measures $(\mathbb{P}_{(y,x)})_{(y,x) \in \mathcal{X}^2}$ on $(\Omega, \mathscr{A})$ is called a *coupling* if $Y_{\mathbb{Z}_{\geqslant 0}} = (Y_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ and $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ are Markov chains, each with transition matrix $\mathrm{P}_{\bullet,\bullet} = (\mathrm{P}_{x,y})_{x,y \in \mathcal{X}}$ on $\mathcal{X}$.

§19.18 **Lemma**. *Let $((Y_z, X_z))_{z \in \mathbb{Z}_{\geqslant 0}}$ be a coupling with family of probability measures $(\mathbb{P}_{(y,x)})_{(y,x) \in \mathcal{X}^2}$ on $(\Omega, \mathscr{A})$ where each Markov chain, $Y_{\mathbb{Z}_{\geqslant 0}} = (Y_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ and $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z \in \mathbb{Z}_{\geqslant 0}}$, is irreducible, positive recurrent and aperiodic with transition matrix $\mathrm{P}_{\bullet,\bullet} = (\mathrm{P}_{x,\tilde{x}})_{x,\tilde{x} \in \mathcal{X}}$ on a countable state space $\mathcal{X}$ and with invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$. If $X_{\mathbb{Z}_{\geqslant 0}}$ and $Y_{\mathbb{Z}_{\geqslant 0}}$ are independent, then the bivariate Markov chain $((Y_z, X_z))_{z \in \mathbb{Z}_{\geqslant 0}}$ is irreducible, positive recurrent and aperiodic with transition matrix*

$$\mathrm{P}_{(\bullet,\bullet),(\bullet,\bullet)} = (\mathrm{P}_{(y_1,x_1),(y_2,x_2)} := \mathrm{P}_{y_1,y_2}\mathrm{P}_{x_1,x_2})_{(y_1,x_1),(y_2,x_2) \in \mathcal{X}^2}$$

and with product probability measure $\pi \otimes \pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}^2})$ as invariant distribution. Moreover, for all $y \in \mathcal{X}$ the hitting time $\tau = \inf\{n \in \mathbb{Z}_{\geqslant 0} : X_n = Y_n\}$ is $\mathbb{P}_{\delta_y \otimes \pi}$-a.e. finite.

§19.19 **Proof** of Lemma §19.18. is given in the lecture. □

§19.20 **Theorem**. *Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ be an irreducible, positive recurrent and aperiodic Markov chain with countable state space $\mathcal{X}$, family of probability measures $(\mathbb{P}_x)_{x \in \mathcal{X}}$ on $(\Omega, \mathscr{A})$, and invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$. For all $x, \tilde{x} \in \mathcal{X}$ we have*

$$\lim_{n \to \infty} \mathbb{P}_x(X_n = \tilde{x}) = \pi(\{\tilde{x}\}).$$

§19.21 **Proof** of Theorem §19.20. is given in the lecture. □

End LE21
2026/06/29

§19.22 **Theorem**. *Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ be an irreducible and positive recurrent Markov chain with countable state space $\mathcal{X}$, family of probability measures $(\mathbb{P}_x)_{x \in \mathcal{X}}$ on $(\Omega, \mathscr{A})$, and invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$. Then the following four statements are equivalent:*

(C1) *$X_{\mathbb{Z}_{\geqslant 0}}$ is aperiodic;*

(C2) $\|\mathbb{P}^{X_n}_x - \pi\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0$ *for all $x \in \mathcal{X}$;*

(C3) *There exists $x \in \mathcal{X}$ with* $\|\mathbb{P}^{X_n}_x - \pi\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0$;

(C4) $\|\mathbb{P}^{X_n}_\mu - \pi\|_{\mathrm{TV}} \xrightarrow{n \to \infty} 0$ *for all $\mu \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$.*

§19.23 **Proof** of Theorem §19.22. is given in the lecture. □

Chapter 6

Ergodic theory

§20 Stationary and ergodic processes

Ergodic theory is the study of laws of large numbers for possibly dependent, but stationary, random variables.

§20.01 **Reminder.** Let $\mathcal{T} \subseteq \mathbb{R}$ be closed under addition (e.g. $\mathcal{T} = \mathbb{Z}_{\geqslant 0}$) and for $s \in \mathcal{T}$ denote by

$$\vartheta_s : \mathcal{X}^{\mathcal{T}} \ni x_{\mathcal{T}} = (x_t)_{t \in \mathcal{T}} \mapsto \vartheta_s(x_{\mathcal{T}}) := x_{\mathcal{T}+s} := (x_t)_{t \in \mathcal{T}+s} = (x_{t+s})_{t \in \mathcal{T}} \in \mathcal{X}^{\mathcal{T}}$$

the *shift* (-operator). A stochastic process $X_{\mathcal{T}} = (X_t)_{t \in \mathcal{T}}$ with state space $\mathcal{X}$ is called *stationary*, if $X_{\mathcal{T}}$ and $X_{\mathcal{T}+s} := \vartheta_s(X_{\mathcal{T}})$ are identically distributed (i.e. $\mathbb{P}^{X_{\mathcal{T}}} = \mathbb{P}^{X_{\mathcal{T}+s}}$), symbolically $X_{\mathcal{T}} \overset{d}{=} X_{\mathcal{T}+s}$, for every $s \in \mathcal{T}$ (see Definition §09.08).

