L^2 as a Hilbert space and conditional expectations

Robert Baumgarth[1]

[1]Fakultät für Mathematik und Informatik, Augustusplatz 10, 04109 Leipzig, Germany

robert.baumgarth@math.uni-leipzig.de

23rd November 2021

Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $\mathscr{F} \subset \mathscr{A}$ a sub-σ-algebra.

For any random variable $X : \Omega \to \mathbb{R}$ and measurable function $u : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}u(X) = \int_\Omega u(X)\mathrm{d}\mathbb{P} = \int_\mathbb{R} u(x)\mathbb{P}(X \in \mathrm{d}x),$$

where $\mathbb{P}(X \in \mathrm{d}x) = \mathbb{P} \circ X^{-1}(\mathrm{d}x) = \mathbb{P}_X(\mathrm{d}x)$ is the distribution of X.

If $X \sim f(x)\mathrm{d}x$ is a density for X, then $\mathbb{P}(X \in \mathrm{d}x) = f(x)\mathrm{d}x$ and

$$\mathbb{E}u(X) = \int_\mathbb{R} u(x)f(x)\mathrm{d}x.$$

If $X \sim \sum_{n\in\mathbb{N}_0} p_n\delta_n$ is discrete, then p_n is e.g. the Bernoulli, Binomial, Poisson distribution etc. and $\delta_\omega : \mathscr{A} \to \{0, 1\}$ is given by $\delta_\omega(A) = \mathbb{1}_A(\omega)$ $(A \in \mathscr{A})$ the Dirac measure. It follows

$$\int f(y)\delta_x(\mathrm{d}y) = f(x) \quad\Longrightarrow\quad \mathbb{E}u(X) = \sum_{n\in\mathbb{N}} u(n)p_n = \sum_{n\in\mathbb{N}} u(n)\mathbb{P}(X = n).$$

The space of (equivalence classes of) square-integrable random variables is denoted by

$$\begin{aligned} \mathscr{L}^2(\Omega, \mathscr{A}, \mathbb{P}) &:= \left\{ X : \Omega \to \mathbb{R} \quad \text{ZV}, \quad \int_\Omega X^2\mathrm{d}\mathbb{P} < \infty \right\}, \\ \mathsf{L}^2(\Omega, \mathscr{A}, \mathbb{P}) &:= \mathscr{L}^2(\Omega, \mathscr{A}, \mathbb{P})/\sim, \end{aligned}$$

where

$$X \sim Y \qquad\Longleftrightarrow\qquad X = Y \quad \mathbb{P}\text{-f.s.}$$

Recall: L^2 is actually the space of equivalence classes, with which we can calculate as we do with functions. In what follows, we will always identify equivalence classes and their representatives. If we would like to emphasise Ω, $\mathscr{A}$ or $\mathbb{P}$, we shortly write

$$\mathsf{L}^2(\Omega), \quad \mathsf{L}^2(\mathbb{P}), \quad \mathsf{L}^2(\mathscr{A}) \qquad \text{etc. instead of} \qquad \mathsf{L}^2(\Omega, \mathscr{A}, \mathbb{P}).$$

1 Orthogonal projection

For a comprehensive introduction for this subsection, we refer the reader to [Sch17a] ∬ 20-22 and ∬ 17 treating the more general case of a σ-finite measure space and [Wil91]. First, let us recall a central theorem in the study of the geometry of Hilbert spaces. A **Hilbert space** $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}})$ is a complete inner product space, i.e. an inner product space where every Cauchy sequence converges. Let $\|\cdot\| = \langle\cdot,\cdot\rangle_{\mathscr{H}}^{1/2}$ the norm corresponding to the inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}}$ and $\mathbb{K} := \{\mathbb{R}, \mathbb{C}\}$.

