# Probability Theory

Wilhelm Stannat

Technische Universität Darmstadt

Winter Term 2007/08

Fifth part - corrected version

This text is a summary of the lecture on Probability Theory held at the TU Darmstadt in Winter Term 2007/08.

Please email all misprints and mistakes to
`stannat@mathematik.tu-darmstadt.de`

# Bibliography


1. Bauer, H., *Probability theory*, de Gruyter, 1996.
2. Bauer, H., *Maß- und Integrationstheorie*, de Gruyter, 1996.
3. Billingsley, P., *Probability and Measure*, Wiley, 1995.
4. Billingsley, P., *Convergence of probability measures*, Wiley, 1999.
5. Dudley, R.M., *Real analysis and probability*, Cambridge University Press, 2002.
6. Elstrodt, J., *Maß- und Integrationstheorie*, Springer, 2005.
7. Feller, W., *An introduction to probability theory and its applications*, Vol. 1 & 2, Wiley, 1950.
8. Halmos, P.R., *Measure Theory*, Springer, 1974.
9. Klenke, A., *Wahrscheinlichkeitstheorie*, Springer, 2006.
10. Shiryaev, A.N., *Probability*, Springer, 1996.

# 4 Conditional expectations

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{A}_0 \subset \mathcal{A}$ be a sub-$\sigma$-algebra.

## 1 Definitions

**Definition 1.1.** Let $X \geqslant 0$ be a r.v. A r.v. $X_0 \geqslant 0$ is said to be (a version of) the *conditional expectation of $X$ given $\mathcal{A}_0$* if

(i) $X_0$ is $\mathcal{A}_0$-measurable.

(ii) $\mathbb{E}[Y_0 \cdot X] = \mathbb{E}[Y_0 \cdot X_0]$ for all $\mathcal{A}_0$-measurable r.v. $Y_0 \geqslant 0$.

**Proposition 1.2.** *Let $X$ and $\mathcal{A}_0$ be as above. Then*

*(i) A r.v. $X_0$ satisfying (i) and (ii) of the previous definition exists. (see Subsection 3 below).*

*(ii) Any two random variables satisfying (i) and (ii) coincide $P$-a.s.*

*Notation:*

$$X_0 =: \mathbb{E}[X \,|\, \mathcal{A}_0].$$

**Remark 1.3.** *(i) "extreme cases"*

$$\mathbb{E}\big[X \;\big|\; \{\emptyset, \Omega\}\big] = \mathbb{E}[X] \qquad \mathbb{E}[X \,|\, \mathcal{A}] = X$$

*(ii) Let $X$ be a r.v., not necessarily nonnegative. We can decompose $X = X^+ - X^-$. If*

$$\min\big(\mathbb{E}[X^+ \,|\, \mathcal{A}_0]\,,\, \mathbb{E}[X^- \,|\, \mathcal{A}_0]\big) < \infty \quad P-a.s.$$

*we define*

$$\mathbb{E}[X \,|\, \mathcal{A}_0] := \mathbb{E}[X^+ \,|\, \mathcal{A}_0] - \mathbb{E}[X^- \,|\, \mathcal{A}_0]\,.$$

*Note that*

$$\begin{aligned} X \in \mathcal{L}^1 \quad &\Leftrightarrow \quad \mathbb{E}[X \,|\, \mathcal{A}_0] \in \mathcal{L}^1 \\ &\Leftrightarrow \quad \begin{cases} \mathbb{E}[X^+] = \mathbb{E}\big[\mathbb{E}[X^+ \mid \mathcal{A}_0]\big] < \infty \\ \mathbb{E}[X^-] = \mathbb{E}\big[\mathbb{E}[X^- \mid \mathcal{A}_0]\big] < \infty, \end{cases} \end{aligned}$$

*(iii) For any* $A \in \mathcal{A}$ *let*

$$P[A \,|\, \mathcal{A}_0] := \mathbb{E}[1_A \,|\, \mathcal{A}_0]\,.$$

$P[A \,|\, \mathcal{A}_0]$ *is said to be the* conditional probability given $\mathcal{A}_0$.