In the sequel we write $\vartheta^s := \vartheta_s$ and $\vartheta := \vartheta_1$. For $z \in \mathcal{T} = \mathbb{Z}$ we note that $\vartheta^z \circ \vartheta^{-z} = \vartheta^0 = \mathrm{id}_{\mathcal{X}^{\mathbb{Z}}}$ and $\vartheta^{|z|} = \vartheta \circ \cdots \circ \vartheta$ ($|z|$-times). Moreover, $X_{\mathbb{Z}} \overset{d}{=} \vartheta^z(X_{\mathbb{Z}})$ for all $z \in \mathbb{Z}$ is equivalent to $X_{\mathbb{Z}} \overset{d}{=} \vartheta(X_{\mathbb{Z}})$. □

§20.02 **Example.**

(a) If $\{X_t : t \in \mathcal{T}\}$ are i.i.d. random variables, then $X_{\mathcal{T}} = (X_t)_{t \in \mathcal{T}}$ is stationary. Dismissing the independence assumption, i.e., $\mathbb{P}^{X_t} = \mathbb{P}^{X_s}$ holds for every $t, s \in \mathcal{T}$, in general $X_{\mathcal{T}}$ is not stationary. For example, consider i.d. random variables $\{X_t : t \in \mathbb{Z}_{\geqslant 0}\}$ where $X_1 = X_n$ for all $n \in \mathbb{N}$ but $X_0 \neq X_1$. Then $X_{\mathbb{Z}_{\geqslant 0}}$ is not stationary.

(b) Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ be an irreducible and positive recurrent Markov chain with countable state space, family of probability measures $(\mathbb{P}_x)_{x \in \mathcal{X}}$ on $(\Omega, \mathscr{A})$, and invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$. If π is the initial distribution, i.e. $\mathbb{P}_\pi^{X_{\mathbb{Z}_{\geqslant 0}}} = \sum_{x \in \mathcal{X}} \pi(\{x\}) \mathbb{P}_x^{X_{\mathbb{Z}_{\geqslant 0}}}$ is the distribution of $X_{\mathbb{Z}_{\geqslant 0}}$, then $X_{\mathbb{Z}_{\geqslant 0}} \overset{d}{=} \vartheta(X_{\mathbb{Z}_{\geqslant 0}})$ and hence $X_{\mathbb{Z}_{\geqslant 0}}$ is stationary.

(c) Let $\{X_z : z \in \mathbb{Z}\}$ be i.i.d. real random variables, $m \in \mathbb{N}$, and $\mathfrak{c}_i \in \mathbb{R}$ for all $[\![m]\!]$. Then $Y_z := \sum_{i \in [\![m]\!]} \mathfrak{c}_i X_{z-i}$, $z \in \mathbb{Z}$, defines a stationary process $Y_{\mathbb{Z}} = (Y_z)_{z \in \mathbb{Z}}$, that is called the moving average of $X_{\mathbb{Z}} = (X_z)_{z \in \mathbb{Z}}$ with weights $(\mathfrak{c}_i)_{i \in [\![m]\!]} \in \mathbb{R}^m$.

More general, let $g \in \mathcal{M}(\mathscr{B}^{\mathbb{Z}})$ be real and Borel-measurable (e.g. $g(x_{\mathbb{Z}}) = \sum_{i \in [\![m]\!]} \mathfrak{c}_i x_{0-i}$ for $x_{\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$). For $z \in \mathbb{Z}$ we set

$$g_z := g \circ \vartheta^z : \mathbb{R}^{\mathbb{Z}} \ni x_{\mathbb{Z}} \mapsto g_z(x_{\mathbb{Z}}) := g(\vartheta^z(x_{\mathbb{Z}})) \in \mathbb{R}$$

(e.g. $g_z(x_{\mathbb{Z}}) = \sum_{i \in [\![m]\!]} \mathfrak{c}_i x_{z-i}$ for $x_{\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$). Then $Y_{\mathbb{Z}} = (Y_z := g_z(X_{\mathbb{Z}}))_{z \in \mathbb{Z}}$ defines a stationary process . □

In the sequel $(\Omega, \mathscr{A}, \mathbb{P})$ denotes a probability space and $T : \Omega \to \Omega$ a measurable map, i.e. $T \in \mathcal{M}(\mathscr{A}, \mathscr{A})$.

§20.03 **Definition.** $T \in \mathcal{M}(\mathscr{A}, \mathscr{A})$ is called *measure preserving* (maßerhaltend) or $\mathbb{P}$*-preserving* if

$$\mathbb{P}^T(A) = \mathbb{P}(T^{-1}(A)) = \mathbb{P}(A) \quad \text{for all } A \in \mathscr{A}.$$

In this case $(\Omega, \mathscr{A}, \mathbb{P}, T)$ is called a (measure preserving) *dynamical system*. □

§20.04 **Example**. Let $(\mathcal{X}, \mathscr{B}_{\mathcal{X}})$ be a Polish space equipped with its Borel-σ-algebra.

(a) For a $\mathcal{X}$-valued random variable $f \in \mathcal{M}(\mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}})$ and a measure preserving map T on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ the process $Y_{\mathbb{Z}_{\geqslant 0}} := (Y_z := f \circ T^z)_{z \in \mathbb{Z}_{\geqslant 0}}$ is stationary.

(b) Let $\Pi_{\mathbb{Z}_{\geqslant 0}} = (\Pi_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ be the coordinate process on $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathbb{P}) = (\Omega, \mathscr{A}, \mathbb{P})$ (see Remark §09.03) and ϑ the shift operator as recalled in Reminder §20.01. Then

(b1) $\Pi_z = \Pi_0 \circ \vartheta^z$ for all $z \in \mathbb{Z}_{\geqslant 0}$ and

(b2) $\Pi_{\mathbb{Z}_{\geqslant 0}}$ is a stationary process if and only if $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathbb{P}, \vartheta)$ is a dynamical system.