Theorem 1.1 (Projection theorem). *Let $C \neq \emptyset$ be a closed convex subset of the Hilbert space $\mathscr{H}$. For every $h \in \mathscr{H}$ there is a unique minimiser $u \in C$ such that*

$$\|h - u\| = \inf_{w \in C} \|h - w\| .$$

This element $u = P_C h$ is called **(orthogonal) projection** *of h onto C.*

A **continuous linear functional** on $\mathscr{H}$ is a map $\Lambda : \mathscr{H} \to \mathbb{K}$, $h \mapsto \Lambda(h)$ which is linear,

$$\Lambda(\alpha g + \beta h) = \alpha\Lambda(g) + \beta\Lambda(h) \qquad \text{for all } \alpha, \beta \in \mathscr{K}\ \forall g, h \in \mathscr{H}$$

and satisfies

$$|\Lambda(g - h)| \leqslant c(\Lambda)\, \|g - h\| \qquad \text{for all } g, h \in \mathscr{H}$$

with a constant $c(\Lambda) \geqslant 0$ independent of $g, h \in \mathscr{H}$. In fact, all linear functional on $\mathscr{H}$ arise in this way:

Theorem 1.2 (Riesz representation theorem). *For each continuous linear functional φ on the Hilbert space $\mathscr{H}$ there exists a unique $g \in \mathscr{H}$ such that*

$$\Lambda_g(h) := \langle h, g\rangle , \qquad \text{for all } h \in \mathscr{H},$$

and $\left\|\Lambda_g\right\|_{\mathscr{H}} = \|g\|_{\mathscr{H}}$. Conversely, given $h \in \mathscr{H}$, then $h \mapsto \langle h, g\rangle$ is a continuous linear functional with operator norm $\|g\|_{\mathscr{H}}$.

2 Conditional expectation

A prototypical example of a Hilbert space is $\mathscr{H} = \mathsf{L}^2(\mathscr{A})$, i.e. the space of all functions whose (absolute) 2^{nd} moment is integrable with inner product, resp. norm

$$\langle u, v\rangle_2 := \int uv \, \mathrm{d}\mathbb{P} \qquad \text{resp.} \qquad \|u\|_2 := \left(\int |u|^2 \, \mathrm{d}\mathbb{P}\right)^{1/2} .$$

Given a sub-σ-algebra $\mathscr{F} \subset \mathscr{A}$, the idea of a conditional expectation is to make a random variable $u \in \mathsf{L}^2(\mathscr{A})$ also measurable with respect to the coarser σ-algebra, i.e. $u \in \mathscr{F}$. Formally, construct an object $\mathbb{E}^{\mathscr{F}} \in \mathscr{F}$ that is $\mathscr{F}$-measurable by definition.

Definition 2.1. Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $\mathscr{F} \subset \mathscr{A}$ be a sub-σ-algebra. The **conditional expectation** (▷ bedingte Erwartung, espérance conditionnelle) of $u \in \mathsf{L}^2(\mathscr{A})$ relative to $\mathscr{F}$ is the orthogonal projection onto the closed subspace $\mathsf{L}^2(\mathscr{F})$

$$\mathbb{E}^{\mathscr{F}} : \mathsf{L}^2(\mathscr{A}) \to \mathsf{L}^2(\mathscr{F}), \quad u \mapsto \mathbb{E}^{\mathscr{F}} u.$$

Remark 2.2 (Properties of $\mathbb{E}^{\mathscr{F}}$). Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space and $\mathscr{F} \subset \mathscr{A}$ be a sub-σ-algebra. The conditional expectation $\mathbb{E}^{\mathscr{F}}$ has the following properties, for all $u, v \in \mathsf{L}^2(\mathscr{A})$, almost surely:

(i) $\mathbb{E}^{\mathscr{F}} \in \mathsf{L}^2(\mathscr{F})$

(ii) $\left\|\mathbb{E}^{\mathscr{F}} u\right\|_{\mathsf{L}^2(\mathscr{F})} \leqslant \|u\|_{\mathsf{L}^2(\mathscr{A})}$ (Contraction)

(iii) $\left\langle \mathbb{E}^{\mathscr{F}} u, w\right\rangle = \left\langle u, \mathbb{E}^{\mathscr{F}} w\right\rangle = \left\langle \mathbb{E}^{\mathscr{F}} u, \mathbb{E}^{\mathscr{F}} w\right\rangle$ (Symmetry)

(iv) $\mathbb{E}^{\mathscr{F}} u$ is the unique minimiser in $\mathsf{L}^2(\mathscr{F})$ such that

$$\left\|u - \mathbb{E}^{\mathscr{F}} u\right\|_{\mathsf{L}^2(\mathscr{A})} = \min_{g \in \mathsf{L}^2(\mathscr{F})} \|u - g\|_{\mathsf{L}^2(\mathscr{A})}$$