*(iv)* discrete case *Let* $B_i \in \mathcal{A}$, $i \in \mathbb{N}$, *be pairwise disjoint with* $\Omega = \bigcup_{i\in\mathbb{N}} B_i$ *such that* $\mathcal{A}_0 = \sigma\{B_i \,|\, i \in \mathbb{N}\}$. *Then for any r.v.* $X \geqslant 0$*:*

$$\mathbb{E}[X \,|\, \mathcal{A}_0] = \sum_{i\in\mathbb{N}:P(B_i)>0} \underbrace{\mathbb{E}[X \,|\, B_i]}_{:=\frac{1}{P(B_i)}\cdot\mathbb{E}[X\cdot 1_{B_i}]} \cdot 1_{B_i}.$$

*where*

$$\mathbb{E}[X \,|\, B_i] = \frac{1}{P[B_i]}\int_{B_i} X\,dP$$

*denotes the* elementary *conditional expectation of* $X$ *given* $B_i$.

*(v) Let* $(\Omega', \mathcal{A}')$ *be a measurable space and* $Y : \Omega \to \Omega'$ *be* $\mathcal{A}/\mathcal{A}'$*-measurable. Let* $\mathcal{A}_0 := \sigma(Y)$ *and* $X \geqslant 0$ *be a r.v. on* $\Omega$. *The factorization lemma then implies that there exists a function* $f_X : \Omega' \to \bar{\mathbb{R}}_+$, *such that* $P - a.s.$

$$\mathbb{E}[X \,|\, Y] := \mathbb{E}\big[X \,\big|\, \sigma(Y)\big] = f_X \circ Y$$

*Notation:*

$$\mathbb{E}[X \,|\, Y = \omega'] := f_X(\omega') \qquad \omega' \in \Omega'\,.$$

$$\begin{tikzcd}
 & {(\Omega,\mathcal{A}) \quad \mathcal{A}_0 := \sigma(Y) \subset \mathcal{A}} \arrow[r, "Y"] \arrow[d, "\mathbb{E}[X|Y]"] & (\Omega',\mathcal{A}') \arrow[dl, "f_X"] \\
(\Omega,\mathcal{A}) \arrow[r, "X"] & \mathbb{R}_+ &
\end{tikzcd}$$

*In particular,* $Y^{-1}(A') \in \mathcal{A}_0 = \sigma(Y)$ *for all* $A' \in \mathcal{A}'$ *and*

$$\int_{Y^{-1}(A')} X \,\mathrm{d}P \overset{\textit{1.1(ii)}}{=} \int_{Y^{-1}(A')} f_X(Y)\,\mathrm{d}P = \int_\Omega 1_{A'}(Y)\, f_X(Y)\,\mathrm{d}P$$
$$= \int_{A'} f_X \,\mathrm{d}(P \circ Y^{-1}),$$

*Hence,* $f_X$ *is* $P \circ Y^{-1}$*-a.s. unique.*

## 2 Properties of the conditional expectation

**(a) Linearity and monotonicity**

$$\mathbb{E}[c_1 X_1 + c_2 X_2 \,|\, \mathcal{A}_0] = c_1 \cdot \mathbb{E}[X_1 \,|\, \mathcal{A}_0] + c_2 \cdot \mathbb{E}[X_2 \,|\, \mathcal{A}_0]$$

$$X \leqslant Y \quad P-a.s. \quad \Rightarrow \quad \mathbb{E}[X \,|\, \mathcal{A}_0] \leqslant \mathbb{E}[Y \,|\, \mathcal{A}_0]\,.$$

**(b) Convergence theorems** *B. Levi, monotone convergence*

$$0 \leqslant X_1 \leqslant X_2 \leqslant \ldots \quad P-a.s. \quad \Rightarrow \quad \mathbb{E}\big[\lim_{n\to\infty} X_n \,\big|\, \mathcal{A}_0\big] = \lim_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{A}_0]\,.$$

*Fatou*

$$X_n \geq 0 \quad \forall\, n \in \mathbb{N} \quad \Rightarrow \quad \mathbb{E}\big[\liminf_{n\to\infty} X_n \,\big|\, \mathcal{A}_0\big] \leqslant \liminf_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{A}_0].$$

*Lebesgue, dominated convergence* $|X_n| \leqslant Y \in \mathcal{L}^1$ for all $n \in \mathbb{N}$ and $X_n \to X$ P-a.s. Then

$$\mathbb{E}\big[\lim_{n\to\infty} X_n \,\big|\, \mathcal{A}_0\big] = \lim_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{A}_0].$$

