Chapter V

Conditional Expectations, Martingales, Ergodicity

1 Conditional Expectations

'Access to the martingale concept is afforded by one of the truely basic ideas of probability theory, that of conditional expectation.', see Bauer (1996, p. 109).

Recall the elementary *conditional probability*

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)}, \quad A, B \in \mathfrak{A}, \quad P(B) > 0.$$

Alternately, we can say, that given B, we consider a probability measure $P(\cdot \,|\, B)$ with P–density $1/P(B) \cdot 1_B$.

Next, for a random variable $X \in \mathfrak{L}^1(\Omega, \mathfrak{A}, P)$, we define the elementary *conditional expectation*

$$\mathrm{E}(X \,|\, B) = \frac{1}{P(B)} \cdot \mathrm{E}(1_B \cdot X) = \int X \, dP(\cdot \,|\, B).$$

For $A \in \mathfrak{A}$, we have

$$\mathrm{E}(1_A \,|\, B) = P(A \,|\, B).$$

A first generalization: Let I be finite or countable, and let $(B_i)_{i \in I}$ be a partition of Ω, $B_i \in \mathfrak{A}$ with $P(B_i) > 0$. Then

$$\mathfrak{G} = \left\{ \bigcup_{j \in J} B_j : J \subset I \right\}$$

is the σ–algebra generated by the B_i. Then we define a $\mathfrak{G}$–measurable mapping by

$$\mathrm{E}(X \,|\, \mathfrak{G})(\omega) = \sum_{i \in I} \mathrm{E}(X \,|\, B_i) \cdot 1_{B_i}(\omega), \qquad \omega \in \Omega. \tag{1}$$

We have the property

$$\int_{B_j} \mathrm{E}(X \,|\, \mathfrak{G}) \, dP = \mathrm{E}(X \,|\, B_j) \cdot P(B_j) = \int_{B_j} X \, dP,$$

and thus for every $G \in \mathfrak{G}$

$$\int_G \mathrm{E}(X \mid \mathfrak{G})\, dP = \int_G X\, dP.$$

Intuitively speaking, we refined the idea of the expectation as a mean; we defined a refined, 'localized mean'. The way we localize is through a σ–algebra; the localized mean is $\mathfrak{G}$–measurable. The larger the σ–algebra, the finer our refined mean. A second, just as valid, point of view is that passing from X to $\mathrm{E}(X \mid \mathfrak{G})$ is a coarsening; again, the strictness of this coarsening is described by the σ–algebra $\mathfrak{G}$. We already met the underlying idea – it was prevalent in the proof of the Radon–Nikodym Theorem.

Example 1. Extremal cases: If $|I| = 1$, we have $\mathfrak{G} = \{\emptyset, \Omega\}$ and

$$\mathrm{E}(X \mid \mathfrak{G}) = \mathrm{E}(X).$$

On tho other hand, if Ω is countable and $\mathfrak{G} = \mathfrak{P}(\Omega)$, we have

$$\mathrm{E}(X \mid \mathfrak{G}) = X.$$

Now the real thing: Let $X \in \mathfrak{L}^1(\Omega, \mathfrak{A}, P)$ and let $\mathfrak{G} \subset \mathfrak{A}$ be a sub–σ–algebra.

Definition 1. A random variable $Z \in \mathfrak{L}^1(\Omega, \mathfrak{A}, P)$ with

(i) Z is $\mathfrak{G}$-measurable;

(ii) $\forall\, G \in \mathfrak{G} : \int_G Z\, dP = \int_G X\, dP$

is called (a version of) *the conditional expectation of X, given (or w.r.t.)* $\mathfrak{G}$. Notation: $Z = \mathrm{E}(X \mid \mathfrak{G})$.
If $X = 1_A$ with $A \in \mathfrak{A}$ we say also that Z is (a version of) *the conditional probability of A, given (or w.r.t.)* $\mathfrak{G}$ Notation: $Z = P(A \mid \mathfrak{G})$.

