Rolf Steyer

Introduction to Probability and Conditional Expectation

Version: May 21, 2021

University of Jena

Preface

In 2017, together with Werner Nagel, I published a book entitled *Probability and Conditional Expectation — Fundamentals for the Empirical Sciences*. We started this project with the idea to have a short but precise reference to all concepts and propositions that are relevant for understanding, in terms of probability theory, the mathematical essence of what we are talking about in applied statistical research. From my point of view, this project was quite successful, except for one point: Instead of the originally intended 50 to 100 pages, we ended up with more than 550 pages. The present book condenses the previous one to about 150 pages. The price was omitting some chapters and including proofs only if not found in Steyer and Nagel (2017). The result is a relatively brief introduction to probability and conditional expectation. However, I also added some new propositions that proved to be useful in my work on causality and on latent variables.

Overview

This book comprises some chapters on important concepts of probability theory. In chapter 1, we start with the components of a *probability space*, which can be used for the mathematical representation of a concrete random experiment. In chapter 2, we introduce the concept of a *random variable* as a measurable mapping and deal with concepts related to a random variable such as its *distribution*, *density*, and *distribution function*. Chapter 3 deals with the concepts of *expectation*, *covariance*, *correlation*, and *linear quasi-regression*. Chapter 4 is devoted to the concept of a *conditional expectation*, chapter 5 to conditional expectations with respect to a conditional-probability measure, and chapter 6 to *conditional independence* of events and of random variables given a random variable.

For Whom is it?

This book has been written for two kinds of readers. The first are applied statisticians and empirical researchers who want to understand, in terms of probability theory, what they estimate and test in their empirical studies. The second kind of readers are mathematicians who want to understand in these terms what applied statisticians and empirical researchers estimate and which kind of hypotheses they test in their research. Both kinds of readers are potential contributors to the methodology of empirical sciences.

How to Use it

This book is written such that it should be studied in the sequence of chapters. Chapter 5 might be omitted at first reading. It is a prerequisite, however, for studying causal effects. For a second reading it is certainly useful to consult Steyer and Nagel (2017), not only for many proofs but also for a more extensive treatment of some of the topics. Whenever we refer to a definition, theorem, section, or any other part of this book we use the abrevi-ation SN. For example, SN-Definition 2.1 SN-Equation (2.41) refer to Definition 2.1 and Equation (2.41) in chapter 2 of Steyer and Nagel (2017).

Acknowledgements

As mentioned above, this book heavily builds on Steyer and Nagel (2017). Therefore, thanks are due to all who helped writing and checking that volume. In the first place this is Ivailo Partchev who prepared the LaTeX framework, many of the figures, tables, and boxes. Some of the figures have been produced by Désirée Thielemann, who also cared for references, read some of the chapters, and hinted at errors. For support with respect to LaTeX, finding errors or suggesting other improvements I also thank Karoline Bading, Marcel Bauer, Sonja Hahn, Gregor Kappler, Christoph Kiefer, Andreas Neudecker, Axel Mayer, Erik Sengewald, Jan Plötner, and Tom Landes. Finally, this book too, has been written with the help of Werner Nagel who not only served as a coauthor of the previous book but also helped proving some of the new theorems.

Jena, Germany, May 1, 2021

Contents

List of Figures

List of Tables

Chapter 1
Probability and Conditional Probability

We start with the definition of a *probability space* and its components, the *set of possible outcomes*, also called the *sample space*, the *set of possible events*, and the *probability measure* assigning a probability to each possible event. Then we turn to *conditional probability* and the most important theorems related to conditional probability: the *multiplication rule*, the *theorem of total probability*, and *Bayes' Theorem*. Next, we define *independence of events* and *independence of sets of events* with respect to a probability measure. A section on *conditional independence of events given an event* concludes this chapter.

1.1 Probability Space

In this section we introduce the three components of a probability space, a *set of possible outcomes* (of the random experiment), a *set of possible events* to be considered in such a random experiment, and a *probability measure* that assigns a probability to each of these possible events. All three components have a certain mathematical structure. The triple of these components is the mathematical representation of a *random* or *chance experiment*, which itself is not a mathematical term, but the kind of empirical phenomenon that we consider applying probability theory and talking about true probabilities, true dependencies (e. g., true correlations), true effects (e. g., true mean differences), true causal dependencies and true causal effects. Adding the term 'true' in this context, we simply emphasize that we will not talk about *estimates* of correlations, mean differences, and so on in a data sample but about *what we would like to estimate* using a data sample.[1] These true parameters are also those terms that we refer to in our hypotheses and theories in the empirical sciences. However, they are also what we are interested in if we have to act in practice. Once this is said, we can drop the term 'true' in the sequel.

Example 1.1 [Joe and Ann With Randomized Assignment] In order to illustrate the practical relevance of probabilities consider Table 1.1 specifying a *random experiment* that is composed of three parts.

(1) A person is sampled from a set of two persons, Joe and Ann, with identical probabilities for each person u, that is, with probability $P(U{=}u) = .5$.
(2) Both persons obtain treatment $(X{=}1)$ with probability $P(X{=}1|U{=}u) = .4$. (This number reflects randomized assignment to treatment that does not depend on the sampled person and its attributes.)
(3) Given that Joe is sampled and *not treated*, the probability of success is .7, that is,

[1] Another term that is often used instead of 'true mean' is 'population mean'. However, the term 'population' tends to evoke even more misleading connotations than the term 'true'.

Table 1.1. Joe and Ann with randomized assignment – compressed table

Person u	$P(U=u)$	$P(X=1 \mid U=u)$	$P(Y=1 \mid U=u, X=0)$	$P(Y=1 \mid U=u, X=1)$
Joe	.5	.4	.7	.8
Ann	.5	.4	.2	.4

$$P(Y=1 \mid U=Joe, X=0) = .7.$$

Given that Joe is sampled and *treated*, the probability of success is .8, that is,

$$P(Y=1 \mid U=Joe, X=1) = .8.$$

In contrast, given that Ann is sampled and *not treated*, the probability of success is .2, that is,

$$P(Y=1 \mid U=Ann, X=0) = .2.$$

Finally, given Ann is sampled and *treated*, the probability of success is .4, that is,

$$P(Y=1 \mid U=Ann, X=1) = .4.$$

It is the first two numbers that are relevant for making a decision about a treatment for Joe, and the last two numbers that are relevant for making this decision for Ann. However, these numbers can only be meaningfully defined and fully understood in a formal or mathematical framework, the probability space $(\Omega, \mathscr{A}, P)$. ◁

1.1.1 Set of Possible Outcomes

The *set of possible outcomes* (of a random experiment), denoted by Ω, also called *sample space*, is the first component of a probability space. If we think about a specific random experiment, then we should know its set of possible outcomes. Otherwise we run the risk of misunderstandings. The mathematical structure of Ω is simply the structure of a set. This means that we have to know its elements, the *possible outcomes*. We use the term *possible* outcomes in order to communicate and keep in mind that we are always talking about a random experiment from the *pre-factual* or *a priori perspective*, or that is, from the perspective *before* it is actually conducted. Even if a random experiment is already executed, then, talking about the probabilities of certain events, we do *as if* it had not yet been conducted. Only in this way does it make sense to talk about the *probability* of an event.

Example 1.2 [Joe and Ann With Randomized Assignment] Table 1.2 refers to the same random experiment as Table 1.1. However, this new table presents the same information in a more explicit way. The first column in Table 1.2 contains the eight elements of the set Ω of possible outcomes of the random experiment considered. Hence, in this example,

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_8\} = \big\{ (Joe, no, -),\ (Joe, no, +), \ldots,\ (Ann, yes, +) \big\}, \tag{1.1}$$

Table 1.2. Joe and Ann with randomized assignment

Outcomes ω_i (Unit, Treatment, Success)	$P(\{\omega_i\})$	$P^{X=0}(\{\omega_i\})$	$P^{X=1}(\{\omega_i\}) = P^B(\{\omega_i\})$	Observables: Person variable U	Treatment variable X	Outcome variable Y	Conditional probabilities: $P(Y=1 \mid X, U)$	$P(Y=1 \mid X)$	$P(X=1 \mid U)$	$P^{X=0}(Y=1 \mid U)$	$P^{X=1}(Y=1 \mid U)$
$\omega_1 = (Joe, no, -)$	.09	.15	0	*Joe*	0	0	.7	.45	.4	.7	.8
$\omega_2 = (Joe, no, +)$	.21	.35	0	*Joe*	0	1	.7	.45	.4	.7	.8
$\omega_3 = (Joe, yes, -)$	.04	0	.1	*Joe*	1	0	.8	.6	.4	.7	.8
$\omega_4 = (Joe, yes, +)$	.16	0	.4	*Joe*	1	1	.8	.6	.4	.7	.8
$\omega_5 = (Ann, no, -)$	.24	.4	0	*Ann*	0	0	.2	.45	.4	.2	.4
$\omega_6 = (Ann, no, +)$	.06	.1	0	*Ann*	0	1	.2	.45	.4	.2	.4
$\omega_7 = (Ann, yes, -)$	.12	0	.3	*Ann*	1	0	.4	.6	.4	.2	.4
$\omega_8 = (Ann, yes, +)$	.08	0	.2	*Ann*	1	1	.4	.6	.4	.2	.4

Note. The probabilities of the elementary events $P(\{\omega_i\})$ are fictive. The terms illustrated in this table are introduced in this and the next chapters. The event B occurring in the term $P^B(\{\omega_i\})$ denotes the event that the drawn person (whoever this is) is treated (see Example 1.36).

that is, Ω consists of the eight triples $(Joe, no, -)$, $(Joe, no, +)$, …, $(Ann, yes, +)$. If we define $\Omega_U = \{Joe, Ann\}$, $\Omega_X = \{no, yes\}$, and $\Omega_Y = \{-, +\}$, then Ω can also be written as the Cartesian product

$$\Omega = \Omega_U \times \Omega_X \times \Omega_Y. \tag{1.2}$$

In this example, the set Ω of possible outcomes has eight elements, namely all triples whose first component is an element of Ω_U, whose second component is an element of Ω_X, and whose third component is an element of Ω_Y (see the first column of Table 1.2). The other terms occuring in Table 1.2 will be introduced in the sections and chapters to come. ◁

Example 1.3 [Nonorthogonal Factors] Another example is in Table 1.3. If we define the sets

$$\Omega_U = \{Tom, Tim, Joe, Jim, Ann, Eva, Sue, Mia\},$$
$$\Omega_X = \{control, treatment\,1, treatment\,2\},$$

and

$$\Omega_Y = \mathbb{R},$$

then Ω is the Cartesian product of these three sets, that is,

$$\Omega = \Omega_U \times \Omega_X \times \Omega_Y.$$

In contrast to Example 1.2, now the set Ω has an uncountable number of elements, which would be necessary if we want to consider a response variable Y that can take on as a

Table 1.3. Random experiment with two nonorthogonal factors

Person u	Educational status z	$P(U=u)$	$P(X=1 \mid U=u)$	$P(X=2 \mid U=u)$	$E(Y \mid U=u, X=0)$	$E(Y \mid U=u, X=1)$	$E(Y \mid U=u, X=2)$
Tom	*low*	1/8	10/60	3/60	120	100	80
Tim	*low*	1/8	18/60	9/60	120	100	80
Joe	*med*	1/8	26/60	17/60	90	90	70
Jim	*med*	1/8	26/60	17/60	100	100	80
Ann	*med*	1/8	26/60	17/60	120	100	100
Eva	*med*	1/8	26/60	17/60	130	110	110
Sue	*hi*	1/8	12/60	44/60	60	100	140
Mia	*hi*	1/8	16/60	36/60	60	100	140

value any real number, representing, for example, the degree of success. This is also the reason why this example can only be presented in the compressed form (see Table 1.3), which does not show the possible values of Y but only its conditional expectation values $E(Y \mid X=x, U=u)$, which will be introduced in chapter 3 [see Eq. (3.23)]. Note that in many empirical applications in the empirical sciences, Ω_Y may just be a subset of $\mathbb{R}$. $\triangleleft$

1.1.2 Set of Possible Events

The second component of a probability space, denoted by $\mathscr{A}$, is a *set of possible events*. It is a set of subsets of the set Ω of possible outcomes that has the properties of a *σ-algebra*. An element A of a σ-algebra $\mathscr{A}$ is called a *measurable set*. In the context of a probability measure on this σ-algebra, a measurable set is also called an *event*. We use A^c to denote the *complement* of a measurable set A, that is $A^c := \Omega \setminus A$, where $\Omega \setminus A$ denotes the *set difference*, which, by definition, consists of all elements of Ω that are not elements of the set A.