**(c) contraction properties** *Jensen's inequality* $X \in \mathcal{L}^1$ and $u$ concave (!) function on $\mathbb{R}$. Then

$$\mathbb{E}\big[u(X) \,\big|\, \mathcal{A}_0\big] \leqslant u\big(\mathbb{E}[X \,|\, \mathcal{A}_0]\big) \quad P-a.s.$$

*contraction on* $\mathcal{L}^p$
In particular, for $p \geqslant 1$ and $X \in \mathcal{L}^p$

$$\big\|\mathbb{E}[X \,|\, \mathcal{A}_0]\big\|_p \leqslant \|X\|_p.$$

It follows that the mapping

$$X \mapsto \mathbb{E}[X \,|\, \mathcal{A}_0]$$

is continuous on $\big(L^p, \|\cdot\|_p\big)$

**(d) smoothing properties** Let $X \geqslant 0$, and $Y_0 \geqslant 0$ be $\mathcal{A}_0$-measurable. Then

$$\mathbb{E}[Y_0 \cdot X \,|\, \mathcal{A}_0] = Y_0 \cdot \mathbb{E}[X \,|\, \mathcal{A}_0] \quad P-a.s.$$

*Tower property* (in german: Projektivität)
in particular: let $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \mathcal{A}$ be $\sigma$-algebras

$$\Rightarrow \qquad \mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{A}_1] \,\big|\, \mathcal{A}_0\big] = \mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{A}_0] \,\big|\, \mathcal{A}_1\big] = \mathbb{E}[X \,|\, \mathcal{A}_0] \quad P-a.s.$$

**(e) conditional expectation and independence**

**Proposition 2.1.** *Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ be $\sigma$-algebras and $X \in \mathcal{L}^1$. Let $\sigma(\mathcal{A}_1, \mathcal{A}_2)$ (resp. $\sigma(\mathcal{A}_1, X)$) be the $\sigma$-algebra generated by $\mathcal{A}_1$ and $\mathcal{A}_2$ (resp. $\mathcal{A}_1$ and $X$). Then*

$$\sigma(\mathcal{A}_1, X) \text{ independent of } \mathcal{A}_2$$
$$\Rightarrow \quad \mathbb{E}\big[X \bigm| \sigma(\mathcal{A}_1, \mathcal{A}_2)\big] = \mathbb{E}[X \,|\, \mathcal{A}_1] \quad P\text{-a.s.}$$

*In particular*

$$X \text{ independent of } \mathcal{A}_0 \quad \Rightarrow \quad \mathbb{E}[X \,|\, \mathcal{A}_0] = \mathbb{E}[X].$$

The proof of the proposition follows from the next proposition.

**Proposition 2.2.** *Let $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$ be $\sigma$-algebras and $X \in \mathcal{L}^1$, $X \geq 0$. Then the following statements are equivalent:*

*(i)* $\mathbb{E}\big[X \bigm| \sigma(\mathcal{A}_1, \mathcal{A}_2)\big] = \mathbb{E}[X \,|\, \mathcal{A}_1]$.

*(ii)*

$$\mathbb{E}[X \cdot Y \,|\, \mathcal{A}_1] = \mathbb{E}[X \,|\, \mathcal{A}_1] \cdot \mathbb{E}[Y \,|\, \mathcal{A}_1].$$

*for all $Y \geq 0$ $\sigma(\mathcal{A}_1, \mathcal{A}_2)$-measurable.*

*(iii)*

$$\mathbb{E}[X \cdot X_2 \,|\, \mathcal{A}_1] = \mathbb{E}[X \,|\, \mathcal{A}_1] \cdot \mathbb{E}[X_2 \,|\, \mathcal{A}_1].$$

*for all $X_2 \geq 0$ $\mathcal{A}_2$-measurable.*

*Proof.* Exercise. □

**Example 2.3** (Markov chain)**.** Let $P_\mu$ be the distribution of a Markov chain $X_0, X_1, \ldots$ on $\Omega = S^{\{0,1,\ldots\}}$ with initial distribution $\mu$ and transition probabilities $p(x, \mathrm{d}y)$ on $(S, \mathcal{S})$. Let $\mathcal{A}_n := \sigma(X_0, \ldots, X_n)$, $\hat{\mathcal{A}}_n := \sigma(X_i \,|\, i \geqslant n)$ and $\vartheta^n$ be the shift by $n$, i.e.