We stress that both quantities, conditional expectation and probability, are *random variables*. Further, it is important that unlike in the elementary case, we allow the σ–algebra to contain nontrivial sets of zero measure.

Theorem 1. In the above situation, there exists a conditional expectation; two conditional expectations coincide P–a.s.

Proof. **Existence:**
Case 1: $X \geq 0$. Then

$$Q(G) := \int_G X\, dP, \qquad G \in \mathfrak{G},$$

defines (see Theorem II.7.1) a finite measure on $(\Omega, \mathfrak{G})$. Further, $Q \ll P|_{\mathfrak{G}}$. We apply the Radon–Nikodym Theorem: There is a density, in other words a $\mathfrak{G}$-measurable mapping $Z : \Omega \to [0, \infty[$ such that

$$\forall\, G \in \mathfrak{G} : \quad Q(G) = \int_G Z\, dP.$$

This Z obviously is a conditional expectation.
Case 2: X arbitrary. Then there are conditional expectations Z^+, Z^- for X^+, X^-; $Z = Z^+ - Z^-$ is a conditional expectation for X.
Uniqueness:
As a general fact, if Z, Z' are $\mathfrak{G}$–measurable and

$$\int_G Z\, dP \leq \int_G Z'\, dP \qquad \forall G \in \mathfrak{G}\ ,$$

then $Z \leq Z'$ a.s.. (Compare the proof of Theorem II.7.3.) □

In the sequel we will write $X = Y/X \leq Y$ etc. iff we have $X = Y/X \leq Y$ etc. a.s..
The theorem was suspiciously simple to prove. Indeed, there is a pitfall: It is in general not trivial to explicitly *calculate* the conditional expectation. We will collect a bunch of helpful tools and try to develop an intuition about the conditional expectation along the way.

Remark 1. 1. Two extremal cases: If X itself is $\mathfrak{G}$–measurable, then X itself qualifies as conditional expectation; hence $\mathrm{E}(X \,|\, \mathfrak{G}) = X$. If, on the other hand, $\mathfrak{G}$ consists only of sets A with $\mathbb{P}(A) \in \{0, 1\}$, then the unconditional expectation $\mathrm{E}(X)$ qualifies as conditional expectation, hence $\mathrm{E}(X \,|\, \mathfrak{G}) = \mathrm{E}(X)$.

2. Since always $\Omega \in \mathfrak{G}$, we have

$$\mathrm{E}(X) = \int_\Omega \mathrm{E}(X \,|\, \mathfrak{G})\, dP = \mathrm{E}(\mathrm{E}(X \,|\, \mathfrak{G})).$$

In the special case of (1) ($X = 1_A$ with $A \in \mathfrak{A}$) this yields the classical formula of total probability, i.e.,

$$P(A) = \sum_{i \in I} P(A \,|\, B_i) \cdot P(B_i).$$

Lemma 1. For fixed $\mathfrak{G}$, the conditional probability

$$\mathrm{E}(\cdot \,|\, \mathfrak{G}) : L^1(\Omega, \mathfrak{A}, P) \to L^1(\Omega, \mathfrak{G}, P)$$

is positive, linear and continuous. Further, if $X_n \uparrow X$, $\mathrm{E}(X \,|\, \mathfrak{G}) \uparrow \mathrm{E}(X \,|\, \mathfrak{G})$.

Proof. **Linearity:** If Z is a conditional expectation for X and Z' a condtional expectation for Y, then $Z + Z'$ is a conditional expectation for $X + Y$.
Positiveness: If $X \geq 0$ and Z is a conditional expectation for X, then $G = \{Z < 0\} \in \mathfrak{G}$ and

$$\int_G Z = \int_G X \geq 0\ ;$$

hence $Z \geq 0$.
Continuity: Let $X = X^+ - X^-$ and Z^+, Z^- be conditional expectations of X^+, X^-. Then $Z^+, Z^- \geq 0$, and

$$\mathrm{E}\,|Z| \leq \mathrm{E}\, Z^+ + \mathrm{E}\, Z^- = \mathrm{E}\, X^+ + \mathrm{E}\, X^- = \mathrm{E}\,|X|\ .$$