Definition 1.4 [**σ-Algebra, Measurable Set, and Measurable Space**]
A set $\mathscr{A}$ of subsets of a nonempty set Ω is called a σ-algebra *(or* σ-field*) on Ω, if:*

(a) $\Omega \in \mathscr{A}$.
(b) If $A \in \mathscr{A}$*, then* $A^c \in \mathscr{A}$.
(c) If $A_1, A_2, \ldots \in \mathscr{A}$*, then* $\bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$.

An element A of a σ-algebra $\mathscr{A}$ is called a measurable set*. The pair $(\Omega, \mathscr{A})$ is called a* measurable space*.*

Remark 1.5 [Closure With Respect to Set Operations] Condition (c) means that a σ-algebra is closed with respect to *countable unions* of sets $A_1, A_2, \ldots \in \mathscr{A}$. However, in conjunction with (a) and (b) this implies that a σ-algebra is also closed with respect to *finite* unions of sets $A_1, \ldots, A_n \in \mathscr{A}$. That is, if $A_1, \ldots, A_n \in \mathscr{A}$, then $A_1 \cup \ldots \cup A_n \in \mathscr{A}$. Furthermore, although condition (c) explicitly requires only that a σ-algebra is closed with respect to countable unions, Definition 1.4 implies that a σ-algebra is closed also with respect to countable intersections $A_1 \cap A_2 \cap \ldots$ of elements of $\mathscr{A}$. Definition 1.4 also implies that unions such as $A_1 \cup A_2$, intersections $A_1 \cap A_2$, and set differences $A_1 \setminus A_2$ are elements of $\mathscr{A}$. In other words, if A_1 and A_2 are elements of $\mathscr{A}$, then $A_1 \cup A_2$, $A_1 \cap A_2$, and $A_1 \setminus A_2$ are elements of $\mathscr{A}$ as well, provided that $\mathscr{A}$ is a σ-algebra (for more details see SN-section 1.2.) ◁

An important property of σ-algebras is formulated in the following theorem (see SN-Th. 1.12 for a proof).

Theorem 1.6 [The Intersection of σ-Algebras is a σ-Algebra]
Let I be a nonempty (finite, countable, or uncountable) index set and let all $\mathscr{A}_i$, $i \in I$, be σ-algebras on the set Ω. Then the intersection $\bigcap_{i \in I} \mathscr{A}_i$ of these σ-algebras is also a σ-algebra on Ω.

Theorem 1.6 allows us to define the σ-*algebra generated by a set system on* Ω, that is, by a set of subsets of Ω.

Definition 1.7 [σ-Algebra Generated by a Set System]
Let $\mathscr{E}$ be a set of subsets of Ω and let $(\mathscr{A}_i, i \in I)$ be the family of all σ-algebras on Ω that contain $\mathscr{E}$ as a subset. Then we define

$$\sigma(\mathscr{E}) := \bigcap_{i \in I} \mathscr{A}_i \tag{1.3}$$

and call it the σ-algebra generated by $\mathscr{E}$. The set $\mathscr{E}$ is also called a generating system of $\sigma(\mathscr{E})$.

Example 1.8 [Intersection of Two σ-Algebras] If Ω is a set, then $\mathscr{A}_0 = \{\Omega, \emptyset\}$ is a σ-algebra on Ω. If $A \subset \Omega$, then $\mathscr{A}_1 = \{A, A^c, \Omega, \emptyset\}$ is a σ-algebra on Ω (see Exercises 1-1 and 1-2). Their intersection $\mathscr{A}_0 \cap \mathscr{A}_1 = \{\Omega, \emptyset\}$ is a σ-algebra on Ω, too. The σ-algebra $\mathscr{A}_1$ is generated by the set system $\{A\}$ on Ω, but also by the set systems $\{A^c\}$, $\{A, A^c\}$, $\{A, A^c, \Omega\}$, and $\{A, A^c, \emptyset\}$, for instance (see Exercise 1-3). ◁

Remark 1.9 [A Property of a Set System and its Generated σ-Algebra] Assume $(\Omega, \mathscr{A})$ is a measurable space. Then

$$\mathscr{E} \subset \mathscr{A} \quad \Rightarrow \quad \sigma(\mathscr{E}) \subset \mathscr{A}. \tag{1.4}$$

That is, if the generating system $\mathscr{E}$ is subset of a σ-algebra $\mathscr{A}$, then its generated σ-algebra $\sigma(\mathscr{E})$ is also a subset of $\mathscr{A}$ (see Exercise 1-4). ◁

Remark 1.10 [Monotonicity of Generated σ-Algebras] Let $\mathscr{E}_1$ and $\mathscr{E}_2$ be set systems on Ω. Then

$$\mathscr{E}_1 \subset \mathscr{E}_2 \quad \Rightarrow \quad \sigma(\mathscr{E}_1) \subset \sigma(\mathscr{E}_2) \tag{1.5}$$

(see SN-Exercise 1-9). ◁

Remark 1.11 [Union of Two σ-Algebras] Let $(\Omega, \mathscr{A})$ be a measurable space. Then

$$\mathscr{C}, \mathscr{D} \subset \mathscr{A} \quad \Rightarrow \quad \mathscr{C} \cup \mathscr{D} \subset \mathscr{A} \tag{1.6}$$

$$\Rightarrow \quad \sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{A} \tag{1.7}$$

(see Exercise 1-5). ◁

Example 1.12 [Joe and Ann With Randomized Assignment] We may choose $\mathscr{A} = \mathscr{P}(\Omega)$ to be the *power set*, that is, the set of all subsets of the set Ω that has been specified in Equation (1.1). The power set of a set Ω is always a σ-algebra on Ω. The power set contains 2^n elements, where n is the number of elements of Ω. In the example presented in Table 1.2, $n = 8$. Hence, in this example, the power set consists of $2^n = 256$ elements. ◁

Example 1.13 [Nonorthogonal Factors] In this example, we cannot use the power set of Ω as a σ-algebra !, because this would lead to contradictions (see SN-Rem. 1.8). Instead, we use the *product of the σ-algebras* $\mathscr{A}_U = \mathscr{P}(\Omega_U)$, $\mathscr{A}_X = \mathscr{P}(\Omega_X)$, and the *Borel σ-algebra* $\mathscr{B}$ on the set $\mathbb{R}$ of real numbers (see Def. 1.15.) ◁

Remark 1.14 [Borel σ-Algebra] The *Borel σ-algebra* $\mathscr{B}$ on $\mathbb{R}$ is defined by $\mathscr{B} = \sigma(\mathscr{I})$, where $\mathscr{I} = \{(a, b]\colon a, b \in \mathbb{R},\ a < b)\}$ is the set system on $\mathbb{R}$ of all left half-open intervals in $\mathbb{R}$. The Borel σ-algebra contains as elements all singletons $\{\alpha\}$, $\alpha \in \mathbb{R}$, as well as all (open, half open, and closed) intervals, their countable unions and intersections (for more details see SN-sect. 1.2.2). ◁

If we consider composite random experiments that consist of several simple random experiments (see Examples 1.2 and 1.3), then the *product of σ-algebras* is an important concept. The following definition is adapted from SN-Def. 1.31.

Definition 1.15 [Product σ-Algebra]
Let $(\Omega_1, \mathscr{A}_1), \ldots, (\Omega_m, \mathscr{A}_m)$ be measurable spaces. Then

$$\mathscr{A}_1 \otimes \ldots \otimes \mathscr{A}_m := \bigotimes_{i=1}^{m} \mathscr{A}_i := \sigma\Big(\Big\{ \mathop{\times}_{i=1}^{m} A_i \colon A_i \in \mathscr{A}_i, i = 1, \ldots, m \Big\}\Big) \tag{1.8}$$

is called the product of the σ-algebras $\mathscr{A}_1, \ldots, \mathscr{A}_m$ and $(\Omega, \mathscr{A})$ the product of the measurable spaces $(\Omega_1, \mathscr{A}_1), \ldots, (\Omega_m, \mathscr{A}_m)$ if $\Omega = \Omega_1 \times \ldots \times \Omega_m$ and $\mathscr{A} = \mathscr{A}_1 \otimes \ldots \otimes \mathscr{A}_m$.

Note that the product σ-algebra $\mathscr{A}_1 \otimes \ldots \otimes \mathscr{A}_m$ *is not* the Cartesian product $\mathscr{A}_1 \times \ldots \times \mathscr{A}_m$. Instead, the product σ-algebra is generated by the set system of all Cartesian products of elements of the σ-algebras $\mathscr{A}_1, \ldots, \mathscr{A}_m$ (see Def. 1.7).

Example 1.16 [Countable Sets and Product σ-Algebra] Let $\Omega_1, \ldots, \Omega_m$ be finite or countable sets and $\mathscr{A}_1, \ldots, \mathscr{A}_m$ be their power sets. Then

$$\bigotimes_{i=1}^{m} \mathscr{A}_i = \mathscr{P}\Big(\mathop{\times}_{i=1}^{m} \Omega_i \Big), \tag{1.9}$$

that is, $\mathscr{A}_1 \otimes \ldots \otimes \mathscr{A}_m$ is the power set on $\Omega = \Omega_1 \times \ldots \times \Omega_m$ (for a proof see SN-Exercise 1-14). ◁

Example 1.17 [Joe and Ann With Randomized Assignment] Consider again the set

$$\Omega = \Omega_U \times \Omega_X \times \Omega_Y$$

of possible outcomes, where $\Omega_U = \{\mathit{Joe}, \mathit{Ann}\}$, $\Omega_X = \{\mathit{yes}, \mathit{no}\}$, and $\Omega_Y = \{-, +\}$. Furthermore, consider the σ-algebras $\mathscr{A}_1 = \mathscr{P}(\Omega_U)$, $\mathscr{A}_2 = \mathscr{P}(\Omega_X)$, and $\mathscr{A}_3 = \mathscr{P}(\Omega_Y)$, respectively. Hence, we specified the measurable spaces $(\Omega_U, \mathscr{A}_1)$, $(\Omega_X, \mathscr{A}_2)$, and $(\Omega_Y, \mathscr{A}_3)$. The product σ-algebra $\mathscr{A}_1 \otimes \mathscr{A}_2 \otimes \mathscr{A}_3$ is the σ-algebra generated by the set system

$$\mathscr{E} = \{A_1 \times A_2 \times A_3 \colon A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2, A_3 \in \mathscr{A}_3\},$$

and, according to Example 1.16, the σ-algebra generated by $\mathscr{E}$ is identical to the power set of Ω, which consists of 256 elements. ◁

1.1.3 Probability Measure

The last component of a probability space is a *probability measure*, which assigns a probability to each element of a σ-algebra. This concept has been introduced by Kolmogorov (1933/1977) (for the English version of this book see Kolmogorov, 1956). In the following definition we use $[0,1]$ to denote the *closed interval* of all real numbers between 0 and 1, inclusively.