(v) $u = w \implies \mathbb{E}^{\mathscr{F}} u = \mathbb{E}^{\mathscr{F}} w$

(vi) $\mathbb{E}^{\mathscr{F}}(\alpha u + \beta w) = \alpha \mathbb{E}^{\mathscr{F}} u + \beta \mathbb{E}^{\mathscr{F}} w$ for all $\alpha, \beta \in \mathbb{R}$ (Linearity)

(vii) If $\mathscr{G} \subset \mathscr{F}$ is a another sub-σ-algebra, then $\mathbb{E}^{\mathscr{G}} \mathbb{E}^{\mathscr{F}} u = \mathbb{E}^{\mathscr{G}} u$ (Tower property)

(viii) $\mathbb{E}^{\mathscr{F}}(gu) = g \mathbb{E}^{\mathscr{F}} u$ for all $g \in \mathsf{L}^\infty(\mathscr{F})$ (Pull out)

(ix) $\mathbb{E}^{\mathscr{F}} g = g$ for all $g \in \mathsf{L}^2(\mathscr{F})$

(x) $0 \leqslant u \leqslant 1 \implies 0 \leqslant \mathbb{E}^{\mathscr{F}} u \leqslant 1$ (Markov property)

(xi) $u \leqslant w \implies \mathbb{E}^{\mathscr{F}} u \leqslant \mathbb{E}^{\mathscr{F}} w$ (Monotony)

(xii) $\left|\mathbb{E}^{\mathscr{F}} u\right| \leqslant \mathbb{E}^{\mathscr{F}} |u|$ (Δ-inequality)

(xiii) $\mathbb{E}^{\{\emptyset, \Omega\}} u = \mathbb{E} u$

(xiv) $\mathbb{E}\mathbb{E}^{\mathscr{F}} u = \mathbb{E} u$ (Tower property)

(xv) $0 \leqslant u_n \uparrow u \implies \mathbb{E}^{\mathscr{F}} u_n \uparrow \mathbb{E}^{\mathscr{F}} u$ (conditional Beppo Levi)

(xvi) $u_n \geqslant 0 \implies \mathbb{E}^{\mathscr{F}} \left(\liminf u_n\right) \leqslant \liminf \mathbb{E}^{\mathscr{F}} u_n$ (conditional Fatou)

(xvii) For all $n \in \mathbb{N}$, $\left|u_n\right| \leqslant w$, $\mathbb{E} w < \infty$ and

$$u_n \xrightarrow{\text{a.s.}} u \implies \mathbb{E}^{\mathscr{F}} u_n \to \mathbb{E}^{\mathscr{F}} u \qquad \text{(conditional dominated convergence)}$$

(xviii) $c : \mathbb{R} \to \mathbb{R}$ convex and $\mathbb{E}\,|c(u)| < \infty \implies \mathbb{E}^{\mathscr{F}} c(u) \geqslant c\left(\mathbb{E}^{\mathscr{F}}\right)$ (conditional Jensen)

Theorem 2.3. *Let $\mathscr{F} \subset \mathscr{A}$ be sub-σ-algebra. Then*

$$\left\|\mathbb{E}^{\mathscr{F}} u\right\|_{\mathsf{L}^1} \leqslant \|u\|_{\mathsf{L}^1} \qquad \text{for all } u \in \mathsf{L}^1(\mathscr{F}),$$

and $\mathbb{E}^{\mathscr{F}}$ can be extended to $\mathsf{L}^1(\mathscr{F})$ by continuity and Properties (v)-(xiv) carry over to $u \in \mathsf{L}^1(\mathscr{F})$. It is common to write $\mathbb{E}(u \mid \mathscr{F})$ instead of $\mathbb{E}^{\mathscr{F}} u$.