Linearity entails continuity.
Monotone convergence: Denote $Z_n = \mathrm{E}(X_n \,|\, \mathfrak{G})$, $Z = \mathrm{E}(X \,|\, \mathfrak{G})$. By linearity and positiveness, we know Z_n is monotonically increasing; by continuity and monotone convergence of the unconditional expectation we know that

$$\mathrm{E}\,|Z - Z_n| \leq \mathrm{E}\,|X - X_n| = \mathrm{E}\,X - \mathrm{E}\,X_n \to 0 \,.$$

Hence $Z_n \to Z$ in $\mathfrak{L}^1$, by ?? there is a subsequence tending to Z a.s.; since Z_n is monotone, this entails that $Z_n \uparrow Z$ a.s.. □

Lemma 2 (Factor out measurable parts). Let Y $\mathfrak{G}$-measurable, $X \cdot Y \in \mathfrak{L}^1$. Then

$$\mathrm{E}(X \cdot Y \,|\, \mathfrak{G}) = Y \cdot \mathrm{E}(X \,|\, \mathfrak{G}).$$

Proof. Obviously, $Y \cdot \mathrm{E}(X \,|\, \mathfrak{G})$ ist $\mathfrak{G}$-meßbar.
Case 1: $Y = 1_C$ for $C \in \mathfrak{G}$. Then for $G \in \mathfrak{G}$

$$\int_G Y \cdot \mathrm{E}(X \,|\, \mathfrak{G})\, dP = \int_{G \cap C} \mathrm{E}(X \,|\, \mathfrak{G})\, dP = \int_{G \cap C} X\, dP = \int_G X \cdot Y\, dP.$$

Case 2: Y simple. Use linearity of conditional expectation.
Case 3: Y positive. Use monotone convergence of conditional expectation.
Case 4: Y arbitrary. Split $Y = Y^+ - Y^-$. □

It is noteworthy that our old friend, algebraic induction, works just as well with conditional expectations, as sketched in the above proof.

Lemma 3 (Coarsening/Towering Lemma). Let $\mathfrak{G}_1 \subset \mathfrak{G}_2 \subset \mathfrak{A}$ be σ-algebras. Then

$$\mathrm{E}(\mathrm{E}(X \,|\, \mathfrak{G}_1) \,|\, \mathfrak{G}_2) = \mathrm{E}(X \,|\, \mathfrak{G}_1) = \mathrm{E}(\mathrm{E}(X \,|\, \mathfrak{G}_2) \,|\, \mathfrak{G}_1).$$

Proof. The first equality is within the scope of Remark 1. For the second equality, fix $G \in \mathfrak{G}_1 \subset \mathfrak{G}_2$; then

$$\int_G \mathrm{E}(\mathrm{E}(X \,|\, \mathfrak{G}_2) \,|\, \mathfrak{G}_1)\, dP = \int_G \mathrm{E}(X \,|\, \mathfrak{G}_2)\, dP = \int_G X\, dP.$$

□

We say that X and $\mathfrak{G}$ are independent iff $(\sigma(X), \mathfrak{G})$ are independent.

Lemma 4 (Independence Lemma). If $X, \mathfrak{G}$ are independent, then

$$\mathrm{E}(X \,|\, \mathfrak{G}) = \mathrm{E}(X).$$

Proof. Let $G \in \mathfrak{G}$. Then X, 1_G independent. Hence

$$\int_G X\, dP = \mathrm{E}(X \cdot 1_G) = \mathrm{E}(X)\,\mathrm{E}(1_G) = \int_G \mathrm{E}(X)\, dP \,.$$

This shows that $\mathrm{E}(X)$ qualifies for the conditional expectation. □

Theorem 2 (Jensen's inequality). Let $J \subset \mathbb{R}$ be an interval such that $X(\omega) \in J$ for all $\omega \in \Omega$. Further, let $\varphi : J \to \mathbb{R}$ be convex such that $\varphi \circ X \in \mathfrak{L}^1$. Then $\mathrm{E}(X \mid G) \in J$ a.s., and

$$\varphi \circ \mathrm{E}(X \mid \mathfrak{G}) \leq \mathrm{E}(\varphi \circ X \mid \mathfrak{G}).$$