Definition 1.18 [Probability Measure]
Let $(\Omega, \mathscr{A})$ be a measurable space. Then the function $P\colon \mathscr{A} \to [0,1]$ is called a probability measure on $(\Omega, \mathscr{A})$, if the following conditions hold:

(a) $P(\Omega) = 1$ *(standardization).*
(b) $P(A) \geq 0, \ \forall A \in \mathscr{A}$ *(nonnegativity).*
(c) $A_1, A_2, \ldots \in \mathscr{A}$ *are pairwise disjoint* $\Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ *(σ-additivity).*

Remark 1.19 [Probability of an Event and Probability Space] Let P be a probability measure on $(\Omega, \mathscr{A})$. Then the triple $(\Omega, \mathscr{A}, P)$ is called a *probability space* and a value $P(A)$ of P is called the *probability* of (the event) A. If a probability space $(\Omega, \mathscr{A}, P)$ is used to represent a random experiment, then it contains all information about this random experiment. That is, everything we can ever learn about this random experiment can be computed from the probabilities $P(A)$, $A \in \mathscr{A}$. The most important properties of a probability measure are gathered in Box 1.1. (This box has been adopted from SN-chapter 4. For the general concept of a measure and its properties as well as for other measures than probability measures see SN-ch. 1.) ◁

Remark 1.20 [Elementary Event and Event] Let $(\Omega, \mathscr{A}, P)$ be a probability space. If $A \in \mathscr{A}$, then A is called a (possible) *event*, and if $\{\omega\} \in \mathscr{A}$, then it is called a (possible) *elementary event*. Note the distinction between an *outcome* $\omega \in \Omega$ and an *elementary event* $\{\omega\} \in \mathscr{A}$. Also note that the term *event* is only used in the context of a probability space $(\Omega, \mathscr{A}, P)$.

Box 1.1 Rules of Computation for Probabilities

Let $(\Omega, \mathscr{A}, P)$ be a probability space and, for $\alpha_i \in \mathbb{R}$ define $\sum_{i=1}^{\infty} \alpha_i := \lim_{n\to\infty} \sum_{i=1}^{n} \alpha_i$.
If $A_i \cap A_j = \emptyset$, for all $i, j \in \mathbb{N}$, then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i). \qquad (\sigma\text{-}additivity) \qquad \text{(i)}$$

If $A_i \cap A_j = \emptyset$, for all $i, j \in \{1, \dots, n\}$, $n \in \mathbb{N}$, then

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i). \qquad (finite\ additivity) \qquad \text{(ii)}$$

If $A, B \in \mathscr{A}$, then:

$$P(A) = P(A\cap B) + P(A \setminus B) \qquad \text{(iii)}$$
$$P(A^c) = 1 - P(A) \qquad \text{(iv)}$$
$$P(A) \le P(B), \quad \text{if } A \subset B \qquad (monotonicity) \qquad \text{(v)}$$
$$P(A \setminus B) = P(A) - P(A\cap B) \qquad \text{(vi)}$$
$$P(A\cup B) = P(A) + P(B) - P(A\cap B) \qquad \text{(vii)}$$
$$P(A) = 1 \Rightarrow P(A\cap B) = P(B) \qquad \text{(viii)}$$
$$P(A) = 0 \Rightarrow P(A\cup B) = P(B). \qquad \text{(ix)}$$

Let $A \in \mathscr{A}$ and let $\Omega_0 \subset \Omega$ be finite or countable and $P(\Omega_0) = 1$.
If, for all $\omega \in \Omega_0$, $\{\omega\} \in \mathscr{A}$, then

$$P(A) = \sum_{\omega \in A\cap\Omega_0} P(\{\omega\}). \qquad \text{(x)}$$

If $A_1, A_2, \ldots \in \mathscr{A}$, then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) \le \sum_{i=1}^{\infty} P(A_i). \qquad (\sigma\text{-}subadditivity) \qquad \text{(xi)}$$

Otherwise $A \in \mathscr{A}$ is called a *measurable set*. For simplicity we often drop the long-winded term 'possible' if we talk about outcomes and events. Nevertheless, in applications we continue considering a random experiment from the pre-factual perspective (see again the introductory remarks to sect. 1.1.1). ◁

Remark 1.21 [Probability Space and Random Experiment] Note again the disctinction between a *probability space* and a *random experiment*. The term probability space is a mathematical concept. It does not have any empirical meaning unless we interpret Ω by saying that it represents the set of possible outcomes of a concrete random experiment. In contrast, 'random experiment' is a term of our colloquial language referring to an empirical phenomenon that we might be interested in. Giving Ω a concrete empirical meaning, the σ-algebra $\mathscr{A}$ also obtains a concrete empirical meaning: the set of possible events to be considered *in this concrete random experiment*. Correspondingly, the probability measure P assigns to each of these possible events $A \in \mathscr{A}$ its probability $P(A)$ that refers *to this specific random experiment*. Hence, propositions about probabilities refer to a concrete random experiment and cannot offhandedly be generalized to another random experiment. ◁

Example 1.22 [Joe and Ann With Randomized Assignment] The second column of Table 1.2 displays the assignment of the probabilities to all eight elementary events. In this example, each nonempty element $A \in \mathscr{A}$ is a union of the elementary events $\{\omega_i\}$, $\omega_i \in \Omega$, and because a measure is additive [see Box 1.1 (ii)], the probability measure $P\colon \mathscr{A} \to [0,1]$ is uniquely specified by the second column of Table 1.2 [see Box 1.1 (x)]. The probabilities of all other events we may consider in this random experiment can be computed from the probabilities of the eight elementary events (see Exercise 1-6). Because Ω has been fixed in Example 1.2 and $\mathscr{A}$ in Example 1.12, now the probability space $(\Omega, \mathscr{A}, P)$ is completely specified. ◁

1.2 Conditional Probability

A probability space contains all information about the random experiment considered. However, this information is not easily grasped unless it is processed. A first conceptual tool that helps in processing this information is the concept of a *conditional probability* of an event A given an event B.

1.2.1 Conditional Probability of an Event A Given an Event B

The conditional probability of an event A given an event B can be used to describe the *dependency* of A on B with respect to a probability measure P on $\mathscr{A}$. This concept is also used in order to introduce the concept of a *conditional probability measure*.

Definition 1.23 [Conditional Probability]
Let $(\Omega, \mathscr{A}, P)$ be a probability space, $A, B \in \mathscr{A}$, and $P(B) > 0$. Then

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \tag{1.10}$$

is called the B-conditional probability of A or the conditional probability of A given B (with respect to P).

Example 1.24 [Joe and Ann With Randomized Assignment] Consider again Table 1.2, define $\Omega_U = \{Joe, Ann\}$ and $\Omega_X = \{yes, no\}$, and let

$$C = \Omega_U \times \Omega_X \times \{+\} = \{(Joe, no, +), (Joe, yes, +), (Ann, no, +), (Ann, yes, +)\}$$

be the event that the *drawn person is successful* (no matter who is drawn and whether or not he or she is treated). Furthermore, let

$$B = \Omega_U \times \{yes\} \times \Omega_Y = \big\{(Joe, yes, -), (Joe, yes, +), (Ann, yes, -), (Ann, yes, +)\big\}$$

denote the event that the *drawn person is treated*. Then

$$C \cap B = \Omega_U \times \{yes\} \times \{+\} = \big\{(Joe, yes, +), (Ann, yes, +)\big\}.$$

Additivity of a probability measure [see Box 1.1 (ii)] and Table 1.2 yield

$$\begin{aligned} P(B) &= P\big(\{(Joe, yes, -)\}\big) + P\big(\{(Joe, yes, +)\}\big) + P\big(\{(Ann, yes, -)\}\big) + P\big(\{(Ann, yes, +)\}\big) \\ &= .04 + .16 + .12 + .08 = .4 \end{aligned}$$

and

$$P(C \cap B) = P\big(\{(Joe, yes, +)\}\big) + P\big(\{(Ann, yes, +)\}\big) = .16 + .08 = .24.$$

Applying Equation (1.10) then results in

$$P(C \mid B) = \frac{P(C \cap B)}{P(B)} = \frac{.24}{.4} = .6.$$

Similarly, conditioning on the event B^c that the *drawn person is not treated* yields

$$P(C \mid B^c) = \frac{P(C \cap B^c)}{P(B^c)} = \frac{P(\Omega_U \times \{no\} \times \{+\})}{P(\Omega_U \times \{no\} \times \Omega_Y)} = \frac{.21 + .06}{.09 + .21 + .24 + .06} = .45.$$

In this example, the difference $P(C \mid B) - P(C \mid B^c) = .6 - .45 = .15$ *can* be used to evaluate the average effect of the treatment. In fact, if we additionally condition on the event

$$A = \{Joe\} \times \Omega_X \times \Omega_Y = \big\{(Joe, no, -), (Joe, no, +), (Joe, yes, -), (Joe, yes, +)\big\}$$

that *Joe is drawn*, then we receive

$$P(C \mid B \cap A) - P(C \mid B^c \cap A) = .8 - .7 = .1.$$

Furthermore, additionally conditioning on A^c yields

$$P(C \mid B \cap A^c) - P(C \mid B^c \cap A^c) = .4 - .2 = .2.$$

Obviously, the average of the two individual treatment effects of Joe and Ann is .15. This is also called the *average treatment effect*. ◁

Example 1.25 [Joe and Ann With Self-Selection] In contrast, if we compute the corresponding conditional probabilities for the random experiment presented in Table 1.4, then we receive $P(C \mid B) = .42$ and $P(C \mid B^c) = .6$. Their difference is

$$P(C \mid B) - P(C \mid B^c) = .6 - .42 = -.18,$$

while the two *individual treatment effects* remain unchanged if compared to Table 1.2. They are again

$$P(C \mid B \cap A) - P(C \mid B^c \cap A) = .8 - .7 = .1$$

for Joe and

$$P(C \mid B \cap A^c) - P(C \mid B^c \cap A^c) = .4 - .2 = .2.$$

for Ann. Hence, in the example of Table 1.4, the difference $P(C \mid B) - P(C \mid B^c)$ is completely misleading if interpreted as the average effect of the treatment. In this example, this difference has no causal meaning. ◁

Table 1.4. Joe and Ann with self-selection – explicit table

Possible outcomes ω_i		Observables			Conditional probabilities		
Unit, Treatment, Success	$P(\{\omega_i\})$	Person variable U	Treatment variable X	Response variable Y	$P(X=1 \mid U)$	$P(Y=1 \mid X)$	$P(Y=1 \mid X, U)$
ω_1 = (*Joe, no,* −)	.144	*Joe*	0	0	.04	.6	.7
ω_2 = (*Joe, no,* +)	.336	*Joe*	0	1	.04	.6	.7
ω_3 = (*Joe, yes,* −)	.004	*Joe*	1	0	.04	.42	.8
ω_4 = (*Joe, yes,* +)	.016	*Joe*	1	1	.04	.42	.8
ω_5 = (*Ann, no,* −)	.096	*Ann*	0	0	.76	.6	.2
ω_6 = (*Ann, no,* +)	.024	*Ann*	0	1	.76	.6	.2
ω_7 = (*Ann, yes,* −)	.228	*Ann*	1	0	.76	.42	.4
ω_8 = (*Ann, yes,* +)	.152	*Ann*	1	1	.76	.42	.4

Example 1.26 [Nonorthogonal Factors] In the fourth column of Table 1.3 we displayed the conditional treatment probabilities $P(X=1 \mid U=u)$. These conditional probabilities are the conditional probabilities of the event

$$\{X=1\} := \{\omega \in \Omega\colon X(\omega) = 1\}$$

given the event

$$\{U=Tom\} := \{\omega \in \Omega\colon U(\omega) = Tom\}$$

(see Example 1.3 for the definition of the sets Ω_U, Ω_X, and Ω_Y). Then

$$P(X=1 \mid U=Tom) = 10/60$$

(see the first row in the fourth column of Table 1.3). The notation $P(X=1 \mid U=u)$ will be introduced in Remark 3.22. ◁

1.2.2 Multiplication Rule

Now we treat some theorems involving conditional probabilities. The first one shows how the probability $P(A_1 \cap \ldots \cap A_n)$ can be factorized into a product of an (unconditional) probability and conditional probabilities.

Remark 1.27 [Multiplication Rule for Two and for Three Events] For two events A_1 and A_2, the multiplication rule is

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 \mid A_1), \tag{1.11}$$

provided that $P(A_1) > 0$. This equation directly follows from the definition of the conditional probability $P(A_2 \mid A_1)$ [see Eq. (1.10)].