Theorem 2.4 (Classical definition of $\mathbb{E}^{\mathscr{F}}$). *Let $\mathscr{F} \subset \mathscr{A}$ be a sub-σ-algebra. For $X \in \mathsf{L}^1(\mathscr{A})$ and $Y \in \mathsf{L}^1(\mathscr{A})$ it is then equivalent:*

(i) $Y = \mathbb{E}(X \mid \mathscr{F})$ *a.s.* – Y **is a version of the conditional expectation**

(ii) $\displaystyle\int_F Y \mathrm{d}\mathbb{P} = \int_F X \mathrm{d}\mathbb{P}$ for all $F \in \mathscr{F}$,

where (ii) holds on any $\cap$-stable generator $\mathscr{F} = \sigma(\mathscr{G})$ for $\mathscr{F}$.

Theorem 2.5 (Independence). *Let $\mathscr{H}, \mathscr{F} \subset \mathscr{A}$ be two sub-σ-algebras and $X : \Omega \to \mathbb{R}$ a random variable. Then*

$$\mathscr{H} \perp\!\!\!\perp \sigma(\sigma(X), \mathscr{F}) \qquad \Longrightarrow \qquad \mathbb{E}(X \mid \sigma(\mathscr{F}, \mathscr{H})) = \mathbb{E}(X \mid \mathscr{F}).$$

In particular, $$X \perp\!\!\!\perp \mathscr{H} \qquad \Longrightarrow \qquad \mathbb{E}(X \mid \mathscr{H}) = \mathbb{E}X.$$

The proof of the following lemma follows from usual approximations arguments for Lebesgue integrals, cf. e.g. [SP14, Appendix A.2].

Lemma 2.6. *Assume that $\mathscr{X}, \mathscr{Y} \subset \mathscr{A}$ are σ-algebras and let $X : (\Omega, \mathscr{A}) \to (C, \mathscr{C}) \in \mathscr{X}/\mathscr{C}$ and $Y : (\Omega, \mathscr{A}) \to (D, \mathscr{D}) \in \mathscr{Y}/\mathscr{D}$ be two random variables such that $\mathscr{X} \perp\!\!\!\perp \mathscr{Y}$. Then*

$$\mathbb{E}(\Phi(X, Y) \mid \mathscr{X}) = \mathbb{E}\Phi(x, Y)\big|_{x=X} = \mathbb{E}(\Phi(X, Y) \mid X)$$

holds for all bounded $\mathscr{C} \times \mathscr{D}/\mathscr{B}(\mathbb{R})$-measurable functions $\Phi : C \times D \to \mathbb{R}$. If $\Psi : E \times \Omega \to \mathbb{R}$ is bounded and $\mathscr{C} \times \mathscr{Y}/\mathscr{B}(\mathbb{R})$-measurable, then

$$\mathbb{E}(\Psi(X(\cdot), \cdot) \mid \mathscr{X}) = \mathbb{E}\Psi(x, \cdot)\big|_{x=X} = \mathbb{E}(\Psi(X(\cdot), \cdot) \mid X).$$

Corollary 2.7. *Assume that $\mathscr{X}, \mathscr{Y} \subset \mathscr{A}$ are σ-algebras and let $X : (\Omega, \mathscr{A}) \to (C, \mathscr{C}) \in \mathscr{X}/\mathscr{C}$ and $Y : (\Omega, \mathscr{A}) \to (D, \mathscr{D}) \in \mathscr{Y}/\mathscr{D}$ be two random variables such that $\mathscr{X} \perp\!\!\!\perp \mathscr{Y}$. Then*

$$\mathbb{E}\Phi(X, Y) = \int \mathbb{E}\Phi(x, Y)\mathbb{P}(X \in \mathrm{d}x) = \mathbb{E}\int \Phi(x, Y)\mathbb{P}(X \in \mathrm{d}x)$$

holds for all bounded $\mathscr{C} \times \mathscr{D}/\mathscr{B}(\mathbb{R})$-measurable functions $\Phi : C \times D \to \mathbb{R}$. If $\Psi : E \times \Omega \to \mathbb{R}$ is bounded and $\mathscr{C} \times \mathscr{Y}/\mathscr{B}(\mathbb{R})$-measurable, then

$$\mathbb{E}\Psi(X(\cdot), \cdot) = \int \mathbb{E}\Psi(x, \cdot)\mathbb{P}(X \in \mathrm{d}x) = \mathbb{E}\int \Psi(x, \cdot)\mathbb{P}(X \in \mathrm{d}x).$$

3 (Classical) conditional probability

For the following sections, see e.g. [Sch17b, § 14].