$$\vartheta^n((x_0, x_1, x_2, \ldots)) = (x_n, x_{n+1}, \ldots)$$

so that in particular $X_k \circ \vartheta^n = X_{n+k}$. Then the Markov property, applied at time $n$, implies

$$\mathbb{E}_\mu[\psi \circ \vartheta^n \,|\, \mathcal{A}_n] = \mathbb{E}_{X_n}[\psi] \tag{4.1}$$

for all $\psi \geq 0$ $\mathcal{A}$-measurable. It follows for $X \geqslant 0$, $\hat{\mathcal{A}}_n$-measurable, that

$$\mathbb{E}_\mu[X \,|\, \mathcal{A}_n] = \mathbb{E}_\mu[X \,|\, X_n]\,.$$

According to Proposition 2.2 this is equivalent to say that for all $\mathcal{A}_n$-measurable r.v. $Y \geqslant 0$

$$\mathbb{E}_\mu[Y \cdot X \,|\, X_n] = \mathbb{E}_\mu[Y \,|\, X_n] \cdot \mathbb{E}_\mu[X \,|\, X_n],$$

i.e., given the present $\sigma(X_n)$, the "future" $\hat{\mathcal{A}}_n$ is independent from the past $\mathcal{A}_n$. This is called the *elementary Markov property*.

**(f) Best approximation in $\mathcal{L}^2$.** Let $X \in \mathcal{L}^2$ be a r.v. Then

$$\mathbb{E}\Big[\big(X - \mathbb{E}[X \,|\, \mathcal{A}_0]\big)^2\Big] \leqslant \mathbb{E}\big[(X - Y_0)^2\big]$$

for all $Y_0 \in \mathcal{L}^2$, $\mathcal{A}_0$-measurable.

# 3 Existence

**(a) Hilbert space method.** Let $L^2 := \mathcal{L}^2/\sim$, with equivalence relation $X \sim Y$ meaning that $X = Y$ $P$-a.s. For given $X \in \mathcal{L}^2$, let $\bar{X}$ denote the corresponding equivalence class, i.e. $Y \in \bar{X}$ if and only if $X = Y$ $P$-a.s. Any $Y \in \bar{X}$ is called a representative of the equivalence class $\bar{X}$. Given two equivalence classes $\bar{X}, \bar{Y} \in L^2$ define its scalar product by

$$(\bar{X}, \bar{Y}) := \mathbb{E}[X \cdot Y],$$

where $X$ (resp. $Y$) is a representative of $\bar{X}$ (resp. $\bar{Y}$). Then $\big(L^2, (\ ,\ )\big)$ is a *Hilbert space*.

Let $\mathcal{L}_0^2 := \mathcal{L}^2(\Omega, \mathcal{A}_0, P)$ ($\subset \mathcal{L}^2$) be the subspace of square-integrable r.v. that are measurable w.r.t. the smaller $\sigma$-algebra $\mathcal{A}_0$. and let $L_0^2 := \mathcal{L}_0^2/\sim$. Then $L_0^2$ is a closed subspace of $L^2$ (by Riesz-Fisher (see Proposition 1.8.14), because any $L^2$-Cauchy sequence $(X_n)$ has a subsequence converging $P$-a.s., so that $X_n$ $\mathcal{A}_0$-measurable for all $n$ implies that its $L^2$-limit $X$ is $\mathcal{A}_0$-measurable too.) According to paragraph (f) of the preceeding subsection, we have that $\mathbb{E}[X \,|\, \mathcal{A}_0]$ for $X \in \mathcal{L}^2$ is a representative of the orthogonal projection $\bar{\pi}(\bar{X})$ of $\bar{X} \in L^2$ onto $L_0^2$. Using the existence of the orthogonal projection, we can now define the conditional expectation $\mathbb{E}[X \,|\, \mathcal{A}_0]$ as follows:

**Step 1:** For $X \in \mathcal{L}^2$ define

$$X_0 := \pi(\bar{X}) \quad \big(:= \mathcal{A}_0\text{-measurable representative } \bar{\pi}(\bar{X})\big)$$