Proof. If $a \leq X \leq b$ a.s., $a \leq \mathrm{E}(X \mid G) \leq b$ a.s. by monotonicity of the conditional expectation. Further, we note that for a countable family $(Y_n)_{n \in \mathbb{N}}$ of integrable r.v. such that $\sup_{n \in \mathbb{N}} Y_n$ is integrable, we have

$$\mathrm{E}(\sup_{n \in \mathbb{N}} Y_n \mid \mathfrak{G}) \geq \sup_{n \in \mathbb{N}} \mathrm{E}(Y_n \mid \mathfrak{G}) \qquad a.s..$$

Let now $\varphi : J \to \mathbb{R}$ be convex; then there is a sequence a_n of linear mappings such that, for all $x \in J$, $\varphi(x) = \sup_n a_n(x)$. Thus, we can estimate

$$\begin{aligned} \mathrm{E}(\varphi(X) \mid \mathfrak{G}) &= \mathrm{E}(\sup_n a_n(X) \mid \mathfrak{G}) \\ &\geq \sup_n \mathrm{E}(a_n(X) \mid \mathfrak{G}) \\ &= \sup_n a_n(\mathrm{E}(X \mid \mathfrak{G})) \\ &= \varphi(\mathrm{E}(X \mid G)) \ . \end{aligned}$$

□

Remark 2. Special case: $J = \mathbb{R}$ and $\varphi(u) = |u|^{p/q}$ with $1 \leq q \leq p$. Then

$$(\mathrm{E}(|X|^q \mid \mathfrak{G}))^{1/q} \leq (\mathrm{E}(|X|^p \mid \mathfrak{G}))^{1/p}$$

for $X \in \mathfrak{L}^p$; further,

$$\mathrm{E}(|\,\mathrm{E}(X \mid \mathfrak{G})|^p) \leq \mathrm{E}[\mathrm{E}(|X|^p \mid \mathfrak{G})] = \mathrm{E}\,|X|^p \ . \qquad (2)$$

Estimate (2) shows that $\mathrm{E}(\cdot \mid \mathfrak{G}) : L^p \to L^p$ is a continuous linear operator with norm 1; by Remark 1, it is idempotent (i.e., a projection). In particular, for $p = 2$, this means that $\mathrm{E}(\cdot \mid \mathfrak{G})$ is the *orthogonal* projection on the closed linear subspace $L^2(\Omega, \mathfrak{G}, P)$.

Conditional expectations are particularly interesting if $\mathfrak{G}$ is encoding 'knowledge obtainable by evaluating a r.v. '$Y : \Omega \to \Omega'$. This is formalized by setting $\mathfrak{G} = \sigma(Y)$.

Definition 2. Let $X : \Omega \to \mathfrak{R}$, $Y : \Omega \to \Omega'$ be measurable. The *conditional expectation of X given Y* is

$$\mathrm{E}(X \mid Y) := \mathrm{E}(X \mid \sigma(Y)).$$

The most fundamental insight about conditional expectations given Y is that **they are functions of** Y; indeed, by Theorem II.2.8, there exists a measurable mapping $g : \Omega' \to \mathbb{R}$ such that the $\sigma(Y)$–measurable r.v. $\mathrm{E}(X \mid Y)$ can be factorized as

$$\mathrm{E}(X \mid Y) = g(Y) \ .$$

Further, any two such mappings g are equal P_Y–a.s..

The next definition is mildly confusing at first, but is most useful if applied properly.

Definition 3. In the above situation, $g(y)$ is called the *conditional expectation of X given $Y = y$,* written

$$\mathrm{E}(X \,|\, Y = y) = g(y) \; .$$

Analogously,

$$P(A \,|\, Y = y) := \mathrm{E}(\mathbf{1}_A \,|\, Y = y) \; .$$

Note that we do *not* naïvely condition on the event $\{Y = y\}$, since this usually is an event of probability zero. However, if $\{Y = y\}$ has positive probability, the definitions coincide with the above elementary ones of conditional probabilities and expectations.