(c) Let $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathbb{P}, \vartheta)$ be a dynamical system and $f \in \mathcal{M}(\mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}})$ a $\mathcal{X}$-valued random variable on $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathbb{P})$, then $X_{\mathbb{Z}_{\geqslant 0}} = (X_z := f \circ \vartheta^z)_{z \in \mathbb{Z}_{\geqslant 0}}$ is stationary and ϑ is $\mathbb{P}^{X_{\mathbb{Z}_{\geqslant 0}}}$-preserving, i.e. $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathbb{P}^{X_{\mathbb{Z}_{\geqslant 0}}}, \vartheta)$ a dynamical system. □

§20.05 **Example** *(§20.02 (b) continued)*. Let $X_{\mathbb{Z}_{\geqslant 0}}$ be an irreducible and positive recurrent Markov chain with countable state space $(\mathcal{X}, 2^{\mathcal{X}})$, family of probability measures $(\mathbb{P}_x)_{x \in \mathcal{X}}$ on $(\Omega, \mathscr{A})$, and invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$. If π is the initial distribution, then ϑ is $\mathbb{P}_\pi^{X_{\mathbb{Z}_{\geqslant 0}}}$-preserving, i.e. $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}}, \mathbb{P}_\pi^{X_{\mathbb{Z}_{\geqslant 0}}}, \vartheta)$ a dynamical system. □

§20.06 **Definition**. An event $A \in \mathscr{A}$ is called

invariant, if $T^{-1}(A) = A$ ($\Leftrightarrow T^{-1}(A) \triangle A = \emptyset$),

$\mathbb{P}$*(-almost)-invariant*, if $\mathbb{P}(T^{-1}(A) \triangle A) = 0$ ($\Leftrightarrow \mathbb{1}_{T^{-1}(A)} = \mathbb{1}_A$ $\mathbb{P}$-a.e.)

We denote by

$\mathscr{I}_T := \{A \in \mathscr{A} : T^{-1}(A) = A\}$ the σ-algebra of all invariant events,

$\mathscr{I}_T^{\mathbb{P}} := \{A \in \mathscr{A} : \mathbb{P}(T^{-1}(A) \triangle A) = 0\}$ the σ-algebra of all $\mathbb{P}$-invariant events. □

§20.07 **Remark**. $\mathscr{I}_T$ and $\mathscr{I}_T^{\mathbb{P}}$ are indeed σ-algebras (verify!) and evidently $\mathscr{I}_T \subseteq \mathscr{I}_T^{\mathbb{P}}$.

For each $A^\circ \in \mathscr{I}_T$ and $A \in \mathscr{A}$ with $\mathbb{P}(A^\circ \triangle A) = 0$ we have $A \in \mathscr{I}_T^{\mathbb{P}}$. Conversely, for each $A \in \mathscr{I}_T^{\mathbb{P}}$ there exists $A^\circ \in \mathscr{I}_T$ with $\mathbb{P}(A^\circ \triangle A) = 0$. □

§20.08 **Reminder**. $\mathscr{T} = \{\emptyset, \Omega\}$ and $\mathscr{T}^{\mathbb{P}} := \{A \in \mathscr{A} : \mathbb{P}(A) \in \{0,1\}\}$ denotes, respectively, the trivial and $\mathbb{P}$-trivial σ-algebra. A sub-σ-algebra $\mathscr{I} \subseteq \mathscr{A}$ is called $\mathbb{P}$-trivial if $\mathbb{P}(A) \in \{0,1\}$ for all $A \in \mathscr{I}$, i.e. $\mathscr{I} \subseteq \mathscr{T}^{\mathbb{P}}$. □

§20.09 **Definition**. A dynamical system $(\Omega, \mathscr{A}, \mathbb{P}, T)$ is called

ergodic, if $\mathscr{I}_T$ is $\mathbb{P}$-trivial or in equal $\mathscr{I}_T^{\mathbb{P}}$ is $\mathbb{P}$-trivial.

In this case we call T $\mathbb{P}$*-ergodic*. □

§20.10 **Lemma**. *Let $(\Omega, \mathscr{A}, \mathbb{P}, T)$ be a dynamical system.*

(i) *For any real $f \in \mathcal{M}(\mathscr{A})$ we have*

(ia) $f \circ T = f$ *if and only if f is $\mathscr{I}_T$-measurable, i.e. $f \in \mathcal{M}(\mathscr{I}_T)$*

(ib) $f \circ T = f$ $\mathbb{P}$-a.e. *if and only if f is $\mathscr{I}_T^{\mathbb{P}}$-measurable, i.e. $f \in \mathcal{M}(\mathscr{I}_T^{\mathbb{P}})$*

(ii) $(\Omega, \mathscr{A}, \mathbb{P}, T)$ *is ergodic if and only if every $f \in \mathcal{M}(\mathscr{I}_T^{\mathbb{P}})$ is* $\mathbb{P}$-a.s. *constant.*

§20.11 **Proof** of Lemma §20.10. is given in the lecture. □

§20.12 **Example** *(§20.04 continued)*. Let $\Pi_{\mathbb{Z}_{\geqslant 0}} = (\Pi_z)_{z \in \mathbb{Z}_{\geqslant 0}}$ be the coordinate process on $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathbb{P})$, $f \in \mathcal{M}(\mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}})$ and $X_{\mathbb{Z}_{\geqslant 0}} = (X_z := f \circ \vartheta^z)_{z \in \mathbb{Z}_{\geqslant 0}}$.