It follows for all $Y_0 \in \mathcal{L}_0^2$ that

$$\begin{aligned}\mathbb{E}[Y_0 \cdot X] &= \mathbb{E}[Y_0 \cdot X_0] + \mathbb{E}\big[Y_0 \cdot (X - X_0)\big] \\ &= \mathbb{E}[Y_0 \cdot X_0] + \underbrace{(\bar{Y}_0, \overline{X - X_0})}_{=0}.\end{aligned} \tag{4.2}$$

Hence

$$\mathbb{E}[Y_0 \cdot X] = \mathbb{E}[Y_0 \cdot X_0]$$

for all $Y_0 \geq 0$ $\mathcal{A}_0$-measurable, so that

$$X_0 = \mathbb{E}[X \,|\, \mathcal{A}_0]$$

Similar to the last subsection

$$X \leqslant Y \quad P\text{-a.s.} \quad \Rightarrow \quad \pi(\bar{X}) \leqslant \pi(\bar{Y}) \quad P\text{-a.s.}$$

**Step 2:** For general $X \geqslant 0$, not necessarily in $\mathcal{L}^2$, consider $X \wedge n \in \mathcal{L}^2$. Monotonicity implies that

$$Z_0 := \lim_{n\to\infty} \pi(\overline{X \wedge n})$$

exists $P$-a.s. ($\mathcal{A}_0$-measurable !) Monotone convergence implies that for any $Y_0 \geqslant 0$, $\mathcal{A}_0$-measurable,

$$\mathbb{E}[Y_0 \cdot Z_0] = \lim_{n\to\infty} \mathbb{E}\big[Y_0 \cdot \pi(\overline{X \wedge n})\big] = \lim_{n\to\infty} \mathbb{E}[Y_0 \cdot X \wedge n] = \mathbb{E}[Y_0 \cdot X].$$

It follows that $Z_0 = \mathbb{E}[X \,|\, \mathcal{A}_0]$, hence the existence of the conditional expectation.

**(b) Radon-Nikodym theorem** Throughout the whole paragraph let $(\Omega, \mathcal{A})$ be a measurable space.

**Definition 3.1.** Let $\mu$, $\nu$ be two finite measures on $(\Omega, \mathcal{A})$. Then $\nu$ is said to be *absolutely continuous w.r.t.* $\mu$ (notation: $\nu \ll \mu$), if

$$\mu(N) = 0\,, N \in \mathcal{A} \quad \Rightarrow \quad \nu(N) = 0\,.$$

In other words: every $\mu$-null set is a $\nu$-null set (but not necessarily conversely!).

**Example 3.2.** Let $\mu$ be a finite measure on $\Omega$ and $f \in \mathcal{L}^1$, $f \geq 0$. Define the (finite) measure

$$\nu(A) := \int_A f\,d\mu := \int 1_A f\,d\mu\,, \quad A \in \mathcal{A}\,. \tag{4.3}$$

Then $\nu \ll \mu$.

The theorem of Radon-Nikodym (see Proposition 3.4 below) tells us that conversely, if $\nu \ll \mu$ there exists an $\mathcal{A}$-measurable nonnegative function $f : \Omega \to \mathbb{R}_+$ satisfying (4.3). $f$ is $\mu$-a.s. uniquely determined and called the *density* of $\nu$ w.r.t. $\mu$ (Notation: $\frac{d\nu}{d\mu}$).

The Radon-Nikodym theorem will be used to obtain a second, independent proof for the existence of the conditional expectation. We will prove the theorem in the case of finite measures.

**Lemma 3.3.** *Let $\sigma$ and $\tau$ be finite (positive) measures on a measurable space $(\Omega, \mathcal{A})$ with $\sigma(\Omega) < \tau(\Omega)$. Then there exists a measurable set $\Omega' \in \mathcal{A}$ satisfying:*

*(i)* $\sigma(\Omega') < \tau(\Omega')$.

*(ii)* $\sigma(A) \leqslant \tau(A)$ *for all* $A \in \Omega' \cap \mathcal{A} := \{A \subset \Omega' \,|\, A \in \mathcal{A}\}$.