Definition 3.1. Let $\mathscr{F} \subset \mathscr{A}$ eine sub-σ-algebra. Then

$$\mathbb{P}(A \mid \mathscr{F}) := \mathbb{E}(\mathbb{1}_A \mid \mathscr{F}) \qquad \text{for all } A \in \mathscr{A}$$

is the **conditional probability (with respect to $\mathscr{F}$).** (▷ **bedingte Wahrscheinlichkeit, probabilité conditionnelle**)

Remark 3.2. (a) Actually, it should read $\mathbb{P}(A \mid \mathscr{F})(\omega)$, i.e. $\mathbb{P}(A \mid \mathscr{F})$ is a random variable!

(b) $A \mapsto \mathbb{P}(A \mid \mathscr{F})$ is, in general, not a measure! The problem lies in the σ-additivity. As an element in L^2, $\mathbb{P}(A \mid \mathscr{F})$ is only (uniquely) determined up to $\mathbb{P}$-null sets $N \subset \Omega$ and may depend on A.

Mind: Two partitions $A = \bigcup A_j = \bigcup_j B_j$ may give different null sets. Since there may be *uncountably many* of such, the null sets may explode.

(c) Eminent exceptions of (b):

(i) $A \in \mathscr{F} \Longrightarrow \mathbb{P}(A \mid \mathscr{F}) = \mathbb{E}(\mathbb{1}_A \mid \mathscr{F}) = \mathbb{1}_A = \delta_{\{\omega\}}(A)$ and $A \mapsto \mathbb{1}_A(\omega)$ is a measure.

(ii) $A \perp\!\!\!\perp \mathscr{F} \Longrightarrow \mathbb{P}(A \mid \mathscr{F}) = \mathbb{E}(\mathbb{1}_A \mid \mathscr{F}) = \mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$ and $A \mapsto \mathbb{P}(A)$ is a measure.

Example 3.3. We recover the definition of the classical «conditional probability»

$$A \mapsto \mathbb{P}(A \mid B) := \begin{cases} \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}, & \mathbb{P}(B) > 0, \\ 0, & \mathbb{P}(B) = 0, \end{cases}$$

in the following way. First we want to find a relation between $\mathbb{P}(A \mid B)$ and $\mathbb{P}(A \mid \mathscr{F})$:
Let $\mathscr{F} := \sigma(\{B\}) = \{\emptyset, B, B^c, \Omega\}$. Then we have

$$\mathsf{L}^2(\mathscr{F}) = \left\{\alpha \mathbb{1}_B + \beta \mathbb{1}_{B^c} \;:\; \alpha, \beta \in \mathbb{R}\right\},$$

and hence for $X \in \mathsf{L}^1(\mathscr{A})$

$$\mathbb{E}^{\mathscr{F}} X = \alpha \mathbb{1}_B + \beta \mathbb{1}_{B^c} \qquad \text{(suitable } \alpha, \beta).$$

To find α and β, we calculate

$$\begin{aligned} & \mathbb{E}^{\mathscr{F}}(\mathbb{1}_B X) \overset{\text{pull}}{\underset{\text{out}}{=}} \mathbb{1}_B \mathbb{E}^{\mathscr{F}} X = \mathbb{1}_B \left(\alpha \mathbb{1}_B + \beta \mathbb{1}_{B^c}\right) = \alpha \mathbb{1}_B \\ \Longrightarrow \quad & \mathbb{E}(\mathbb{1}_B X) = \mathbb{E}\mathbb{E}^{\mathscr{F}}(\mathbb{1}_B X) = \mathbb{E}(\alpha \mathbb{1}_B) = \alpha \mathbb{P}(B) \\ \Longrightarrow \quad & \alpha = \frac{1}{\mathbb{P}(B)} \mathbb{E}(\mathbb{1}_B X) = \frac{1}{\mathbb{P}(B)} \int_B X \mathrm{d}\mathbb{P} = \int X \underbrace{\frac{\mathrm{d}\mathbb{P}(\bullet \cap B)}{\mathbb{P}(B)}}_{= \mathrm{d}\mathbb{P}(\bullet \mid B)}. \end{aligned}$$