For three events A_1, A_2, and A_3, the multiplication rule is

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2), \tag{1.12}$$

provided that $P(A_1 \cap A_2) > 0$. This equation follows from the definition of the conditional probability

$$P(A_3 \mid A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}, \tag{1.13}$$

inserting Equation (1.11) for $P(A_1 \cap A_2)$, and solving the resulting equation for the joint probability $P(A_1 \cap A_2 \cap A_3)$. ◁

For n events $A_1, \ldots, A_n$, the multiplication rule is formulated in the following theorem. (For a proof see SN-Theorem 4.22).

Theorem 1.28 [Multiplication Rule]
Let $(\Omega, \mathscr{A}, P)$ *be a probability space, let* $A_1, \ldots, A_n \in \mathscr{A}$, *where* $2 \le n \in \mathbb{N}$, *and assume* $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. *Then*

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1) \cdot \prod_{j=2}^{n} P\left(A_j \,\middle|\, \bigcap_{i=1}^{j-1} A_i\right). \tag{1.14}$$

1.2.3 Theorem of Total Probability

In the next theorem, called the *theorem of total probability*, it is shown how the probability of an event $B \subset A_1 \cup \ldots \cup A_n$ can additively be decomposed into the products $P(B \mid A_i) \cdot P(A_i)$ of conditional and unconditional probabilities. In this theorem it is assumed that the events $A_1, \ldots, A_n$ are *pairwise disjoint*, that is, $A_i \cap A_j = \emptyset$, for all $i, j = 1, \ldots, n$, with $i \neq j$. (For a proof see SN-Theorem 4.25.)

Theorem 1.29 [Theorem of Total Probability]
Let $(\Omega, \mathscr{A}, P)$ *be a probability space, assume that the events* $A_1, \ldots, A_n \in \mathscr{A}$ *are pairwise disjoint, and* $B \subset \bigcup_{i=1}^{n} A_i$. *Then*

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i). \tag{1.15}$$

If, additionally, $P(A_i) > 0$, *for all* $i = 1, \ldots, n$, *then*

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i) \cdot P(A_i). \tag{1.16}$$

Equation (1.15) can be illustrated by Figure 1.1. If the sizes of the areas represent probabilities, then the figure visualizes that $P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)$. The crucial points are:

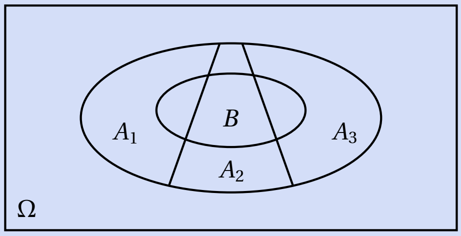

Figure 1.1. Venn diagram illustrating the theorem of total probability

(a) If the events $A_1, \ldots, A_n$ are pairwise disjoint (which in Fig. 1.1 is true for A_1, A_2, A_3) and $B \subset A_1 \cup A_2 \cup A_3$, then $B \cap A_1, \ldots, B \cap A_n$ are pairwise disjoint as well.
(b) The probability measure P is additive [see Eq. (1.15)].

1.2.4 Bayes' Theorem

The next theorem, called *Bayes' theorem*, reveals how the A_i-conditional probabilities $P(B\,|\,A_i)$ are related to the B-conditional probabilities $P(A_i|B)$. Using the definitions of the conditional probabilities $P(A_i|B)$ and $P(B\,|\,A_i)$ yields

$$P(A_i|B) \;=\; \frac{P(B\,|\,A_i)\cdot P(A_i)}{P(B)}. \tag{1.17}$$

Inserting Equation (1.16) for $P(B)$ yields the following theorem.

Theorem 1.30 [Bayes' Theorem]
Let $(\Omega, \mathscr{A}, P)$ be a probability space, $B \in \mathscr{A}$, and $P(B) > 0$. Under the assumptions of Theorem 1.29 and assuming $P(A_i) > 0$, for all $i = 1, \ldots, n$,

$$P(A_i|B) \;=\; \frac{P(B\,|\,A_i)\cdot P(A_i)}{\sum_{j=1}^{n} P(B\,|\,A_j)\cdot P(A_j)}, \quad \forall\; i = 1, \ldots, n. \tag{1.18}$$

Example 1.31 [Infection and a Test of Being Infected] Consider the following random experiment: A person u is sampled from large but finite set Ω_U of persons. The sampled person is either infected by a specified virus or not, and the same person is tested for the virus, the test being either positive (+) or negative (−). Hence, the set of possible outcomes is

$$\Omega \;=\; \Omega_U \times \Omega_T,$$

where $\Omega_T = \{+, -\}$. If Ω_I denotes the *subset of all infected persons* and Ω_N the *subset of all persons that are not infected*, both at the time of sampling the person, then $\Omega_U = \Omega_I \cup \Omega_N$.

As the set $\mathscr{A}$ of possible events we consider the power set of Ω. Furthermore, we consider the event

$$A = \Omega_I \times \Omega_T$$

that *the sampled person is infected*, and the event

$$B = \Omega_U \times \{+\}$$

that the *test is positive*. We focus on the crucial question: *What is the probability $P(A|B)$ that the person is infected given that his or her test is positive?* Note that it is this probability that informs us about the diagnostic quality of the test. We obtain the answer applying Equation (1.18), which, in this example, simplifies to

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)}, \tag{1.19}$$

where A^c denotes the event that *the person is not infected*. Equation 1.19 shows that the crucial probability $P(A|B)$ that the person is infected given that his or her test is positive depends on $P(A)$, the so-called *incidence*. Of course, the incidence depends on the phase of the pandemic in which the test is made. In order to illustrate this point, we consider three random experiments that differ from each other only by their probability measures P_1, P_2 and P_3. We assume that these three probability measures differ only by their incidences $P_1(A)$, $P_2(A)$, and $P_3(A)$, and by their implications.

First Random Experiment. In the first random experiment we assume that the test is taken at a phase of the pandemic in which the incidence is $P_1(A) = .01$. Furthermore, we assume that each person $u \in \Omega_U$ has the same probability of being sampled and we know that $P_1(B|A) = .98$. In the context of this example, this conditional probability is called the *sensitivity* of the test. Suppose that we also know $P_1(B^c|A^c) = .99$, called the *specificity* in this context. This is the probability that the sampled person is negatively tested given that he or she is not infected.

Because A^c is the complement of A, we obtain $P_1(A^c) = 1 - P_1(A) = .99$. Furthermore, for the analog reason, $P_1(B|A^c) = 1 - P_1(B^c|A^c) = 1 - .99 = .01$. Hence, Equation (1.19) yields

$$P_1(A|B) = \frac{.98 \cdot .01}{.98 \cdot .01 + .01 \cdot .99} = .4974619$$

and

$$P_1(A^c|B) = 1 - .4974619 = .5025381.$$

Furthermore, Equation (1.16) yields

$$P_1(B) = P_1(B|A) \cdot P_1(A) + P_1(B|A^c) \cdot P_1(A^c) = .98 \cdot .01 + .01 \cdot .99 = .0197,$$

which is the probability that a test is positive. Therefore, if $1{,}500{,}000$ persons are tested *in this weak of the first phase of the pandemic*, then we expect that

$$P_1(B) \cdot 1{,}500{,}000 = .0197 \cdot 1{,}500{,}000 = 29{,}550$$

are positive. The number of positive tests in a weak per $100{,}000$ inhabitants in a country (like Germany on April 26, 2021) with a population of $83{,}720{,}586$ is

$$\frac{29{,}550}{83{,}720{,}586} \cdot 100{,}000 \approx 35.$$

Finally, we expect that

$$P_1(A^c|B)\cdot 29,550 \;=\; .5025381\cdot 29,550 \;\approx\; 14,850$$

are not infected, although their tests are positive.

Second Random Experiment. Now we consider a second random experiment that differs from the previous one only by the phase of the pandemic, in which we test for the virus. This time we assume that the incidence is $P_2(A) = .1$, whereas sensitivity $P_2(B\,|\,A) = .98$ and specificity $P_2(B^c|A^c) = .99$ are the same as in the first random experiment. Then $P_2(A^c) = 1 - P_2(A) = .9$, whereas all other probabilities in Equation (1.19) remain unchanged. Now this equation yields

$$P_2(A\,|\,B) \;=\; \frac{.98\cdot .1}{.98\cdot .1 \,+\, .01\cdot .9} \;\approx\; .9158879$$

and

$$P_2(A^c|B) \;\approx\; 1 - .9158879 \;\approx\; .08411215.$$

Furthermore, in this example, Equation (1.16) yields

$$P_2(B) \;=\; P_2(B\,|\,A)\cdot P_2(A) \,+\, P_2(B\,|\,A^c)\cdot P_2(A^c) \;=\; .98\cdot .1 \,+\, .01\cdot .9 \;=\; .107.$$

This is the probability, that *in this phase of the pandemic*, the test is positive. Hence, if 1,500,000 persons are tested in this second phase of the pandemic, then we expect that

$$P_2(B)\cdot 1,500,000 \;=\; .107\cdot 1,500,000 \;=\; 160,500$$

are positive. The number of positive tests in a weak per 100,000 inhabitants in a country with a population of 83,720,586 is then

$$\frac{160,500}{83,720,586}\cdot 100,000 \;\approx\; 192.$$

Finally, we expect that

$$P_2(A^c|B)\cdot 160,500 \;=\; .08411215\cdot 160,500 \;=\; 13,500$$

are not infected, although they are positive.

Third Random Experiment. Finally, we consider a third random experiment. Again, this random experiment differs from the previous ones only by the phase of the pandemic, in which we test for the virus. This time we assume that absolutely nobody is infected anymore. This means that the incidence is $P_3(A) = 0$. This implies that $P_3(A^c) = 1 - P_3(A) = 1$ and sensitivity is not defined any more. We still assume that specificity is still $P_3(B^c|A^c) = P_3(B^c) = .99$. Hence, $P_3(B\,|\,A^c) = P_3(B) = 1 - .99 = .01$. Applying Equation (1.10) yields

$$P_3(A\,|\,B) \;=\; \frac{P_3(A\cap B)}{P_3(B)} \;=\; \frac{0}{.01} \;=\; 0. \tag{1.20}$$

Hence, if 1,500,000 persons per week are tested *in this third phase of the pandemic*, then we expect that $P_3(B)\cdot 1,500,000 = .01\cdot 1,500,000 = 15,000$ tests are positive although not a single person is infected. The number of positive tests in a weak per 100,000 inhabitants in a country (like Germany) with a population of 83,720,586 is

$$\frac{15,000}{83,720,586}\cdot 100,000 \;\approx\; 18.$$

Substantive Implications. These examples already show that the diagnostic quality of testing for the virus and the probability $P(B)$ of a test being positive heavily depend on the indicence $P(A)$ at the time of testing. Obviously, these numbers and their implications for quarantining and other policies for stemming the pancemic should be kept in mind when deciding about these policies.