*Proof.* (i) Let $\delta := \tau - \sigma$ (i.e., $\delta(A) := \tau(A) - \sigma(A)$ for all $A \in \mathcal{A}$). $\delta$ is bounded on $\mathcal{A}$, since

$$-\sigma(\Omega) \leqslant \delta(A) \leqslant \tau(\Omega)\,.$$

Define inductively sequences

$$(A_n)_{n\in\mathbb{N}\cup\{0\}}\ , \quad (\Omega_n)_{n\in\mathbb{N}\cup\{0\}}$$

as follows:

Let $A_0 := \emptyset$, $\Omega_0 := \Omega \setminus A_0$ ($= \Omega$, and, given $A_0, \ldots, A_n$ and $\Omega_0, \ldots, \Omega_n$, we have that

$$\alpha_n := \inf_{A\in\Omega_n\cap\mathcal{A}} \delta(A) \leqslant 0\,(\text{ since } \delta(\emptyset) = 0).$$

If $\alpha_n = 0$, let $A_{n+1} := \emptyset$ and $\Omega_{n+1} := \Omega_n \setminus A_{n+1}$ ($= \Omega_n$).

If $\alpha_n < 0$ choose $A_{n+1} \in \Omega_n \cap \mathcal{A}$ with $\delta(A_{n+1}) \leqslant \frac{\alpha_n}{2}$ and let $\Omega_{n+1} := \Omega_n \setminus A_{n+1}$.

It follows that the $A_n$, $n \in \mathbb{N}$, are pairwise disjoint, hence

$$\sum_{n=0}^{\infty} \delta(A_n) \quad \left(= \tau\Big(\bigcup_{n\geq 0} A_n\Big) - \sigma\Big(\bigcup_{n\geq 0} A_n\Big)\right)$$

is convergent, so that

$$\lim_{n\to\infty} \delta(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \alpha_n = 0.$$

Let

$$\Omega' := \bigcap_{n\geqslant 0} \Omega_n.$$

Since $(\Omega_n)$ is decreasing, it follows that

$$\delta(\Omega') = \lim_{n\to\infty} \tau(\Omega_n) - \lim_{n\to\infty} \sigma(\Omega_n) = \lim_{n\to\infty} \delta(\Omega_n) > \delta(\Omega)$$

because

$$\delta(\Omega_{n+1}) \geq \delta(\Omega_n) - \delta(A_{n+1}) \geq \delta(\Omega_n) \geq \delta(\Omega_0) = \delta(\Omega)\,.$$

This proves (i).

(ii) Let $A \in \Omega' \cap \mathcal{A}$. Then $A \in \Omega_n \cap \mathcal{A}$ for all $n$, hence $\delta(A) \geqslant \alpha_n$ for all $n$, which implies $\delta(A) \geqslant \lim_{n\to\infty} \alpha_n = 0$. □

**Proposition 3.4** (Radon-Nikodym)**.** *Let $\mu$ and $\nu$ be finite (positive) measures on the measurable space $(\Omega, \mathcal{A})$. Then the following statements are equivalent:*

*(i) There exists $f \geq 0$ $\mathcal{A}$-measurable ($\mu$-a.s. uniquely determined) such that $\nu = f\cdot\mu$ (i.e., $\nu(A) = \int_A f\,d\mu$ for all $A \in \mathcal{A}$).*

*(ii) $\nu \ll \mu$ (i.e., $\mu(N) = 0$ for $N \in \mathcal{A}$ implies $\nu(N) = 0$).*

*Proof.* (i)⇒(ii) obvious.

(ii)⇒(i)

Let $G$ be the collection of all $\mathcal{A}$-measurable numerical functions $g \geqslant 0$ on $\Omega$ satisfying

$$g \cdot \mu \leqslant \nu,$$

i.e., $\nu(A) \geqslant \int_A g \, \mathrm{d}\mu$ for all $A \in \mathcal{A}$. Note that $g \equiv 0 \in G$. Note that $G$ is stable under taking $\sup$, because for $g, h \in G$

$$\int_A \sup(g,h) \, \mathrm{d}\mu = \int_{A\cap\{g\geqslant h\}} g \, \mathrm{d}\mu + \int_{A\cap\{g<h\}} h \, \mathrm{d}\mu$$
$$\leqslant \nu\big(A \cap \{g \geqslant h\}\big) + \nu\big(A \cap \{g < h\}\big) = \nu(A) \qquad \forall\, A \in \mathcal{A}\,.$$

Let

$$\gamma := \sup_{g\in G} \int g \, \mathrm{d}\mu \quad \big(\leqslant \nu(\Omega) < \infty\big).$$