Analogously, we see that

$$\beta = \int X \mathrm{d}\mathbb{P}(\bullet \mid B^c).$$

Hence,

$$\mathbb{E}^{\mathscr{F}} X = \alpha \mathbb{1}_B + \beta \mathbb{1}_{B^c} = \mathbb{1}_B \int X \mathrm{d}\mathbb{P}(\bullet \mid B) + \mathbb{1}_{B^c} \int X \mathrm{d}\mathbb{P}(\bullet \mid B^c).$$

In particular, for $X = \mathbb{1}_B$, then

$$\mathbb{P}(A \mid \mathscr{F}) = \mathbb{P}(A \mid B) \mathbb{1}_B + \mathbb{P}(A \mid B^c) \mathbb{1}_{B^c},$$

where we had that $\mathscr{F} = \{\emptyset, B, B^c, \Omega\}$.
More general: For disjoint $(B_j)_{j \in \mathbb{N}} \subset \mathscr{A}$ with $\bigsqcup_j B_j = \Omega$ and for $\mathscr{H} = \sigma(B_j \;:\; j \in \mathbb{N})$:

$$\begin{aligned} & \mathbb{P}(A \mid \mathscr{H}) = \sum_{j=1}^{\infty} \mathbb{P}(A \mid B_j) \mathbb{1}_{B_j} \\ \Longleftrightarrow \quad & \forall j : \mathbb{P}(A \mid \mathscr{H}) = \mathbb{P}(A \mid B_j) \qquad \mathbb{P}\text{-a.s. on } B_j \end{aligned}$$

Example 3.4. Let $\mathscr{H} = \sigma(B_1, B_2, B_3, \ldots)$, where $\Omega = \bigsqcup_{j\in\mathbb{N}} B_j$ and $\mathbb{P}(B_j) > 0$. Then, for every random variable X with $\mathbb{E}|X| < \infty$,

$$\mathbb{E}(X \mid \mathscr{H}) = \sum_{j=1}^{\infty} \frac{\mathbb{E}(X\mathbb{1}_{B_j})}{\mathbb{P}(B_j)} \mathbb{1}_{B_j}$$

$$\iff \qquad \forall j : \mathbb{E}(X \mid \mathscr{H}) = \frac{\mathbb{E}(X\mathbb{1}_{B_j})}{\mathbb{P}(B_j)} \qquad \mathbb{P}\text{-a.s. on } B_j.$$

Then, for all k, we get

$$\begin{aligned}\int_{B_k} \sum_{j=1}^{\infty} \frac{\mathbb{E}(X\mathbb{1}_{B_j})}{\mathbb{P}(B_j)} \mathbb{1}_{B_j} \mathrm{d}\mathbb{P} &= \sum_{j=1}^{\infty} \frac{\mathbb{E}(X\mathbb{1}_{B_j})}{\mathbb{P}(B_j)} \underbrace{\int_{B_k} \mathbb{1}_{B_j} \mathrm{d}\mathbb{P}}_{= \mathbb{P}(B_k \cap B_j)} \\ &= \frac{\mathbb{E}(X\mathbb{1}_{B_k})}{\mathbb{P}(B_k)} \mathbb{P}(B_k) \\ &= \mathbb{E}(X\mathbb{1}_{B_k}) = \int_{B_k} X \mathrm{d}\mathbb{P}.\end{aligned}$$

The same argument goes through for $B_1 \cup \ldots \cup B_k$ and, hence, we can use Theorem 2.4 (ii). This justifies the well-known classical notation of the conditional expectation

$$\mathbb{E}(X \mid B) := \begin{cases} \frac{\mathbb{E}(X\mathbb{1}_B)}{\mathbb{P}(B)}, & \mathbb{P}(B) > 0, \\ 0, & \mathbb{P}(B) = 0. \end{cases}$$

4 Conditional expectation if $\mathscr{F} = \sigma(Y)$

Let $\mathscr{F}$ now be generated by a single random variable $Y : \Omega \to \mathbb{R}^n$, i.e.

$$\mathscr{F} := \sigma(Y) \equiv Y^{-1}(\mathscr{B}(\mathbb{R}^n)).$$

Then we write for short

$$\mathbb{E}(X \mid Y) := \mathbb{E}(X \mid \sigma(Y))$$
$$\mathbb{E}(X \mid Y_1, Y_2, \ldots, Y_n, \ldots) := \mathbb{E}(X \mid \sigma(Y_j : j \in \mathbb{N}))$$

and analogously for $\mathbb{P}(A \mid Y)$ and $\mathbb{P}(A \mid Y_1, Y_2, \ldots)$.