Example 2. Let $(\Omega, \mathfrak{A}, P) = ([0,1], \mathfrak{B}([0,1]), \lambda)$ and $(\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$, and define

$$X(\omega) = \omega^2, \qquad\qquad Y(\omega) = \begin{cases} 1, & \text{if } \omega \in [0, 1/2], \\ \omega - 1/2, & \text{if } \omega \in \,]1/2, 1]. \end{cases}$$

Then

$$\sigma(Y) = \{A \cup B : A \in \{\emptyset, [0, 1/2]\},\; B \subset \,]1/2, 1]\,,\; B \in \mathfrak{A}\}$$

and it is not difficult to check that

$$\mathrm{E}(X \,|\, Y)(\omega) = \begin{cases} 1/12, & \text{falls } \omega \in [0, 1/2], \\ \omega^2, & \text{falls } \omega \in \,]1/2, 1] \; . \end{cases}$$

This entails

$$\mathrm{E}(X \,|\, Y = y) = \begin{cases} 1/12, & \text{falls } y = 1, \\ (y + 1/2)^2, & \text{falls } y \in \,]0, 1/2]. \end{cases}$$

(Note that $P(\{Y = y\}) = 0$ for all $y \in \,]0, 1/2]$.)

Remark 3. For measurable $A' \subset \Omega'$ we have by the transformation theorem that

$$\mathrm{E}\, X \mathbf{1}_{A'}(Y) = \mathrm{E}(\mathrm{E}(X \,|\, Y) \mathbf{1}_{A'}(Y)) = \mathrm{E}(X \,|\, Y = y)\, P_Y(dy) \tag{3}$$

and in particular

$$P(A \cap \{Y \in A'\}) = \int_{A'} P(A \,|\, Y = y)\, P_Y(dy)$$

for $A \in \mathfrak{A}$. This is a continuous analogue of the formula of total probability. Equation (3) characterizes the function $\mathrm{E}(X \,|\, Y = \cdot)$; if $g' : \Omega' \to \mathbb{R}$ is measurable and satisfies

$$\mathrm{E}\, X \mathbf{1}_{A'}(Y) = \int_{A'} g'(y)\, P_Y(dy) \; , \qquad\qquad \forall A' \in \mathfrak{A}',$$

then $g' = \mathfrak{E}(X \,|\, Y = \cdot)$ P_Y–a.s..

The following theorem reveals a fact of utmost importance for both probability and statistics: $\mathrm{E}(X \,|\, Y)$ is the **best estimator for X using Y** concerning the mean square error. Compare with Übung 10.4 and Lemma 4.

Theorem 3. For $X \in \mathfrak{L}^2$ and any measurable $\varphi : \Omega' \to \mathbb{R}$ we have

$$\mathrm{E}\big(X - \mathrm{E}(X \,|\, Y)\big)^2 \le \mathrm{E}\big(X - \varphi \circ Y\big)^2 \;;$$

equality holds iff $\varphi = \mathrm{E}(X \,|\, Y = \cdot)$ P_Y-a.s..

Proof. Let $Z^* = \mathrm{E}(X \,|\, Y)$ and $Z = \varphi \circ Y$. By ((2)), $Z \in \mathfrak{L}^2$; we can assume that also $Z \in \mathfrak{L}^2$. Then

$$\mathrm{E}(X - Z)^2 = \mathrm{E}(X - Z^*)^2 + \underbrace{\mathrm{E}(Z^* - Z)^2}_{\ge 0} + 2 \cdot \mathrm{E}((X - Z^*)(Z^* - Z)).$$