It should be noted that these numbers also have implications for the interpretion of the counts of persons that 'died with virus'. If we assume that a person is counted as 'died with virus' if he or she has a positive test, then, in the scenario (i. e., in the random experiment) with an incidence of $P_1(A) = .01$, the probability $P_1(A|B) = .4974619$ means that about 49.7% of those who 'died with virus' were infected, whereas about 50.3% of them were not infected and therefore could not have died because of the virus. In contrast, in the scenario with an incidence of $P_2(A) = .1$, the probability $P_2(A|B) = .9158879$ means that about 91.6% percent of the counts of persons who 'died with virus' were actually infected, whereas about 8.4% of them were not. Finally, in the scenario with incidence $P_3(A) = 0$, the probability $P_3(A|B) = 0$ means that not a single person who 'died with virus' was infected or died *because of the virus*. ◁

Example 1.32 [Joe and Ann With Randomized Assignment] Let

$$A = \{(Joe, no, -), (Joe, no, +), (Joe, yes, -), (Joe, yes, +)\}$$

denote the event that *Joe is drawn*,

$$A^c = \{(Ann, no, -), (Ann, no, +), (Ann, yes, -), (Ann, yes, +)\}$$

the event that *Ann is drawn*, and

$$B = \{(Joe, yes, -), (Joe, yes, +), (Ann, yes, -), (Ann, yes, +)\}$$

the event that the *drawn person is treated*. Then

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)} = \frac{.4 \cdot .5}{.4 \cdot .5 + .4 \cdot .5} = .5$$

is the conditional probability that *Joe is drawn* given that the *drawn person is treated* (see Table 1.2). The corresponding probability that *Ann is drawn* given that the *drawn person is treated* is identical in this example, that is, $P(A^c|B) = .5$. Hence, given treatment, each person has the same probability to be drawn. Asking for the probabilities of sampling a person given a treatment condition is what we refer to as the *sampling* of a randomized experiment. This sampling perspective supplements the *assignment perspective*, from which we ask for the treatment probability given a person, for example, $P(B|A) = P(B|A^c) = .4$ (see again Table 1.2). ◁

1.2.5 Conditional-Probability Measure

Just like (unconditional) probabilities, conditional probabilities of events $A \in \mathscr{A}$ given an event $B \in \mathscr{A}$ are values of a probability measure. This is stated in the following theorem. (For a proof see SN-Th. 4.28.)

Theorem 1.33 [Probability Measure P^B on $(\Omega, \mathscr{A})$]
Let $(\Omega, \mathscr{A}, P)$ be a probability space, let $B \in \mathscr{A}$, and assume $P(B) > 0$. Then the function $P^B\colon \mathscr{A} \to [0, 1]$ defined by

$$P^B(A) = P(A \,|\, B), \quad \forall\, A \in \mathscr{A}, \tag{1.21}$$

is a probability measure on $(\Omega, \mathscr{A})$.

According to this theorem, for each event $B \in \mathscr{A}$ with $P(B) > 0$, the function P^B is a probability measure on $(\Omega, \mathscr{A})$. Hence, according to Remark 1.19, the triple $(\Omega, \mathscr{A}, P^B)$ is a probability space.

The function P^B assigns to each event $A \in \mathscr{A}$ its conditional probability given B. Therefore, we also call it the B-*conditional-probability measure* on $(\Omega, \mathscr{A})$. Of course, if B and C are different events, the conditional probabilities $P^B(A) = P(A \,|\, B)$ and $P^C(A) = P(A \,|\, C)$ can differ from each other.

Remark 1.34 [The Special Case $P(B) = 1$] According to Box 1.1 (viii),

$$P(B) = 1 \quad \Rightarrow \quad P(A \cap B) = P(A). \tag{1.22}$$

Hence,

$$P(B) = 1 \quad \Rightarrow \quad P(A \,|\, B) = \frac{P(A \cap B)}{P(B)} = P(A) \tag{1.23}$$

(see Def. 1.23), and

$$P(B) = 1 \quad \Rightarrow \quad P^B = P \tag{1.24}$$

[see Eqs. (1.21), (1.23), and the fact that P^B and P are functions on $\mathscr{A}$]. ◁

Remark 1.35 [Null Sets With Respect to P^B] Also note that

$$P^B(A) = 0, \quad \forall\, A \in \mathscr{A} \text{ with } A \cap B = \emptyset \tag{1.25}$$

holds for the measure P^B defined by Equation (1.21). ◁

Example 1.36 [Joe and Ann With Randomized Assignment] Consider the example presented in Table 1.2. We specify the B-conditional-probability measure $P^B\colon \mathscr{A} \to [0, 1]$ for the event

$$B = \big\{(Joe, yes, -),\ (Joe, yes, +),\ (Ann, yes, -),\ (Ann, yes, +)\big\}$$

that the *drawn person is treated*. Using the probabilities of the elementary events displayed in the second column of Table 1.2 and Box 1.1 (ii), we compute

$$\begin{aligned} P(B) &= P\big(\big\{(Joe, yes, -),\ (Joe, yes, +),\ (Ann, yes, -),\ (Ann, yes, +)\big\}\big) \\ &= P\big(\{(Joe, yes, -)\}\big) + P\big(\{(Joe, yes, +)\}\big) + P\big(\{(Ann, yes, -)\}\big) + P\big(\{(Ann, yes, +)\}\big) \\ &= .04 + .16 + .12 + .08 = .4. \end{aligned}$$

For the first two elementary events,

$$P^B(\{\omega_1\}) \;=\; P^B(\{(\mathit{Joe}, \mathit{no}, +)\}) \;=\; P^B(\{\omega_2\}) \;=\; P^B(\{(\mathit{Joe}, \mathit{no}, -)\}) \;=\; 0,$$

because the intersections $\{(\mathit{Joe}, \mathit{no}, -)\} \cap B$ and $\{(\mathit{Joe}, \mathit{no}, +)\} \cap B$ are the empty set. For the next two elementary events, the B-conditional probabilities are

$$P^B(\{\omega_3\}) \;=\; P^B\big(\{(\mathit{Joe}, \mathit{yes}, -)\}\big) \;=\; \frac{P\big(\{(\mathit{Joe}, \mathit{yes}, -)\} \cap B\big)}{P(B)} \;=\; \frac{.04}{.4} \;=\; .1$$

and

$$P^B(\{\omega_4\}) \;=\; P^B\big(\{(\mathit{Joe}, \mathit{yes}, +)\}\big) \;=\; \frac{P\big(\{(\mathit{Joe}, \mathit{yes}, +)\} \cap B\big)}{P(B)} \;=\; \frac{.16}{.4} \;=\; .4.$$

Next,

$$P^B(\{\omega_5\}) \;=\; P^B\big(\{(\mathit{Ann}, \mathit{no}, -)\}\big) \;=\; P^B(\{\omega_6\}) \;=\; P^B\big(\{(\mathit{Ann}, \mathit{no}, +)\}\big) \;=\; 0,$$

because again $\{(\mathit{Ann}, \mathit{no}, -)\} \cap B = \{(\mathit{Ann}, \mathit{no}, +)\} \cap B = \emptyset$. Finally, for the last two elementary events, the B-conditional probabilities are

$$P^B(\{\omega_7\}) \;=\; P^B\big(\{(\mathit{Ann}, \mathit{yes}, -)\}\big) \;=\; \frac{P\big(\{(\mathit{Ann}, \mathit{yes}, -)\} \cap B\big)}{P(B)} \;=\; \frac{.12}{.4} \;=\; .3$$

and

$$P^B(\{\omega_8\}) \;=\; P^B\big(\{(\mathit{Ann}, \mathit{yes}, +)\}\big) \;=\; \frac{P\big(\{(\mathit{Ann}, \mathit{yes}, +)\} \cap B\big)}{P(B)} \;=\; \frac{.08}{.4} \;=\; .2.$$

These eight probabilities are summarized in the fourth column of Table 1.2.

Except for $\emptyset$, all other events $A \in \mathscr{A}$ are unions of these elementary events. Because the elementary events are *disjoint*, the probabilities of their unions can easily be computed using finite additivity of a probability measure [see Box 1.1 (ii)]. For example, the B-conditional probability of the event

$$C \;=\; \big\{(\mathit{Joe}, \mathit{no}, +),\, (\mathit{Joe}, \mathit{yes}, +),\, (\mathit{Ann}, \mathit{no}, +),\, (\mathit{Ann}, \mathit{yes}, +)\big\}$$

that the sampled person has success is

$$\begin{aligned} P^B(C) &= P^B\big(\big\{(\mathit{Joe}, \mathit{no}, +),\, (\mathit{Joe}, \mathit{yes}, +),\, (\mathit{Ann}, \mathit{no}, +),\, (\mathit{Ann}, \mathit{yes}, +)\big\}\big) \\ &= P^B\big(\{(\mathit{Joe}, \mathit{no}, +)\}\big) + P^B\big(\{(\mathit{Joe}, \mathit{yes}, +)\}\big) + P^B\big(\{(\mathit{Ann}, \mathit{no}, +)\}\big) + P^B\big(\{(\mathit{Ann}, \mathit{yes}, +)\}\big) \\ &= 0 + .4 + 0 + .2 \;=\; .6. \end{aligned}$$

◁

Example 1.37 [Nonorthogonal Factors] Consider again the example presented in Table 1.3. For the event

$$A \;=\; \{\mathit{Tom}, \mathit{Tim}\} \times \Omega_X \times \Omega_Y$$

that the sampled person has status *low*, we specify its B-conditional probability, where

$$B \;=\; \Omega_U \times \{\mathit{treatment}\,1\} \times \Omega_Y$$

is the event that the *drawn person (whoever it is) receives treatment* 1. According to the second row of the last column of Table 1.5, $P(B) = 1/3$. Then, using the probabilities displayed in Table 1.3,

Table 1.5. Conditional expectation values $E(Y|X=x, Z=z)$ given treatment and status

Treatment	Status						
	Low ($Z=low$)		Medium ($Z=med$)		High ($Z=hi$)		
$X=0$	120	(20/120)	110	(17/120)	60	(3/120)	(40/120)
$X=1$	100	(7/120)	100	(26/120)	100	(7/120)	(40/120)
$X=2$	80	(3/120)	90	(17/120)	140	(20/120)	(40/120)
		(30/120)		(60/120)		(30/120)	

Note. Probabilities $P(X=x, Z=z)$, $P(Z=z)$, and $P(X=x)$ in parentheses.

$$P^B(A) = \frac{P\big(\{Tom, Tim\} \times \Omega_X \times \Omega_Y \cap \Omega_U \times \{treatment\,1\} \times \Omega_Y\big)}{P(B)} \qquad [(1.10), (1.21)]$$

$$= \frac{(10/60 + 18/60) \cdot 1/3}{1/3} = \frac{28}{60}. \qquad [P(A \cap B) = P(A|B) \cdot P(B)]$$

This result is consistent with the probabilities displayed in Table 1.5. Again note the distinction between the two measures P and P^B. While $P^B(A) = 28/60$, the probability of A with respect to the measure P is $P(A) = 2/8$ (see again Table 1.3.) ◁

In the following lemma we consider the relationship between conditional probabilities with respect to the measures P^B and P. (For a proof see SN-Lemma 4.30.)

Lemma 1.38 [Conditional Probabilities With Respect to P^B]
Let $(\Omega, \mathscr{A}, P)$ be a probability space. If $A, B, C \in \mathscr{A}$ and $P(B \cap C) > 0$, then

$$P^B(A|C) = P(A|B \cap C). \tag{1.26}$$

According to this lemma, the C-conditional probability of the event A with respect to the B-conditional probability measure P^B is identical to the $(B \cap C)$-conditional probability of A with respect to the probability measure P.

1.3 Independence of Events

From an intuitive point of view, independence of two events A and B should be defined such that the conditional and unconditional probabilities are the same, that is, $P(A|B) = P(A)$ and $P(B|A) = P(B)$. However, this intuitive idea presupposes that $P(A), P(B) > 0$, because otherwise the two conditional probabilities are not defined. The following definition does not rest on this requirement, although it incorporates this intuitive idea. The definition also extends the concept of independence of two events to a family of events. That is, the index set I occuring in Definition 1.39 (ii) may be finite, countable, or uncountable.