Since $G$ is $\sup$-stable, there exists an increasing sequence $(g_n)$ of functions in $G$ such that (by montone integration)

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

Let $f := \lim\limits_{n\to\infty} g_n$. Then

$$\int_A f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_A g_n \, \mathrm{d}\mu \leqslant \nu(A) \quad \forall\, A \in \mathcal{A}\,.$$

Consequently, $f \in G$. In other words: $f$ is a maximum of

$$g \mapsto \int g \, \mathrm{d}\mu \text{ on } G\,.$$

We will show next that $f \cdot \mu = \nu$. Clearly, $f \cdot \mu \leqslant \nu$, since $f \in G$. Define

$$\tau := \nu - f \cdot \mu. \tag{4.4}$$

$\tau$ is a positive, finite measure on $\mathcal{A}$ and it remains to show that $\tau \equiv 0$.

Suppose on the contrary that $\tau(\Omega) > 0$ (so that $\mu(\Omega) > 0$). Let

$$\beta := \frac{1}{2} \cdot \frac{\tau(\Omega)}{\mu(\Omega)} > 0.$$

Then

$$\tau(\Omega) = 2\beta \cdot \mu(\Omega) > \beta \cdot \mu(\Omega).$$

Due to the previous lemma there exists a set $\Omega' \in \mathcal{A}$ satisfying

$$\tau(\Omega') > \beta \cdot \mu(\Omega') \quad \text{and} \quad \tau(A) \geqslant \beta \cdot \mu(A) \tag{4.5}$$

for all $A \in \Omega' \cap \mathcal{A}$. Define $f_0 := f + \beta \cdot 1_{\Omega'}$. Then $f_0$ is $\mathcal{A}$-measurable and for all $A \in \mathcal{A}$

$$\int_A f_0 \, \mathrm{d}\mu = \int_A f \, \mathrm{d}\mu + \beta \cdot \mu(A \cap \Omega') \leqslant \int_A f \, \mathrm{d}\mu + \tau(A) = \nu(A).$$

It follows that $f_0 \in G$. On the other hand

$$\int f_0 \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \beta \cdot \mu(\Omega') = \gamma + \beta \cdot \mu(\Omega') > \gamma\,,$$

where we used the fact that $\nu \ll \mu$ implies $\mu(\Omega') > 0$. This is a contradiction to the fact that $f$ is a maximum of

$$g \mapsto \int g \, \mathrm{d}\mu$$

on $G$. Consequently, $\tau \equiv 0$. ☐

**Remark 3.5.** *(i) Let $\mu$, $\nu$ be finite measures, $\mu \ll \nu$ and $\frac{d\mu}{d\nu}$ be the density. Then*

$$\nu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$$

*but in general not $\mu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$.*

*(ii) Let $\mu$ and $\nu$ be finite measures. $\mu$ and $\nu$ are said to be equivalent (notation: $\mu \sim \nu$) if $\mu \ll \nu$ and $\nu \ll \mu$. It is easy to see that $\mu \sim \nu$ if and only if $\nu \ll \mu$, hence $\nu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$, and in addition $\mu\left(\left\{\frac{d\nu}{d\mu} = 0\right\}\right) = 0$. In this case*

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$

*(iii) Let $\nu$, $\mu$ and $\lambda$ be finite measures, $\nu \ll \mu$ and $\mu \ll \lambda$. Then*

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda} \quad \lambda - a.s.$$

**Application to the construction of the conditional expectation**

Let $\mathcal{A}_0 \subset \mathcal{A}$ be a sub-$\sigma$-algebra, $X \geq 0$, $X \in \mathcal{L}^1(P)$

Then

$$Q(A) := \int_A X \, dP = \int 1_A X \, dP\,, \quad A \in \mathcal{A}_0$$

defines a finite measure on $(\Omega, \mathcal{A}_0)$. Clearly, $Q \ll P_{|\mathcal{A}_0}$, hence

$$\exists \quad X_0 := \frac{dQ}{dP_{|\mathcal{A}_0}} \qquad \mathcal{A}_0\text{- measurable}\,.$$

Note that for $A \in \mathcal{A}_0$, by definition of the density,

$$\int 1_A X \, dP = Q(A) = \int 1_A \frac{dQ}{dP} \, dP = \int 1_A X_0 \, dP\,. \tag{4.6}$$