(a) If ϑ is $\mathbb{P}$-ergodic, then ϑ is also $\mathbb{P}^{X_{\mathbb{Z}_{\geqslant 0}}}$-ergodic, i.e. $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, \mathscr{B}_{\mathcal{X}}^{\mathbb{Z}_{\geqslant 0}}, \mathbb{P}^{X_{\mathbb{Z}_{\geqslant 0}}}, \vartheta)$ an ergodic dynamical system, since $X_{\mathbb{Z}_{\geqslant 0}}^{-1}(\mathscr{I}_\vartheta) \subseteq \mathscr{I}_\vartheta$.

(b) If $\{\Pi_z : z \in \mathbb{Z}_{\geqslant 0}\}$ are i.i.d. random variables, then ϑ is $\mathbb{P}$-ergodic by Kolmogorov's 0-1 law (lecture course EWS, *Satz* §16.15), since $\mathscr{I}_\vartheta \subseteq \bigwedge_{n\in\mathbb{Z}_{\geqslant 0}} \bigvee_{m\in\mathbb{Z}_{\geqslant n}} \sigma(\Pi_m) =: \mathscr{A}_{\Pi_{\mathbb{Z}_{\geqslant 0}}}$ (tail-σ-algebra). □

§20.13 **Example** *(§20.05 continued)*. Let $X_{\mathbb{Z}_{\geqslant 0}}$ be an irreducible and positive recurrent Markov chain with countable state space $(\mathcal{X}, 2^{\mathcal{X}})$, family of probability measures $(\mathbb{P}_x)_{x\in\mathcal{X}}$ on $(\Omega, \mathscr{A})$, and invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$. If π is the initial distribution, then ϑ is $\mathbb{P}_\pi^{X_{\mathbb{Z}_{\geqslant 0}}}$-ergodic, i.e. $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}}, \mathbb{P}_\pi^{X_{\mathbb{Z}_{\geqslant 0}}}, \vartheta)$ an ergodic dynamical system. □

End LE22 2026/07/01

§21 Ergodic theorems

In this section, $(\Omega, \mathscr{A}, \mathbb{P}, T)$ always denotes a (measure preserving) dynamical system. Further let $X_0 \in \mathcal{M}(\mathscr{A})$ be real, Borel-measurable and

$$X_m = X_0 \circ T^m \quad \text{for all } m \in \mathbb{N}.$$

Hence, $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ is a stationary, real-valued stochastic process. We consider the partial sum prozess $S_{\mathbb{Z}_{\geqslant 0}} := (S_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ with $S_0 := 0$ and

$$S_n := \sum_{m\in[\![n]\!]} X_{m-1} = \sum_{m\in[\![0,n-1]\!]} X_m = \sum_{m\in[\![0,n-1]\!]} X_0 \circ T^m \quad \text{for all } n \in \mathbb{N},$$

Ergodic theorems are laws of large numbers for $S_{\mathbb{Z}_{\geqslant 0}} = (S_z)_{z\in\mathbb{Z}_{\geqslant 0}}$.

§21.01 **Lemma** *(Hopf's maximal-ergodic lemma)*. *Let $X_0 \in \mathcal{L}^1(\mathbb{P})$. Define $M_z := \max\{S_m : m \in [\![0, z]\!]\}$ for $z \in \mathbb{Z}_{\geqslant 0}$, and set $M_\infty := \sup\{S_z : z \in \mathbb{Z}_{\geqslant 0}\}$. Then $\mathbb{E}(X_0 \mathbb{1}_{\mathbb{R}_{>0}}(M_n)) \in \mathbb{R}_{\geqslant 0}$ for every $n \in \mathbb{N}$ and by dominated convergence $\mathbb{E}(X_0 \mathbb{1}_{\mathbb{R}_{>0}}(M_\infty)) \in \mathbb{R}_{\geqslant 0}$.*

§21.02 **Proof** of Lemma §21.01. is given in the lecture. □

§21.03 **Birkhoff's ergodic Theorem**. *Let $X_0 \in \mathcal{L}^1(\mathbb{P})$. Then*

(i) $\frac{1}{n} S_n = \frac{1}{n} \sum_{m\in[\![n]\!]} X_{m-1} = \frac{1}{n} \sum_{m\in[\![0,n-1]\!]} X_0 \circ T^m \xrightarrow{\mathbb{P}\text{-a.e.}} \mathbb{E}(X_0 | \mathscr{I}_T)$.

(ii) *if in addition T is $\mathbb{P}$-ergodic, then* $\frac{1}{n} \sum_{m\in[\![n]\!]} X_{m-1} \xrightarrow{\mathbb{P}\text{-a.e.}} \mathbb{E}(X_0)$.

§21.04 **Proof** of Theorem §21.03. is given in the lecture. □

§21.05 **Lemma**. *Let $\{Y_z : z \in \mathbb{Z}_{\geqslant 0}\}$ be identically distributed, real random variables with $Y_0 \in \mathcal{L}^s(\mathbb{P})$ for $s \in \mathbb{R}_{\geqslant 1}$. Define*

$$Z_n := \Big|\frac{1}{n} \sum_{i\in[\![n]\!]} Y_{i-1}\Big|^s \quad \textit{for all } n \in \mathbb{N}.$$

Then $(Z_n)_{n\in\mathbb{N}}$ is uniformly $\mathbb{P}$-integrable.