Box 1.2 Independence of Events

Let $(\Omega, \mathscr{A}, P)$ be a probability space, $A, B, C \in \mathscr{A}$, and $A^c = \Omega \setminus A$ the complement of A. Then:

$$A \perp\!\!\!\perp B \;:\Leftrightarrow\; P(A \cap B) = P(A) \cdot P(B) \tag{i}$$

$$A \perp\!\!\!\perp B \;\Leftrightarrow\; A^c \perp\!\!\!\perp B \tag{ii}$$

$$A \perp\!\!\!\perp B \;\Leftrightarrow\; \sigma(\{A\}) \perp\!\!\!\perp \sigma(\{B\}) \tag{iii}$$

$$A \perp\!\!\!\perp B \perp\!\!\!\perp C \;:\Leftrightarrow\; \begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(A \cap C) &= P(A) \cdot P(C) \\ P(B \cap C) &= P(B) \cdot P(C) \\ P(A \cap B \cap C) &= P(A) \cdot P(B) \cdot P(C) \end{aligned} \tag{iv}$$

$$A \perp\!\!\!\perp B \perp\!\!\!\perp C \;\Rightarrow\; A \perp\!\!\!\perp B,\; A \perp\!\!\!\perp C,\; B \perp\!\!\!\perp C. \tag{v}$$

If $P(B) > 0$, then

$$A \perp\!\!\!\perp B \;\Leftrightarrow\; P(A \mid B) = P(A). \tag{vi}$$

If $P(B), P(B^c) > 0$, then

$$A \perp\!\!\!\perp B \;\Leftrightarrow\; P(A \mid B) = P(A \mid B^c). \tag{vii}$$

If $P(B \cap C), P(B \cap C^c) > 0$, then

$$B \perp\!\!\!\perp C \;\Rightarrow\; P(A \mid B) = P(A \mid B \cap C) \cdot P(C) + P(A \mid B \cap C^c) \cdot P(C^c). \tag{viii}$$

Definition 1.39 [Independence of Events]
Let $(\Omega, \mathscr{A}, P)$ be a probability space.

(i) *We call two events $A, B \in \mathscr{A}$ independent (with respect to P), denoted $A \perp\!\!\!\perp B$, if*

$$P(A \cap B) = P(A) \cdot P(B). \tag{1.27}$$

(ii) *Let I be a nonempty set and let $A_i \in \mathscr{A}$, for all $i \in I$. Then $(A_i, i \in I)$ is called a family of independent events, denoted $\perp\!\!\!\perp (A_i, i \in I)$, if*

$$P\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} P(A_i), \quad \forall \text{ finite nonempty } J \subset I. \tag{1.28}$$

Remark 1.40 [Pairwise and Triple-Wise Independence] For n events $A_1, \dots, A_n$, independence will also be denoted by

$$A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_n.$$

In this case $I = \{1, \dots, n\}$. For three events, for instance, $I = \{1, \dots, 3\}$ and $A_1 \perp\!\!\!\perp A_2 \perp\!\!\!\perp A_3$ means that

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j), \quad i \neq j, \quad i, j = 1, 2, 3, \tag{1.29}$$

(*pairwise independence*) and

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) \tag{1.30}$$

(*triplewise independence*) hold. Equation (1.29) results from Equation (1.28) if we consider the subsets $J_1 = \{1,2\}$, $J_2 = \{1,3\}$, and $J_3 = \{2,3\}$ of I, and Equation (1.30) results from Equation (1.28) if we consider the subset $J_4 = \{1,2,3\}$ of I. For subsets $J_5 = \{1\}$, $J_6 = \{2\}$, $J_7 = \{3\}$, Equation (1.28) yields the trivial equations $P(A_1) = P(A_1)$, $P(A_2) = P(A_2)$, and $P(A_3) = P(A_3)$, respectively.

Note that pairwise independence of three events [i. e., Eq. (1.29)] does not imply their triplewise independence [i.e., Eq. (1.30)]. Furthermore, triplewise independence [i. e., Eq. (1.30)] does not imply pairwise independence. ◁

Remark 1.41 [Conditional Probability and Independent Events] If A and B are independent and $P(B) > 0$, then $P(A \mid B) = P(A)$. Vice versa, if $P(B) > 0$ and $P(A \mid B) = P(A)$, then A and B are independent. For more propositions on independence of events see Box 1.2, which has been adapted from SN-Box 4.2. (For proofs see SN-Exercise 4-9.) ◁

Remark 1.42 [Independence of Two Events With Respect to P^B] Let $(\Omega, \mathscr{A}, P)$ be a probability space, $A, B, C \in \mathscr{A}$, and $P(B) > 0$. Then the events A and C are called *independent with respect to the probability measure P^B*, denoted $A \underset{P^B}{\perp\!\!\!\perp} C$ or $A \perp\!\!\!\perp_{P^B} C$, if

$$P^B(A \cap C) = P^B(A) \cdot P^B(C). \tag{1.31}$$

Hence all properties of independence with respect to the measure P also apply to independence with respect to the measure P^B. We simply have to replace P by P^B. ◁

1.4 Independence of Set Systems

Building on the concept of independence of a family of events [see Def. 1.39 (ii)], now we extend the concept of independence to *set systems*, that is, to sets of events.

Definition 1.43 [Family of Independent Set Systems]
Let $(\Omega, \mathscr{A}, P)$ be a probability space and assume $\mathscr{E}_i \subset \mathscr{A}$, for all $i \in I \neq \emptyset$. Then $(\mathscr{E}_i, i \in I)$ is called a family of independent set systems, denoted $\perp\!\!\!\perp(\mathscr{E}_i, i \in I)$, if $\perp\!\!\!\perp(A_i, i \in I)$ holds for all families $(A_i, i \in I)$ with $A_i \in \mathscr{E}_i$, for all $i \in I$. If $I = \{1, \dots, n\}$, then we also use the notation $\mathscr{E}_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp \mathscr{E}_n$ instead of $\perp\!\!\!\perp(\mathscr{E}_i, i \in I)$.

Remark 1.44 [Independence of an Event and a Set System] Let $(\Omega, \mathscr{A}, P)$ be a probability space. An event $A \in \mathscr{A}$ and a set system $\mathscr{E} \subset \mathscr{A}$ are called *independent*, denoted $A \perp\!\!\!\perp \mathscr{E}$, if $\{A\} \perp\!\!\!\perp \mathscr{E}$. ◁

Remark 1.45 [Independence of σ-Algebras] Note that σ-algebras are special set systems referred to in Definition 1.43. Hence, a family $(\mathscr{A}_i, i \in I)$ of sub-σ-algebras of $\mathscr{A}$ can be independent as well. This fact will be used introducing the concept of *independence of random variables* (see sect. 2.4). ◁

Example 1.46 [Joe and Ann With Randomized Assignment] Consider again Table 1.2, define

$$A \;=\; \{Joe\} \times \Omega_X \times \Omega_Y \;=\; \big\{(Joe, no, -),\, (Joe, no, +)\,,\, (Joe, yes, -),\, (Joe, yes, +)\big\},$$

the event that the *Joe is sampled*, and

$$B \;=\; \Omega_U \times \{yes\} \times \Omega_Y \;=\; \big\{(Joe, yes, -),\, (Joe, yes, +)\,,\, (Ann, yes, -),\, (Ann, yes, +)\big\},$$

the event that the *sampled person (whoever it is) is treated.* Then A and B are independent. This is easily seen, because $P(A) = .5$ and $P(B) = .4$ (see Exercise 1-6). The product of these two probabilities is $P(A) \cdot P(B) = .5 \cdot .4 = .2$, and this is equal to

$$\begin{aligned} P(A \cap B) \;&=\; P\Big(\big\{(Joe, yes, -),\, (Joe, yes, +)\big\}\Big) \\ &= P\big(\{(Joe, yes, -)\}\big) + P\big(\{(Joe, yes, +)\}\big) \\ &= .04 + .16 \;=\; .2. \end{aligned}$$

Hence, in this example, the events A and B are independent [see Def. 1.39 (i)]. This implies that the two σ-algebras $\sigma(\{A\}) = \{A, A^c, \Omega, \emptyset\}$ and $\sigma(\{B\}) = \{B, B^c, \Omega, \emptyset\}$ are independent as well [see Box 1.2 (iii)]. In contrast, in the random experiment displayed in Table 1.4 (Joe and Ann with self-selection), the corresponding events and sets of events are not independent (see Exercise 1-7). ◁

1.5 Conditional Independence of Events Given an Event

Now we extend the concept of independence of events and of sets of events introducing *conditional independence* of events and of *sets of events given an event*, using the probability measure P^B [see Eq. (1.21)].

Definition 1.47 [Conditional Independence of Two Events Given an Event]
Let $(\Omega, \mathscr{A}, P)$ be a probability space, $A, B, C \in \mathscr{A}$, and $P(B) > 0$. Then the events A and C are called B-conditionally independent, denoted $A \perp\!\!\!\perp C \,|\, B$, if

$$P(A \cap C \,|\, B) \;=\; P(A \,|\, B) \cdot P(C \,|\, B). \tag{1.32}$$

Remark 1.48 [Independence of Two Events With Respect to P^B] The definitions of $A \underset{P^B}{\perp\!\!\!\perp} C$ and $A \perp\!\!\!\perp C \,|\, B$ [see Eqs. (1.21) and (1.32)] immediately yield

$$A \perp\!\!\!\perp C \,|\, B \quad \Leftrightarrow \quad A \underset{P^B}{\perp\!\!\!\perp} C. \tag{1.33}$$

◁

Remark 1.49 [A Condition Equivalent to Conditional Independence] If we assume that $P(B \cap C) > 0$, then Equation (1.32) is equivalent to

$$P(A \mid B \cap C) = P(A \mid B) \tag{1.34}$$

[see Box 1.3 (v)]. Exchanging A and C immediately yields: If $P(A \cap B) > 0$, then Equation (1.32) is equivalent to

$$P(C \mid A \cap B) = P(C \mid B). \tag{1.35}$$

◁

Remark 1.50 [Independence and Conditional Independence] Assume that $B \in \mathscr{A}$ and $P(B) > 0$. Then independence of A and C neither implies nor is implied by B-conditional independence of A and C (see SN-Exercise 4-8). However, independence of A, B, and C *does* imply B-conditional independence of A and C [see Box 1.3 (iii)]. For more propositions on conditional independence of events see Box 1.3 (p. 24). (For proofs see SN-Exercise 4-9). ◁

Example 1.51 [Nonorthogonal Factors] In the example presented in Table 1.3, consider the events

$$A = \Omega_U \times \{\textit{treatment}\,1\} \times \Omega_Y$$

that the sampled person receives treatment 1,

$$B = \{\textit{Joe}, \textit{Jim}, \textit{Ann}, \textit{Eva}\} \times \Omega_X \times \Omega_Y$$

that Joe, Jim, Ann, or Eva is sampled, and

$$C = \{\textit{Ann}, \textit{Eva}, \textit{Sue}, \textit{Mia}\} \times \Omega_X \times \Omega_Y$$

that the sampled person is female. For these events, Equation (1.32) holds, because

$$\begin{aligned} P^B(A \cap C) &= P(A \cap C \mid B) = \frac{P(A \cap B \cap C)}{P(B)} = \frac{P(A \mid B \cap C) \cdot P(B \cap C)}{P(B)} \\ &= \frac{26/60 \cdot 2/8}{1/2} = \frac{13}{60} \end{aligned}$$

and

$$P^B(A) \cdot P^B(C) = P(A \mid B) \cdot P(C \mid B) = \frac{26}{60} \cdot \frac{1}{2} = \frac{13}{60}$$

(see Table 1.3). ◁

If we consider B-conditional independence of more than two events, then this leads us to the concept of a family of B-conditionally independent events.

Definition 1.52 [Family of B-Conditionally Independent Events]
Let $(\Omega, \mathscr{A}, P)$ be a probability space, $B \in \mathscr{A}$, $P(B) > 0$, and let P^B be defined by Equation (1.21). Furthermore, let $I \neq \emptyset$, and $A_i \in \mathscr{A}$, for all $i \in I$. Then we call $(A_i, i \in I)$ a family of B-conditionally independent events, denoted $\underset{P^B}{\perp\!\!\!\perp}(A_i, i \in I)$, if

$$P^B\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} P^B(A_i), \quad \forall \textit{ finite nonempty } J \subset I. \tag{1.36}$$

Box 1.3 Conditional Independence of Events

Let $(\Omega, \mathscr{A}, P)$ be a probability space and $A, B, C \in \mathscr{A}$.