Clearly, (4.6) extends to

a) simple functions $Y_0 = \sum_{k=1}^n a_k 1_{A_k}$ by linearity

b) general $Y_0 \geq 0$, $\mathcal{A}_0$-measurable, by taking pointwise limits of increasing simple functions $Y_n \uparrow Y_0$

It follows that $X_0 = \frac{dQ}{dP_{|\mathcal{A}_0}}$ is a version of $\mathbb{E}[X|\mathcal{A}_0]$. For general $X \geq 0$ consider the approximation $X \wedge n \uparrow X$ and use monotonicity.

# 4 Regular conditional probabilities

Consider the mapping

$$A \mapsto P[A \,|\, \mathcal{A}_0] \quad \big(:= \mathbb{E}[1_A \,|\, \mathcal{A}_0]\big).$$

The following properties haven been shown in Subsection 2:

- $0 \leqslant P[A \,|\, \mathcal{A}_0] \leqslant 1$ $P$-a.s.
- $P[\emptyset \,|\, \mathcal{A}_0] = 0$ and $P[\Omega \,|\, \mathcal{A}_0] = 1$ $P$-a.s.
- $A_1 \subset A_2$ implies $P[A_1 \,|\, \mathcal{A}_0] \leqslant P[A_2 \,|\, \mathcal{A}_0]$ $P$-a.s.
- $A_n$, $n \in \mathbb{N}$, pairwise disjoint
  $$\Rightarrow \quad P\Big[\bigcup_{n=1}^\infty A_n \;\Big|\; \mathcal{A}_0\Big] = \sum_{n=1}^\infty P[A_n \,|\, \mathcal{A}_0] \text{ } P\text{-a.s.}$$

Note that this does not yet imply that

$$A \mapsto P[A \,|\, \mathcal{A}_0](\omega), \quad A \in \mathcal{A}, \tag{4.7}$$

defines a probability measure on $\mathcal{A}$ for $P$-a.e. $\omega \in \Omega$.

However, this is true in the discrete case and we may ask under what assumptions this is true in the general case, i.e., under what assumptions is it possible to choose "good" versions of $P[A \,|\, \mathcal{A}_0]$, $A \in \mathcal{A}$, such that (4.7) in fact defines a probability measure on $(\Omega, \mathcal{A})$ at least for $P$-a.e. $\omega \in \Omega$.

**Definition 4.1.** A measurable space $(\Omega, \mathcal{A})$ is said to be a *Borel space*, if there exists a Borel subset $U \in \mathcal{B}(\mathbb{R})$ and a bijection $\varphi : \Omega \to U$ such that both $\varphi$ and $\varphi^{-1}$, are measurable.

**Proposition 4.2.** *Assume that $(\Omega, \mathcal{A})$ is a Borel space and let $P$ be a probability measure on $(\Omega, \mathcal{A})$. Let $\mathcal{A}_0 \subset \mathcal{A}$ be a sub $\sigma$-algebra. Then there exists a transition probability from $(\Omega, \mathcal{A}_0)$ to $(\Omega, \mathcal{A})$ such that for all $A \in \mathcal{A}$*

$$K_{\mathcal{A}_0}(\omega, A) = P[A \,|\, \mathcal{A}_0](\omega) \quad P - a.s.$$

*In other words: $K_{\mathcal{A}_0}(\cdot, A)$ is a version of the conditional probability $P[A \,|\, \mathcal{A}_0]$ for all $A \in \mathcal{A}$. The transition probability $K_{\mathcal{A}_0}$ is called a* regular conditional probability given *$\mathcal{A}_0$.*

*$K_{\mathcal{A}_0}$ is uniquely determined in the following sense: if $\tilde{K}_{\mathcal{A}_0}$ is a second transition probability from $(\Omega, \mathcal{A}_0)$ to $(\Omega, \mathcal{A})$ having these properties, it follows that there exists a $P$-null set $N \in \mathcal{A}$, such that for all $\omega \in \Omega \setminus N$ and all $A \in \mathcal{A}$*

$$K_{\mathcal{A}_0}(\omega, A) = \tilde{K}_{\mathcal{A}_0}(\omega, A).$$

For a proof see Klenke, Wahrscheinlichkeitstheorie, Satz 8.36.