§21.06 **Proof** of Lemma §21.05. is given in the lecture. □

§21.07 **von Neumann's ergodic Theorem**. *Let $s \in \mathbb{R}_{\geqslant 1}$ and $X_0 \in \mathcal{L}^s(\mathbb{P})$. Then*

(i) $\frac{1}{n} S_n = \frac{1}{n} \sum_{m\in[\![n]\!]} X_{m-1} = \frac{1}{n} \sum_{m\in[\![0,n-1]\!]} X_0 \circ T^m \xrightarrow{\mathcal{L}^s(\mathbb{P})} \mathbb{E}(X_0 | \mathscr{I}_T)$.

(ii) *if in addition T is $\mathbb{P}$-ergodic, then* $\frac{1}{n}\sum_{m\in[\![n]\!]} X_{m-1} \xrightarrow{\mathcal{L}^s(\mathbb{P})} \mathbb{E}(X_0)$.

§21.08 **Proof** of Theorem §21.07. is given in the lecture. □

§21.09 **Example** *(§20.13 continued)*. Let $X_{\mathbb{Z}_{\geqslant 0}} = (X_z)_{z\in\mathbb{Z}_{\geqslant 0}}$ be an irreducible and positive recurrent Markov chain with countable state space, family of probability measures $(\mathbb{P}_x)_{x\in\mathcal{X}}$ on $(\Omega,\mathscr{A})$, and invariant distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$. Since $(\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}, 2^{\mathcal{X}^{\mathbb{Z}_{\geqslant 0}}}, \mathbb{P}_\pi^{X_{\mathbb{Z}_{\geqslant 0}}}, \vartheta)$ is an ergodic dynamical system by applying *Birkhoff's ergodic* Theorem §21.03 for all $x \in \mathcal{X}$ we have

$$\frac{1}{n}\sum_{m\in[\![0,n-1]\!]} \mathbb{1}_{\{X_m = x\}} \to \mathbb{P}_\pi(X_0 = x) = \pi(\{x\}) \quad \mathbb{P}_\pi\text{-a.e.} \tag{21.01}$$

In this sense $\pi(\{x\})$ is the average time $X_{\mathbb{Z}_{\geqslant 0}}$ spends in the state x in the long run. Note that due to *von Neumann's ergodic* Theorem §21.07 the convergence in (21.01) holds also in $\mathcal{L}^s(\mathbb{P}_\pi)$ for all $s \in \mathbb{R}_{\geqslant 1}$. Similarly, for each $f \in \mathcal{L}^s(\pi)$ we have

$$\frac{1}{n}\sum_{m\in[\![0,n-1]\!]} f(X_m) \to \mathbb{P}_\pi(f(X_0)) = \pi(f) \quad \mathbb{P}_\pi\text{-a.e. and in } \mathcal{L}^s(\pi). \tag{21.02}$$

The last results are the foundations for *Markov Chain Monte Carlo* (MCMC) methods which eventually aim to generate a sample from a target distribution $\pi \in \mathfrak{M}_{\mathrm{pr}}(2^{\mathcal{X}})$ or to calculate a value $\pi(f)$ for some $f \in \mathcal{L}^s(\pi)$. □

§21.10 **Example**. Let $(\Omega,\mathscr{A},\mathbb{P}_0,T)$ and $(\Omega,\mathscr{A},\mathbb{P}_1,T)$ be ergodic. If $\mathbb{P}_0 \neq \mathbb{P}_1$ then there exists $f \in \mathcal{M}(\mathscr{A})$ with $|f| \leqslant 1$ and $\mathbb{P}_0(f) \neq \mathbb{P}_1(f)$. By *Birkhoff's ergodic* Theorem §21.03 we have

$$\frac{1}{n}\sum_{m\in[\![0,n-1]\!]} f\circ T^m \to \begin{cases} \mathbb{P}_0(f) & \mathbb{P}_0\text{-a.e.} \\ \mathbb{P}_1(f) & \mathbb{P}_1\text{-a.e.} \end{cases}$$

Set $A := \{\sum_{m\in[\![0,n-1]\!]} f\circ T^m \to \mathbb{P}_0(f)\}$ then $\mathbb{P}_0(A) = 1$ and $\mathbb{P}_1(A) = 0$. Therefore, $\mathbb{P}_0 \perp \mathbb{P}_1$ are mutually singular (see Definition §03.01). Consequently, if T is $\mathbb{P}_0$-ergodic and $\mathbb{P}_1$-ergodic then either $\mathbb{P}_0 = \mathbb{P}_1$ or $\mathbb{P}_0 \perp \mathbb{P}_1$. □

End LE23 2026/07/03

Chapter 7

Exponential inequalities

§22 Exponential inequalities

In this section, all random variables are defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and we denote by $\mathbb{E}$ the expectation with respect to $\mathbb{P}$.

§22.01 **Lemma** *(Birgé, 2000). Let X be a real random variable.*

(i) *If there exists $\mathrm{a} \in \mathbb{R}_{>0}$ such that*

$$\log(\mathbb{E}(\exp(\mathrm{t}X))) \leqslant (\mathrm{at})^2 \quad \textit{for all } \mathrm{t} \in \mathbb{R}_{\geqslant 0},$$

then for all $x \in \mathbb{R}_{\geqslant 0}$ we have

$$\mathbb{P}(X \geqslant 2\mathrm{a}\sqrt{x}) \leqslant \exp(-x) \quad \textit{and} \quad \mathbb{P}(X \geqslant x) \leqslant \exp(-\tfrac{x^2}{4\mathrm{a}^2}).$$