If $P(B) > 0$, then

$$A \perp\!\!\!\perp C \mid B \;:\Leftrightarrow\; P^B(A \cap C) = P^B(A) \cdot P^B(C) \;\Leftrightarrow\; A \underset{P^B}{\perp\!\!\!\perp} C \tag{i}$$

$$A \perp\!\!\!\perp C \mid B \;\Leftrightarrow\; A \perp\!\!\!\perp C^c \mid B \tag{ii}$$

$$A \perp\!\!\!\perp B \perp\!\!\!\perp C \;\Rightarrow\; A \perp\!\!\!\perp C \mid B. \tag{iii}$$

If $P(B), P(B^c) > 0$, then

$$A \perp\!\!\!\perp B \;\Leftrightarrow\; P(A \mid B) = P(A \mid B^c). \tag{iv}$$

If $P(B \cap C) > 0$, then

$$A \perp\!\!\!\perp C \mid B \;\Leftrightarrow\; P(A \mid B \cap C) = P(A \mid B). \tag{v}$$

If $P(B \cap C^c) > 0$, then

$$A \perp\!\!\!\perp C \mid B \;\Leftrightarrow\; P(A \mid B \cap C^c) = P(A \mid B). \tag{vi}$$

If $P(B \cap C), P(B \cap C^c) > 0$, then

$$A \perp\!\!\!\perp C \mid B \;\Leftrightarrow\; P(A \mid B \cap C) = P(A \mid B \cap C^c) \tag{vii}$$

$$A \perp\!\!\!\perp C \mid B \;\Rightarrow\; P(A \mid B) = P(A \mid B \cap C) \cdot P(C) + P(A \mid B \cap C^c) \cdot P(C^c). \tag{viii}$$

1.6 Conditional Independence of Set Systems Given an Event

Now we extend the concept of conditional independence given an event to *set systems*. In Remark 1.48 we already noted that B-conditional independence of two events A and C is equivalent to independence of A and C with respect to the measure P^B. Correspondingly, B-conditional independence of a family $(\mathscr{E}_i, i \in I)$ of events is defined as independence of $(\mathscr{E}_i, i \in I)$ with respect to P^B.

Definition 1.53 [Family of B-Conditionally Independent Set Systems]
Let $(\Omega, \mathscr{A}, P)$ be a probability space, $B \in \mathscr{A}$, $P(B) > 0$, and let P^B be defined by Equation (1.21). Furthermore, let I be a nonempty set, and $\mathscr{E}_i \subset \mathscr{A}$, for all $i \in I$. Then we call $(\mathscr{E}_i, i \in I)$ a family of B-conditionally independent set systems, denoted $\perp\!\!\!\perp_{P^B}(\mathscr{E}_i, i \in I)$, if $\perp\!\!\!\perp_{P^B}(A_i, i \in I)$ holds for all families $(A_i, i \in I)$ with $A_i \in \mathscr{E}_i$, for all $i \in I$. If $I = \{1, \dots, m\}$, then we also use the notation $\mathscr{E}_1 \perp\!\!\!\perp_{P^B} \dots \perp\!\!\!\perp_{P^B} \mathscr{E}_m$ instead of $\perp\!\!\!\perp_{P^B}(\mathscr{E}_i, i \in I)$.

Remark 1.54 [Conditional Independence of σ-Algebras] Again, σ-algebras can be such set systems referred to in Definition 1.53. Hence, a family $(\mathscr{A}_i, i \in I)$ of sub-σ-algebras of $\mathscr{A}$ can be B-conditionally independent as well. ◁

Box 1.4 Glossary of new concepts

Ω	*Set of possible outcomes*, also called *sample space*.
$\mathscr{A}$	*Set of possible events*. A set of subsets of Ω satisfying the requirements of a σ-*algebra* on Ω.
P	*Probability measure on* $\mathscr{A}$. A function on $\mathscr{A}$ with values in the closed interval $[0,1]$ satisfying the Kolmogorov axioms of probability.
$(\Omega, \mathscr{A}, P)$	*Probability space*. It consists of the three components listed above. In empirical applications it contains all information about the random experiment considered. All concepts listed below refer to such a probability space.
$P(A \mid B)$	*Conditional probability of the event A given the event B*. If $A, B \in \mathscr{A}$ and $P(B) > 0$, then $P(A \mid B) := P(A \cap B)/P(B)$.
P^B	*B-conditional-probability measure on $\mathscr{A}$*. If $B \in \mathscr{A}$ and $P(B) > 0$, then $P^B(A) := P(A \mid B)$, for all $A \in \mathscr{A}$.
$P^B(A)$	The probability of the event A with respect to the measure P^B.
$A \perp\!\!\!\perp B$	Independence of the events $A, B \in \mathscr{A}$ with respect to the measure P. It is defined by $P(A \cap B) = P(A) \cdot P(B)$.
$\perp\!\!\!\perp (A_i, i \in I)$	*Family of independent events* $(A_i, i \in I)$. Assuming $A_i \in \mathscr{A}$ for all $i \in I$, where $I \neq \emptyset$, it is defined by $$P\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} P(A_i), \quad \forall \text{ finite nonempty } J \subset I.$$
$A_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp A_m$	*Independence of the events* $A_1, \ldots, A_m$. It is defined by $\perp\!\!\!\perp (A_i, i \in I)$ for $I = \{1, \ldots, m\}$.
$\perp\!\!\!\perp (\mathscr{E}_i, i \in I)$	*Family of independent set systems* $(\mathscr{E}_i, i \in I)$. Assuming $\mathscr{E}_i \in \mathscr{A}$ for all $i \in I$, $I \neq \emptyset$, it is defined by independence of all families $(A_i, i \in I)$ with $A_i \in \mathscr{E}_i$, $i \in I$, where $I \neq \emptyset$.
$\mathscr{E}_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp \mathscr{E}_m$	*Independence of the set systems* $\mathscr{E}_1, \ldots, \mathscr{E}_m$. It is defined by $\perp\!\!\!\perp (\mathscr{E}_i, i \in I)$ for $I = \{1, \ldots, m\}$.
$A \perp\!\!\!\perp C \mid B$	*B-conditional independence of the events $A, B \in \mathscr{A}$*. Presuming $P(B) > 0$, it is defined by $P(A \cap C \mid B) = P(A \mid B) \cdot P(C \mid B)$.
$\perp\!\!\!\perp_{P^B} (A_i, i \in I)$	*B-conditional independence of the family $(A_i, i \in I)$ of events*. Assuming $A_i \in \mathscr{A}$ for all $i \in I$, it is defined by independence of the family $(A_i, i \in I)$ with respect to the measure P^B.
$\perp\!\!\!\perp_{P^B} (\mathscr{E}_i, i \in I)$	*B-conditional independence of the set system $(\mathscr{E}_i, i \in I)$*. It is defined by independence of the set system $(\mathscr{E}_i, i \in I)$ with respect to the measure P^B.

1.7 Summary and Conclusions

In this chapter we reviewed some basic concepts of probability theory. Box 1.4 provides a glossary. The emphasis has been on the *structure* or *architecture* of these concepts. We started with the three components of a *probability space* $(\Omega, \mathscr{A}, P)$, the *set Ω of possible outcomes*, the σ-*algebra* $\mathscr{A}$ of *possible events*, which is a set of subsets of Ω with certain properties, and the *probability measure* P, a function on $\mathscr{A}$ with values in the closed interval $[0,1]$, satisfying Kolmogorovs axioms of probability. All other concepts listed in Box 1.4 refer to such a probability space. In an empirical application, such a probability space represents a *random experiment*, that is, the empirical phenomenon considered. (Note, however, that in the framework of Bayesian statistics, a part of the probability measure also represents subjective probabilities.) Example 1.31 on diagnosing an infection via a virus test illustrates that the empirical phenomenon can change within a short period of time. In the example, the crucial change is in the incidence, that is, the probability of being infected, which has dramatic implications for the probability of being infected given a positive test.

In this example, we already used the concept of a *conditional probability* $P(A|B)$ of an event A given an event B, which is the most simple conceptual tool for describing dependencies among events. This notion is also used to define the B-conditional-probability measure on the set $\mathscr{A}$ of events.

Another important concept is *independence of two events* A *and* B, denoted $A \perp\!\!\!\perp B$, which is closely related to conditional probability, because $A \perp\!\!\!\perp B$ implies $P(A|B) = P(A)$. We also defined independence of more than two events introducing the concept of a *family* $(A_i, i \in I)$ *of independent events*. For $I = \{1,2\}$ this is equivalent to independence of A_1 and A_2. For $I = \{1,2,3\}$, however, $\perp\!\!\!\perp (A_i, i \in I)$ is equivalent to pairwise and tripelwise independence of the events A_1, A_2, and A_3 [see Eqs. (1.29) and (1.30)]. Finally, the concept of independence of events and families of events has then been extended to *independence of set systems*, *independence of families of set systems*, and *conditional independence of set systems* and *conditional independence of families of set systems* (see Box 1.4 for details).

1.8 Exercises

▷ **Exercise 1-1** Let Ω be nonempty. Show that $\{\Omega, \emptyset\}$ is a σ-algebra on Ω.

▷ **Exercise 1-2** Let A be nonempty and $A \subset \Omega$. Show that $\{A, A^c, \Omega, \emptyset\}$ is a σ-algebra on Ω.

▷ **Exercise 1-3** List at least two more set systems on Ω that generate the σ-algebra $\mathscr{A}_1 = \{A, A^c, \Omega, \emptyset\}$ and that are not already mentioned in Example 1.8.

▷ **Exercise 1-4** Assume that $(\Omega, \mathscr{A})$ is a measurable space. Prove

$$\mathscr{E} \subset \mathscr{A} \quad \Rightarrow \quad \sigma(\mathscr{E}) \subset \mathscr{A}.$$

▷ **Exercise 1-5** Let $(\Omega, \mathscr{A})$ be a measurable space. Prove

$$\begin{aligned} \mathscr{C}, \mathscr{D} \subset \mathscr{A} \quad &\Leftrightarrow \quad \mathscr{C} \cup \mathscr{D} \subset \mathscr{A} \\ &\Rightarrow \quad \sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{A}. \end{aligned}$$

▷ **Exercise 1-6** Consider the random experiment displayed in Table 1.2 and compute the probability of the event A that Jo is sampled and the probability of the event B that the sampled person (whoever it is) is treated.

▷ **Exercise 1-7** Consider Table 1.4 and show that the events A that Joe is sampled and B that the sampled person is treated are not independent.

Solutions

▷ **Solution 1-1** Obviously, conditions (a) and (b) of Definition 1.4 are satisfied, because $\Omega^c = \emptyset$ and $\emptyset^c = \Omega$, and $\Omega, \emptyset \in \{\Omega, \emptyset\}$. Condition (c) is satisfied as well. If Ω is at least one of the sets $A_1, A_2, \ldots \in \{\Omega, \emptyset\}$, then $\bigcup_{i=1}^{\infty} A_i = \Omega$, which is an element of $\{\Omega, \emptyset\}$. If Ω is not one of the sets $A_1, A_2, \ldots \in \{\Omega, \emptyset\}$, that is, if $A_1, A_2, \ldots = \emptyset, \emptyset, \ldots$, then $\bigcup_{i=1}^{\infty} A_i = \emptyset$, which is an element of $\{\Omega, \emptyset\}$, too.