We employ Lemma 1 and 2:

$$\begin{aligned}\mathrm{E}((X - Z^*)(Z^* - Z)) &= \int_\Omega \mathrm{E}((X - Z^*)(Z^* - Z) \,|\, Y)\, dP \\ &= \int_\Omega (Z^* - Z) \cdot \mathrm{E}((X - Z^*) \,|\, Y)\, dP \\ &= \int_\Omega (Z^* - Z) \cdot \underbrace{(\mathrm{E}(X \,|\, Y) - Z^*)}_{=0}\, dP.\end{aligned}$$

□

Markov kernels have a natural and important connection to conditional expectations. Let $(\Omega, \mathfrak{A})$, $(\Omega', \mathfrak{A}')$, $(\Omega'', \mathfrak{A}'')$ measurable spaces, P a probability measure on Ω, and $Y : \Omega \to \Omega'$, $X : \Omega \to \Omega''$ random elements.

Lemma 5. For a mapping $P_{X \,|\, Y} : \Omega' \times \mathfrak{A}'' \to \mathbb{R}$, TFAE:

(i) $P_{X \,|\, Y}$ Markov-kernel from $(\Omega', \mathfrak{A}')$ to $(\Omega'', \mathfrak{A}'')$ such that

$$P_{(Y,X)} = P_Y \times P_{X \,|\, Y} \;; \tag{4}$$

(ii) for any $y \in \Omega'$, $P_{X \,|\, Y}(y, \cdot)$ is a probability measure on $(\Omega'', \mathfrak{A}'')$ and for arbitrary $A'' \in \mathfrak{A}''$ we have

$$P_{X \,|\, Y}(\cdot, A'') = P(\{X \in A''\} \,|\, Y = \cdot).$$

If these conditions hold, X and Y are independent if and only if

$$P_{X \,|\, Y}(y, \cdot) = P_X$$

P_Y-a.e.(in y).

Proof. Let $A' \in \mathfrak{A}'$ and $A'' \in \mathfrak{A}''$; then by definition of the product measure,

$$(P_Y \times P_{X \,|\, Y})(A' \times A'') = \int_{A'} P_{X \,|\, Y}(y, A'')\, P_Y(dy) \;;$$

on the other hand, by Remark 3, it follows that

$$P_{(Y,X)}(A' \times A'') = \int_{A'} P(\{X \in A''\} \,|\, Y = y)\, P_Y(dy).$$

From this the equivalence of (i) and (ii) easily follows. Assertion about independence: Übung 14.2. □

Definition 4. A Markov kernel $P_{X\,|\,Y}$ from $(\Omega', \mathfrak{A}')$ to $(\Omega'', \mathfrak{A}'')$ with the property (4) is called a *regular conditional probability of X given Y.* The representation (4) is called the *desintegration of the common distribution* $P_{(X,Y)}$.

Remark 4. Let $X = \mathrm{id} : \Omega \to \Omega$, and insert in (4) pairwise disjoint sets $A_1, A_2, \dots \in \mathfrak{A}$. For $A' \in \mathfrak{A}'$ it follows

$$\begin{aligned}\int_{A'} P\left(\bigcup_{i=1}^{\infty} A_i \,|\, Y = y\right) P_Y(dy) &= P\left(\bigcup_{i=1}^{\infty} A_i \cap \{Y \in A'\}\right) \\ &= \sum_{i=1}^{\infty} P(A_i \cap \{Y \in A'\}) \\ &= \sum_{i=1}^{\infty} \int_{A'} P(A_i \,|\, Y = y)\, P_Y(dy) \\ &= \int_{A'} \sum_{i=1}^{\infty} P(A_i \,|\, Y = y)\, P_Y(dy).\end{aligned}$$

Thus, we have equality P_Y–a.s. of

$$P\left(\bigcup_{i=1}^{\infty} A_i \,|\, Y = \cdot\right) = \sum_{i=1}^{\infty} P(A_i \,|\, Y = \cdot).$$

It is important to note that the null set where this equality does not hold is, in general, dependent on the sets A_i.