(ii) *If there exist $\mathrm{a}, \mathrm{b} \in \mathbb{R}_{>0}$ such that*

$$\log(\mathbb{E}(\exp(\mathrm{t}X))) \leqslant \tfrac{(\mathrm{at})^2}{1-\mathrm{bt}} \quad \textit{for all } \mathrm{t} \in (0, \mathrm{b}^{-1}),$$

then for all $x \in \mathbb{R}_{\geqslant 0}$ we have

$$\mathbb{P}(X \geqslant 2\mathrm{a}\sqrt{x} + \mathrm{b}x) \leqslant \exp(-x), \quad \mathbb{P}(X \geqslant x) \leqslant \exp(-\min(\tfrac{x^2}{8\mathrm{a}^2}, \tfrac{x}{2\mathrm{b}}))$$
$$\textit{and} \quad \mathbb{P}(X \geqslant x) \leqslant \exp(-\tfrac{x^2}{4\mathrm{a}^2 + 2\mathrm{b}x}).$$

§22.02 **Proof** of Lemma §22.01. is given in the lecture. □

§22.03 **Remark**. A standard-normally distributed random variable $X \sim \mathrm{N}_{(0,1)}$ satisfies

$$\mathbb{E}(\exp(\mathrm{t}X)) = \tfrac{1}{\sqrt{2\pi}} \int \exp(\mathrm{t}x - \tfrac{x^2}{2})\mathrm{d}x = \exp(\tfrac{\mathrm{t}^2}{2})\tfrac{1}{\sqrt{2\pi}} \int \exp(-\tfrac{(x-\mathrm{t})^2}{2})\mathrm{d}x = \exp(\tfrac{\mathrm{t}^2}{2}).$$

Consequently, by applying Lemma §22.01 (i) we obtain $\mathrm{N}_{(0,1)}(\mathbb{R}_{\geqslant x}) = \mathbb{P}(X \geqslant x) \leqslant \exp(-\frac{x^2}{2})$. On the other hand side, for $x \in \mathbb{R}_{>0}$ using

$$\int_x^\infty (1 + y^{-2})\exp(-\tfrac{y^2}{2})\mathrm{d}y = x^{-1}\exp(-\tfrac{x^2}{2}) \quad \text{and} \quad \int_x^\infty y\exp(-\tfrac{y^2}{2})\mathrm{d}y = \exp(-\tfrac{x^2}{2})$$

we obtain

$$\tfrac{x^2}{1+x^2} \leqslant \sqrt{2\pi}x\exp(\tfrac{x^2}{2})\mathrm{N}_{(0,1)}(\mathbb{R}_{\geqslant x}) \leqslant 1$$

and hence $\mathbb{P}(X \geqslant x) = \frac{\exp(-\frac{x^2}{2})}{\sqrt{2\pi}x}(1 + \mathrm{o}(1))$ as $x \to \infty$ where $\frac{\exp(-\frac{x^2}{2})}{x} = \exp(-\frac{x^2}{2}(1 + \frac{\log(x^2)}{x^2}))$. □

§22.04 **Corollary**. *Let $(X_i)_{i \in [\![n]\!]}$ be independent real random variables.*

(i) *If there exists* $\mathrm{a}_{[\![n]\!]} = (\mathrm{a}_i)_{i\in[\![n]\!]} \in \mathbb{R}^n_{>0}$ *such that for all* $i \in [\![n]\!]$

$$\log(\mathbb{E}(\exp(\mathrm{t}X_i))) \leqslant (\mathrm{a}_i\mathrm{t})^2 \quad \textit{for all } \mathrm{t} \in \mathbb{R}_{\geqslant 0}$$

and setting $\|\mathrm{a}_{[\![n]\!]}\|_2^2 := \sum_{i\in[\![n]\!]} \mathrm{a}_i^2$ *then for all* $x \in \mathbb{R}_{\geqslant 0}$ *we have*

$$\mathbb{P}\Big(\sum_{i\in[\![n]\!]} X_i \geqslant 2\|\mathrm{a}_{[\![n]\!]}\|_2\sqrt{x}\Big) \leqslant \exp(-x) \quad \textit{and} \quad \mathbb{P}\Big(\sum_{i\in[\![n]\!]} X_i \geqslant x\Big) \leqslant \exp\Big(-\frac{x^2}{4\|\mathrm{a}_{[\![n]\!]}\|_2^2}\Big).$$

(ii) *If there exist* $\mathrm{a}_{[\![n]\!]}, \mathrm{b}_{[\![n]\!]} \in \mathbb{R}^n_{>0}$ *such that for all* $i \in [\![n]\!]$

$$\log(\mathbb{E}(\exp(\mathrm{t}X_i))) \leqslant \frac{(\mathrm{a}_i\mathrm{t})^2}{1-\mathrm{b}_i\mathrm{t}} \quad \textit{for all } \mathrm{t} \in (0, \mathrm{b}_i^{-1}),$$

and setting $\|\mathrm{b}_{[\![n]\!]}\|_\infty := \max\{\mathrm{b}_i : i \in [\![n]\!]\}$ *then for all* $x \in \mathbb{R}_{\geqslant 0}$ *we have*

$$\mathbb{P}\Big(\sum_{i\in[\![n]\!]} X_i \geqslant 2\|\mathrm{a}_{[\![n]\!]}\|_2\sqrt{x} + \|\mathrm{b}_{[\![n]\!]}\|_\infty x\Big) \leqslant \exp(-x), \quad \mathbb{P}\Big(\sum_{i\in[\![n]\!]} X_i \geqslant x\Big) \leqslant \exp\Big(-\min\Big(\frac{x^2}{8\|\mathrm{a}_{[\![n]\!]}\|_2^2}, \frac{x}{2\|\mathrm{b}_{[\![n]\!]}\|_\infty}\Big)\Big)$$
$$\textit{and} \quad \mathbb{P}\Big(\sum_{i\in[\![n]\!]} X_i \geqslant x\Big) \leqslant \exp\Big(-\frac{x^2}{4\|\mathrm{a}_{[\![n]\!]}\|_2^2 + 2\|\mathrm{b}_{[\![n]\!]}\|_\infty x}\Big).$$