Recall the factorisation lemma from measure theory:

$$\left.\begin{array}{lcl} Y : (\Omega, \mathscr{A}) & \xrightarrow{\text{mb}} & (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n)) \\ Z : (\Omega, \sigma(Y)) & \xrightarrow{\text{mb}} & (\mathbb{R}, \mathscr{B}(\mathbb{R})) \end{array}\right\} \implies \exists g : \mathbb{R}^n \xrightarrow{\text{mb}} \mathbb{R}, \quad Z = g(Y).$$

Definition 4.1. Let $Y : \Omega \to \mathbb{R}^n$, $X : \Omega \to \mathbb{R}$ random variables and $X \in \mathrm{L}^1(\mathscr{A})$. Then

$$\mathbb{E}(X \mid Y = y) \text{ is the } \textbf{function (!)} \; g : \mathbb{R}^n \to \mathbb{R}$$

such that it holds $\qquad \mathbb{E}(X \mid Y) = g(Y)$ $\mathbb{P}$-a.s..

Problem Well-defined?! $\mathbb{E}(X \mid Y = y)$ should be unique at least $\mathbb{P}_Y$-a.s.

Example 4.2. Let $Y : \Omega \to \{y_1, y_2, ...\}$ a discrete random variable, then

$$\mathbb{E}(X \mid Y = y) = \mathbb{E}(X \mid \{Y = y\}) = \begin{cases} \frac{\mathbb{E}(X \mathbb{1}_{\{y\}}(Y))}{\mathbb{P}(Y=y)}, & \mathbb{P}(Y = y) > 0, \\ 0, & \mathbb{P}(Y = y) = 0. \end{cases}$$

If Y is not discrete, there are (in general) problems if $\mathbb{P}(Y = y) = 0$ for «too many» y (e.g. for densities).

Theorem 4.3. *Let $Y : \Omega \to \mathbb{R}^n$, $X : \Omega \to \mathbb{R}$ random variables and $X \in \mathsf{L}^1(\mathscr{A})$. Then every measurable function $g : \mathbb{R}^n \to \mathbb{R}$ with $g(Y) = \mathbb{E}(X \mid Y)$ satisfies the condition*

$$\int_B g(y)\mathbb{P}(Y \in \mathrm{d}y) = \int_{\{Y \in B\}} X \mathrm{d}\mathbb{P} \qquad \text{for all } B \in \mathscr{B}(\mathbb{R}^n), \tag{4.1}$$

and g is $\mathbb{P}$-a.s. uniquely determined by (4.1). Conversely, for every g with (4.1), we have $g(Y) = \mathbb{E}(X \mid Y)$.

5 Conditional densities

Let $X, Y : \Omega \to \mathbb{R}$ random variables with joint density $f_{X,Y}(x, y)$, i.e.

$$\mathbb{P}(X \leqslant a, Y \leqslant b) = \int_{(-\infty,a]} \int_{(-\infty,b]} f_{X,Y}(x, y)\mathrm{d}y\mathrm{d}x, \qquad a, b \in \mathbb{R}.$$

The **marginal distributions** (▷ **Randverteilung, densités marginales**) , e.g. for X, can be calculated from $f_{X,Y}$ via

$$\mathbb{P}(X \leqslant a) = \mathbb{P}(X \leqslant a, Y < \infty) = \int_{(-\infty,a]} \int_{\mathbb{R}} f_{X,Y}(x, y)\mathrm{d}y\mathrm{d}x,$$

i.e., $X \sim f_X(x)\mathrm{d}x$ and $Y \sim f_Y(y)\mathrm{d}y$, where

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\mathrm{d}y \qquad \text{and} \qquad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\mathrm{d}x.$$

Definition 5.1. Let $(X, Y) \sim f_{X,Y}(x, y)\, \mathrm{d}x\, \mathrm{d}y$. Then we call

$$f_{X|Y}(x \mid y) := \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(y)} & f_Y(y) \neq 0, \\ 0 & \text{else.} \end{cases}$$

conditional density of X given Y. (▷ **bedingte Dichte von X gegeben Y, densité conditionnelle de X à rapport de Y**)