Example 3. Consider a Markov kernel K from $(\Omega', \mathfrak{A}')$ nach $(\Omega'', \mathfrak{A}'')$, and a probability measure μ on $(\Omega', \mathfrak{A}')$. On the product space

$$(\Omega, \mathfrak{A}) = (\Omega' \times \Omega'', \mathfrak{A}' \otimes \mathfrak{A}'')$$

let

$$Y(\omega', \omega'') = \omega', \qquad X(\omega', \omega'') = \omega''.$$

Then, under the probability measure $P := \mu \times K$, the pair (Y, X) of random variables models the result of 'first draw randomly from Ω' according to μ, then draw randomly from Ω'' according to the first result and K'. Our new knowledge yields that K is a regular distribution of X given Y; in particular,

$$K(y, A) = \mathbb{P}\left(X \in A | Y = y\right).$$

Finding a desintegration of a common distribution $P_{(X,Y)}$ thus can be considered the inverse problem to the construction of $P_{(X,Y)}$ with the kernel; it is an important tool in statistics to model the dependence of X from Y. The above relation reveals that this is equivalent to determine the conditional probabilities given Y. This is another reason why efficient methods to estimate and/or approximate conditional expectations and probabilities is of great interest in statistics and probability.

Example 4. Let $(\Omega', \mathfrak{A}') = (\Omega'', \mathfrak{A}'') = (\mathbb{R}, \mathfrak{B})$ and assume that $P_{(X,Y)}$ has the Lebesgue density f. It is then trivial to obtain from Fubini's theorem that P_Y has Lebesgue density

$$h(y) = \int_{\mathbb{R}} f(y, \cdot)\, d\lambda_1, \qquad y \in \mathbb{R}.$$

We claim that the function

$$f(x \,|\, y) = \begin{cases} f(y,x)/h(y) & \text{if } h(y) > 0 \\ 1_{[0,1]}(x) & \text{otherwise .} \end{cases}$$

is a conditional density of X given Y, i.e.,

$$P_{X\,|\,Y}(y, A) := \int_A f(x \,|\, y)\, dx$$

is a regular probability distribution of X given Y. Indeed,

$$\begin{aligned} P_{(Y,X)}(A' \times A'') &= \int_{A'} \int_{A''} f(x \,|\, y)\, \lambda_1(dx) \cdot h(y)\, \lambda_1(dy) \\ &= \int_{A'} P_{X\,|\,Y}(y, A'')\, P_Y(dy). \end{aligned}$$

In Übung 14.2 it will be shown that

$$\mathrm{E}(X \,|\, Y = y) = \int_{\mathbb{R}} x\, P_{X\,|\,Y}(y, dx) = \int_{\mathbb{R}} x \cdot f(x \,|\, y)\, \lambda_1(dx). \tag{5}$$

It is not clear at all whether for given (X, Y), a regular probability distribution (and thus a modelation as two-step experiment using a Markov kernel) always exists. We quote without proof a far–reaching positive result:

Theorem 4. If $(\Omega', \mathfrak{A}') = (M, \mathfrak{B}(M))$ where M is a complete and separable metric space, then for any pair (X, Y) there is a (essentially unique) regular conditional distribution of X given Y.

Proof. See Gänssler, Stute (1977, Kap. V.3) oder Yeh (1995, App. C). □

Theorem 5. If $X \in \mathfrak{L}^1(\Omega, \mathfrak{A}, P)$ and $P_{\mathrm{id}\,|\,Y}$ is a regular conditional probability,

$$\mathrm{E}(X \,|\, Y = y) = \int_\Omega X(\omega)\, P_{\mathrm{id}\,|\,Y}(y, d\omega).$$

Proof. We have to prove that

(i) $\int_\Omega X(\omega)\, P_{\mathrm{id}\,|\,Y}(\cdot, d\omega)$ $\mathfrak{A}'$-measurable,

(ii) $\int_{A'} \int_\Omega X(\omega)\, P_{\mathrm{id}\,|\,Y}(y, d\omega)\, P_Y(dy) = \int_{\{Y \in A'\}} X\, dP$ for $A' \in \mathfrak{A}'$.

This is straightforward with algebraic induction over X. □

We have found in this section final and satisfying answers to the questions posed in Example I.4.