§22.05 **Proof** of Corollary §22.04. is given in the lecture. □

§22.06 **Lemma** *(Hoeffding, 1963). Let* X *be a real random variable. If there exist* $\mathrm{a}, \mathrm{b} \in \mathbb{R}$ *such that*

$$\mathbb{E}(X) = 0 \quad \textit{and} \quad \mathrm{a} \leqslant X \leqslant \mathrm{b},$$

then we have

$$\log(\mathbb{E}(\exp(\mathrm{t}X))) \leqslant \mathrm{t}^2\frac{(\mathrm{b}-\mathrm{a})^2}{8} \quad \textit{for all } \mathrm{t} \in \mathbb{R}_{\geqslant 0}.$$

§22.07 **Proof** of Lemma §22.06. is given in the lecture. □

§22.08 **Theorem** *(Hoeffding's inequaltiy). Let* $(X_i)_{i\in[\![n]\!]}$ *be independent real random variables. If there exist* $\mathrm{a}_{[\![n]\!]}, \mathrm{b}_{[\![n]\!]} \in \mathbb{R}^n$ *such that for all* $i \in [\![n]\!]$

$$\mathbb{E}(X_i) = 0 \quad \textit{and} \quad \mathrm{a}_i \leqslant X_i \leqslant \mathrm{b}_i,$$

then for all $x \in \mathbb{R}_{\geqslant 0}$ *we have*

$$\mathbb{P}\Big(\sum_{i\in[\![n]\!]} X_i \geqslant x\Big) \leqslant \exp\Big(-\frac{2x^2}{\|\mathrm{b}_{[\![n]\!]} - \mathrm{a}_{[\![n]\!]}\|_2^2}\Big).$$

§22.09 **Proof** of Theorem §22.08. is given in the lecture. □

§22.10 **Corollary**. *Let* $(R_i)_{i\in[\![n]\!]} \sim \mathrm{U}^n_{\{-1,1\}}$ *be independent and uniformely on* $\{-1,1\}$*-distributed random variables (Rademacher) and* $\mathrm{c}_{[\![n]\!]} = (\mathrm{c}_i)_{i\in[\![n]\!]} \in \mathbb{R}^n$*, then for all* $x \in \mathbb{R}_{\geqslant 0}$ *we have*

$$\mathbb{P}\Big(\Big|\sum_{i\in[\![n]\!]} \mathrm{c}_i R_i\Big| \geqslant x\Big) \leqslant 2\exp\Big(-\frac{x^2}{2\|\mathrm{c}_{[\![n]\!]}\|_2^2}\Big).$$

Note that $\|\mathrm{c}_{[\![n]\!]}\|_2^2$ *equals the variance of* $\sum_{i\in[\![n]\!]} \mathrm{c}_i R_i$.

§22.11 **Proof** of Corollary §22.10. is given in the lecture. □

§22.12 **Lemma** *(Cramér condition).* *Let X be a real random variable. If there exist $\mathrm{c}, \sigma^2 \in \mathbb{R}_{>0}$ such that*

$$\mathbb{E}(X) = 0 \quad \textit{and} \quad \mathbb{E}(|X|^m) \leqslant \tfrac{m!}{2}\sigma^2 \mathrm{c}^{m-2} \quad \textit{for all } m \in \mathbb{N}_{\geqslant 2}$$

then we have

$$\log(\mathbb{E}(\exp(\mathrm{t}X))) \leqslant \tfrac{\sigma^2 \mathrm{t}^2}{2(1-\mathrm{ct})} \quad \textit{for all } \mathrm{t} \in (0, \mathrm{c}^{-1}).$$

§22.13 **Proof** of Lemma §22.12. is given in the lecture. □

§22.14 **Theorem** *(Bernstein's inequality). Let $(X_i)_{i \in [\![n]\!]}$ be independent real random variables. If there exist $\mathrm{c}_{[\![n]\!]}, \sigma^2_{[\![n]\!]} \in \mathbb{R}^n_{>0}$ such that for all $i \in [\![n]\!]$*

$$\mathbb{E}(X_i) = 0 \quad \textit{and} \quad \mathbb{E}(|X_i|^m) \leqslant \tfrac{m!}{2}\sigma_i^2 \mathrm{c}_i^{m-2} \quad \textit{for all } m \in \mathbb{N}_{\geqslant 2},$$

then for all $x \in \mathbb{R}_{\geqslant 0}$ we have

$$\mathbb{P}\Big(\sum_{i \in [\![n]\!]} X_i \geqslant x\Big) \leqslant \exp\Big(-\tfrac{x^2}{2\|\sigma_{[\![n]\!]}\|_2^2 + 2\|\mathrm{c}_{[\![n]\!]}\|_\infty x}\Big).$$