Conditional densities allow us to compute conditional expectations concretely via the following

Theorem 5.2. *Let $X, Y : \Omega \to \mathbb{R}$ random variables with joint densities $(X, Y) \sim f_{X,Y}(x, y)\mathrm{d}x\mathrm{d}y$. For all measurable functions $h : \mathbb{R}^2 \to \mathbb{R}$ with $\mathbb{E}\,|h(X, Y)| < \infty$, we have*

$$\mathbb{E}(h(X, Y) \mid Y = y) = \int_{\mathbb{R}} h(x, y) f_{X|Y}(x \mid y)\mathrm{d}x \qquad \text{for } \mathbb{P}_Y\text{-a.s. all } y \in \mathbb{R}.$$

If we choose $h(X,Y) = \mathbb{1}_B(X) = \mathbb{1}_{\{X\in B\}}$ in the previous theorem, and use that $\{X \in B\} \in \sigma(X)$, we get

Corollary 5.3. *Let* $(X,Y) \sim f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$ *and* $B \in \mathscr{B}(\mathbb{R})$. *Then*

$$\mathbb{P}(X \in B \mid Y)(\omega) = \int_B f_{X|Y}(x \mid Y(\omega))\mathrm{d}x.$$

6 Conditional expectation and independence

Theorem 6.1. *Let* $X, Y : \Omega \to \mathbb{R}$ *be random variables and* $h : \mathbb{R}^2 \to \mathbb{R}$ *measurable with* $\mathbb{E}\,|h(X,Y)| < \infty$. *If* $X \perp\!\!\!\perp Y$, *then*

$$\mathbb{E}(h(X,Y) \mid Y = b) = \mathbb{E}h(X,b).$$

Remark 6.2. Modifying the argument a little, it also holds that for random variables $X, Y : \Omega \to \mathbb{R}$ with $Y \perp\!\!\!\perp \mathscr{F}$ and X $\mathscr{F}$-measurable. For $h : \mathbb{R}^2 \to \mathbb{R}$ measurable with $\mathbb{E}\,|h(X,Y)| < \infty$, then $\mathbb{E}(h(X,Y) \mid Y = b) = \mathbb{E}h(X,b)$.

7 Regular conditional probability

Definition 7.1. Let $\mathscr{F} \subset \mathscr{A}$ be a sub-σ-algebra and $X : \Omega \to \mathbb{R}^n$ a random variable with values in an arbitrary measure space $(E, \mathscr{E})$. A **regular conditional probability** (▷ **regulare bedingte Verteilung, probabilité conditionnelle régulière**) is a map (Markov kernel) $P(\cdot,\cdot) : \Omega \times \mathscr{B}(\mathbb{R}^n) \to [0,1]$ such that

(a) $\omega \mapsto P(\omega, B)$ is measurable,

(b) $B \mapsto P(\omega, B)$ is $\mathbb{P}$-a.s. a probability measure on $(E, \mathscr{E})$

and $P(\omega, B)$ is a version of $\mathbb{P}(X \in B \mid \mathscr{F})$.

Then it holds the (deep!)

Theorem 7.2. *If X is a random variable with values in a Polish space $(E, \mathscr{E})$ (e.g. $\mathbb{R}^n$), $\mathscr{E} = \mathscr{B}(E)$, then there is a regular conditional probability.*

References

[Sch17a] R. L. Schilling. *Measures, integrals and martingales*. Second. Cambridge University Press, Cambridge, 2017, pp. xvii+476.

[Sch17b] R. L. Schilling. *Wahrscheinlichkeit*. De Gruyter Studium. Eine Einführung für Bachelor-Studenten. [An introduction for undergraduate students]. De Gruyter, Berlin, 2017, pp. x+232.

[SP14] R. L. Schilling and L. Partzsch. *Brownian motion. An introduction to stochastic processes*. Second Edition. De Gruyter Graduate. With a chapter on simulation by Björn Böttcher. De Gruyter, Berlin, 2014, pp. xvi+408.

[Wil91] D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991, pp. xvi+251.