§22.15 **Proof** of Theorem §22.14. is given in the lecture. □

§22.16 **Remark**. For $x \in \mathbb{R}_{>-1}$ the supremum of the function

$$\mathbb{R} \ni \mathrm{t} \mapsto 1 + \mathrm{t}(1 + x) - \exp(\mathrm{t}) \in \mathbb{R}$$

is achieved for $\mathrm{t}_\circ = \log(1 + x)$ and we set

$$\phi(x) := (1 + x)\log(1 + x) - x = \sup\{\mathrm{t}x - \exp(\mathrm{t}) + 1 + \mathrm{t} \colon \mathrm{t} \in \mathbb{R}_{>0}\} \tag{22.01}$$

for $x \in \mathbb{R}_{\geqslant 0}$. □

§22.17 **Lemma**. *Let X be a real random variable. If there exists $\mathrm{a}, \mathrm{c} \in \mathbb{R}_{>0}$ such that*

$$\log(\mathbb{E}(\exp(\mathrm{t}X))) \leqslant \mathrm{a}(\exp(\mathrm{tc}) - 1 - \mathrm{tc}) \quad \textit{for all } \mathrm{t} \in \mathbb{R}_{>0},$$

then for all $x \in \mathbb{R}_{\geqslant 0}$ we have

$$\mathbb{P}(X \geqslant x) \leqslant \exp(-\mathrm{a}\phi(\tfrac{x}{\mathrm{ac}})) \leqslant \exp(-\tfrac{3x}{4\mathrm{c}}\log(1 + \tfrac{2x}{3\mathrm{ac}})) \leqslant \exp(-\tfrac{x^2}{2\mathrm{ac}^2 + 2\mathrm{c}x/3}) \quad \textit{and}$$
$$\mathbb{P}(X \geqslant \sqrt{2\mathrm{ac}^2 x} + \mathrm{c}x/3) \leqslant \exp(-x).$$

§22.18 **Proof** of Lemma §22.17. is given in the lecture. □

§22.19 **Lemma**. *Let X be a real random variable satisfying*

$$\mathbb{E}(X) = 0 \quad \textit{and} \quad |X| \leqslant \mathrm{c} \quad \textit{for } \mathrm{c} \in \mathbb{R}_{>0}.$$

Then setting $\sigma^2 := \mathbb{E}(X^2)$ we have

$$\log(\mathbb{E}(\exp(\mathrm{t}X))) \leqslant \tfrac{\sigma^2}{\mathrm{c}^2}(\exp(\mathrm{tc}) - 1 - \mathrm{tc}) \quad \textit{for all } \mathrm{t} \in \mathbb{R}_{\geqslant 0}.$$

§22.20 **Proof** of Lemma §22.19. is given in the lecture. □

§22.21 **Corollary**. *Let $(X_i)_{i\in[\![n]\!]}$ be independent real random variables satisfying*

$$\mathbb{E}(X_i) = 0 \quad \text{and} \quad |X_i| \leqslant \mathrm{c} \quad \text{for all } i \in [\![n]\!] \text{ and } \mathrm{c} \in \mathbb{R}_{>0}.$$

Then setting $\sigma_n^2 := \frac{1}{n}\sum_{i\in[\![n]\!]}\mathbb{E}(X_i^2)$ and $S_n := \sum_{i\in[\![n]\!]} X_i$ we have

$$\log(\mathbb{E}(\exp(\mathrm{t}S_n))) \leqslant \frac{n\sigma_n^2}{\mathrm{c}^2}(\exp(\mathrm{tc}) - 1 - \mathrm{tc}) \quad \text{for all } \mathrm{t} \in \mathbb{R}_{\geqslant 0}$$

and the same inequality holds for $-S_n$.

§22.22 **Proof** of Corollary §22.21. is given in the lecture. □

§22.23 **Theorem** *(Inequalities of Bennet, Prohorov and Bernstein)*. *Let $(X_i)_{i\in[\![n]\!]}$ be independent real random variables satisfying*

$$\mathbb{E}(X_i) = 0 \quad \text{and} \quad |X_i| \leqslant \mathrm{c} \quad \text{for all } i \in [\![n]\!] \text{ and } \mathrm{c} \in \mathbb{R}_{>0}.$$

Then setting $\sigma_n^2 := \frac{1}{n}\sum_{i\in[\![n]\!]}\mathbb{E}(X_i^2)$ and $S_n := \sum_{i\in[\![n]\!]} X_i$ for all $x \in \mathbb{R}_{\geqslant 0}$ we have

$$\begin{aligned}\mathbb{P}(S_n \geqslant x) &\leqslant \exp\big(-\tfrac{n\sigma_n^2}{\mathrm{c}^2}\phi\big(\tfrac{x\mathrm{c}}{n\sigma_n^2}\big)\big)\\ &\leqslant \exp\big(-\tfrac{3x}{4\mathrm{c}}\log\big(1+\tfrac{2x\mathrm{c}}{3n\sigma_n^2}\big)\big)\\ &\leqslant \exp\big(-\tfrac{x^2}{2n\sigma_n^2+2x\mathrm{c}/3}\big) \quad \text{and}\end{aligned}$$

$$\mathbb{P}(S_n \geqslant \sqrt{2n\sigma^2 x} + \mathrm{c}x/3) \leqslant \exp(-x),$$

where ϕ is defined in (22.01) and the same inequalities hold for $\mathbb{P}(S_n \leqslant -x)$.

§22.24 **Proof** of Theorem §22.23. is given in the lecture. □

End LE24
2026/07/08

Bibliography

Bauer, H. (1992). *Maß- und Integrationstheorie*. 2., überarbeitete Auflage: Berlin etc.: Walter de Gruyter.

Elstrodt, J. (2011). *Maß- und Integrationstheorie.* 7., überarbeitete und ergänzte Auflage: Berlin: Springer.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer, Berlin, Heidelberg.

Klenke, A. (2020). *Wahrscheinlichkeitstheorie*. 4., überarbeitete und ergänzte Auflage: Springer Spektrum.

Neveu, J. (1975). *Discrete-parameter martingales.* English (Translated by T. P. Speed.) Vol. 10. North-Holland mathematical library. Elsevier.

Witting, H. (1985). *Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang.* German. Stuttgart: B. G. Teubner.