▷ **Solution 1-2** Obviously, condition (a) of Definition 1.4 is satisfied, because $\Omega \in \{A, A^c, \Omega, \emptyset\}$. Condition (b) of Definition 1.4 is satisfied as well, because the complement of each element of $\{A, A^c, \Omega, \emptyset\}$ is also an element of this set. Finally, condition (c) is satisfied as well. If Ω is at least one of the sets $A_1, A_2, \ldots \in \{A, A^c, \Omega, \emptyset\}$ or if A and A^c are among the sets $A_1, A_2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i = \Omega$, which is an element of $\{A, A^c, \Omega, \emptyset\}$. If neither Ω nor A or A^c are one of the sets $A_1, A_2, \ldots \in \{A, A^c, \Omega, \emptyset\}$, that is, if $A_1, A_2, \ldots = \emptyset, \emptyset, \ldots$, then $\bigcup_{i=1}^{\infty} A_i = \emptyset$, which is an element of $\{A, A^c, \Omega, \emptyset\}$, too. If A is among the sets $A_1, A_2, \ldots \in \{A, A^c, \Omega, \emptyset\}$ but neither A^c nor Ω, then $\bigcup_{i=1}^{\infty} A_i = A$, which is an element of $\{A, A^c, \Omega, \emptyset\}$. Finally, if A^c is among the sets $A_1, A_2, \ldots \in \{A, A^c, \Omega, \emptyset\}$ but neither A nor Ω, then $\bigcup_{i=1}^{\infty} A_i = A^c$, which is an element of $\{A, A^c, \Omega, \emptyset\}$.

▷ **Solution 1-3** The set systems $\{A, \emptyset\}$, $\{A, \Omega\}$ $\{A^c, \emptyset\}$, and $\{A^c, \Omega\}$ also generate $\mathscr{A}_1 = \{A, A^c, \Omega, \emptyset\}$.

▷ **Solution 1-4** Let $(\Omega, \mathscr{A})$ be a measurable space and $(\mathscr{A}_i, i \in I)$ the family of all σ-algebras on Ω that contain $\mathscr{E}$ as a subset. Then

$$
\begin{aligned}
& \big(\forall\, i \in I\colon\ \mathscr{E} \subset \mathscr{A}_i\big) \wedge \big(\exists\, i \in I\colon \mathscr{A} = \mathscr{A}_i\big) && [\text{def. of } (\mathscr{A}_i, i \in I), \mathscr{A} \text{ is a σ-algebra}] \\
\Rightarrow\ & \Big(\bigcap_{i \in I} \mathscr{A}_i\Big) \subset \mathscr{A} && [\text{Th. } 1.6] \\
\Rightarrow\ & \sigma(\mathscr{E}) \subset \mathscr{A}. && [(1.3)]
\end{aligned}
$$

▷ **Solution 1-5** If $(\Omega, \mathscr{A})$ is a measurable space, then

$$
\begin{aligned}
& \mathscr{C} \cup \mathscr{D} \subset \mathscr{A} \\
\Leftrightarrow\ & \big(\forall\, A \in \mathscr{A}\colon A \in \mathscr{C} \cup \mathscr{D} \ \Rightarrow\ A \in \mathscr{A}\big) \\
\Leftrightarrow\ & \big(\forall\, A \in \mathscr{A}\colon (A \in \mathscr{C} \vee A \in \mathscr{D}) \ \Rightarrow\ A \in \mathscr{A}\big) \\
\Leftrightarrow\ & \big(\forall\, A \in \mathscr{A}\colon (A \in \mathscr{C} \ \Rightarrow\ A \in \mathscr{A}) \wedge (A \in \mathscr{D} \ \Rightarrow\ A \in \mathscr{A})\big) \\
\Leftrightarrow\ & \mathscr{C} \subset \mathscr{A} \wedge \mathscr{D} \subset \mathscr{A}.
\end{aligned}
$$

The proposition $\mathscr{C} \cup \mathscr{D} \subset \mathscr{A} \ \Rightarrow\ \sigma(\mathscr{C} \cup \mathscr{D}) \subset \mathscr{A}$ immediately follows from Proposition (1.4).

▷ **Solution 1-6** According to Box 1.1 (ii), in the random experiment presented in Table 1.2, the probability of the event

$$A \;=\; \big\{(Joe, no, -),\ (Joe, no, +),\ (Joe, yes, -),\ (Joe, yes, +)\big\}$$

that Joe is sampled is

$$
\begin{aligned}
P(A) &= P\big(\{(Joe, no, -)\}\big) + P\big(\{(Joe, no, +)\}\big) + P\big(\{(Joe, yes, -)\}\big) + P\big(\{(Joe, yes, +)\}\big) \\
&= .09 + .21 + .04 + .16 \;=\; .5.
\end{aligned}
$$

Correspondingly, the probability of the event

$$B \;=\; \big\{(Joe, yes, -),\ (Joe, yes, +),\ (Ann, yes, -),\ (Ann, yes, +)\big\}$$

that the sampled person (whoever it is) is treated is

$$
\begin{aligned}
P(B) &= P\big(\{(Joe, yes, -)\}\big) + P\big(\{(Joe, yes, +)\}\big) P\big(\{(Ann, yes, -)\}\big) + P\big(\{(Ann, yes, +)\}\big) \\
&= .04 + .16 + .12 + .08 \;=\; .4.
\end{aligned}
$$

▷ **Solution 1-7** Computing the probabilities of A and B from Table 1.4 in the same way as in Exercise 1-6 yield again $P(A) = .5$ and $P(B) = .4$. However, now $P(A \cap B) = .02$, which is not equal to the product $P(A) \cdot P(B) = .5 \cdot .4 = .2$. Hence, in this example, the events A and B are not independent [see Def. 1.39 (i)].

Chapter 2
Random Variable

Although a probability space contains the complete information about the random experiment considered, this information is not easily grasped. A first concept that helps in processing this information are conditional probabilities. Random variables, their distributions, expectations, variances, and covariances, for example, are other important concepts for processing the information contained in a probability space and, in particular, in a probability measure.

In this chapter, we turn to the concept of a *random variable*, which is a special *measurable mapping*. Important properties of measurable mappings are treated that also apply to random variables. One of these properties is that a measurable mapping *generates a* σ*-algebra*. We introduce special mappings such as an *indicator* of an event, a *projection*, a *multivariate mapping*, and a *composition*. The most important properties of these special mappings are dealt with, which are related to measurability. Then we define the *distribution* of a random variable and turn to *P-equivalence of two random variables*. Among other things, this concept plays a crucial role in understanding the notion of a conditional expectation (see ch. 4). Another section is devoted to *independence of random variables* building on the concepts of independence of sets systems introduced in section 1.4. Finally, we introduce the *probability function of a disrete random variable*, the *distribution function*, and a *density* of a real-valued random variable.

2.1 Measurable Mapping and Random Variable

2.1.1 Definition

In this section, a *random variable on a probability space* $(\Omega, \mathscr{A}, P)$ is defined as a *measurable mapping* on the measurable space $(\Omega, \mathscr{A})$. Hence, the fundamental concept is a measurable mapping, which cannot be understood without the concept of an inverse image.

Remark 2.1 [Inverse Image] The *inverse image of a set* A' *under a mapping* $Y: \Omega \to \Omega'_Y$ (with domain Ω and co-domain Ω'_Y) is the set

$$Y^{-1}(A') := \{\omega \in \Omega: Y(\omega) \in A'\}, \quad A' \subset \Omega'_Y. \tag{2.1}$$

According to this equation, the inverse image $Y^{-1}(A')$ is the subset of Ω for whose elements Y takes on a value in the set A'. Therefore, we also use the notation

$$\{Y \in A'\} := Y^{-1}(A'). \tag{2.2}$$

In a similar vein we introduce the symbols

$$\{Y=y\} := Y^{-1}(\{y\}) \quad \text{and} \quad \{Y \neq y\} := Y^{-1}(\Omega'_Y \setminus \{y\}). \tag{2.3}$$

Hence, $\{Y \neq y\}$ is the set of all elements ω of Ω for which Y takes on a value in the set $\Omega'_Y \setminus \{y\}$, the co-domain of Y without the element y. ◁

Definition 2.2 [Measurable Mapping and Random Variable]
Let $(\Omega, \mathscr{A})$ *and* $(\Omega'_Y, \mathscr{A}'_Y)$ *be measurable spaces. Then the mapping* $Y: \Omega \to \Omega'_Y$ *is called an* $(\mathscr{A}, \mathscr{A}'_Y)$*-measurable mapping on* $(\Omega, \mathscr{A})$ *and* $(\Omega'_Y, \mathscr{A}'_Y)$ *its value space, if*

$$\forall A' \in \mathscr{A}'_Y: \{Y \in A'\} \in \mathscr{A}. \tag{2.4}$$

Furthermore, if P *is a probability measure on* $\mathscr{A}$*, then a measurable mapping on* $(\Omega, \mathscr{A})$ *is also called a random variable on* $(\Omega, \mathscr{A}, P)$.

Remark 2.3 [A Caveat] If we formulate a proposition about a measurable mapping Y that holds irrespective of its value space $(\Omega'_Y, \mathscr{A}'_Y)$, then, for simplicity, we may omit mentioning the value space. For example, we can say that *a random variable on a probability space* $(\Omega, \mathscr{A}, P)$ *is a measurable mapping on the measurable space* $(\Omega, \mathscr{A})$. This proposition holds irrespective of the value space involved. ◁

Remark 2.4 [Notation] We also say that

$$Y: (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$$

is a measurable mapping meaning that $Y: \Omega \to \Omega'_Y$ is an $(\mathscr{A}, \mathscr{A}'_Y)$-measurable mapping on $(\Omega, \mathscr{A})$ and $(\Omega'_Y, \mathscr{A}'_Y)$ is its value space. If there is no ambiguity about $\mathscr{A}'_Y$, then we omit the reference to this σ-algebra and just say that Y is $\mathscr{A}$-measurable or *measurable with respect to* $\mathscr{A}$. If there is no ambiguity about $\mathscr{A}$ and $\mathscr{A}'_Y$, then we also omit the reference to both σ-algebras and say that Y is a measurable mapping. In all these cases we mean that $Y: \Omega \to \Omega'_Y$ is an $(\mathscr{A}, \mathscr{A}'_Y)$-measurable mapping on $(\Omega, \mathscr{A})$ and $(\Omega'_Y, \mathscr{A}'_Y)$ is its value space. ◁

Remark 2.5 [Real-Valued and Numerical Measurable Mappings] If the value space of Y is $(\Omega'_Y, \mathscr{A}'_Y) = (\mathbb{R}, \mathscr{B})$, then Y is called *real-valued*, referring to the set $\mathbb{R}$ of real numbers and the *Borel σ-algebra* $\mathscr{B}$ on this set. Remember, the set $\mathscr{B}$ contains as elements all singletons $\{\alpha\}$, $\alpha \in \mathbb{R}$, and all (open, half-open, and closed) intervals such as $(a,b), (a,b], [a,b), [a,b]$, $a, b \in \mathbb{R}$, $a < b$, as well as their finite or countable unions and intersections. . Similarly, if $(\Omega'_Y, \mathscr{A}'_Y) = (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, then Y is called *numerical*, where $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$ refers to the measurable space consisting of the set $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ and the Borel σ-algebra $\overline{\mathscr{B}}$ on this set. (For more details see SN-section 1.2.2.) Real-valued and numerical measurable mappings are also called *measurable functions*, and in the context of a probability space $(\Omega, \mathscr{A}, P)$ they are also called *real-valued* and *numerical random variables*, respectively. ◁

2.1.2 First Examples

Example 2.6 [Indicator of a Measurable Set] Let $(\Omega, \mathscr{A})$ and $(\mathbb{R}, \mathscr{B})$ be measurable spaces, where $\mathscr{B}$ denotes the Borel σ-algebra on the set $\mathbb{R}$ of real numbers (see Rem. 1.14). Furthermore, let $A \in \mathscr{A}$, that is, let A be a measurable set. Then $1_A: \Omega \to \mathbb{R}$ defined by

$$1_A(\omega) = \begin{cases} 1, \text{ if } \omega \in A \\ 0, \text{ otherwise} \end{cases} \tag{2.5}$$

is a measurable mapping on $(\Omega, \mathscr{A})$ and $(\mathbb{R}, \mathscr{B})$ is its value space. The mapping 1_A is called the *indicator of A*.

There are four different inverse images $1_A^{-1}(B) = \{1_A \in B\}$ of sets $B \in \mathscr{B}$ under 1_A:

$$\forall B \in \mathscr{B}: \quad 1_A^{-1}(B) = \begin{cases} A = \{1_A = 1\}, & \text{if } 0 \notin B \text{ and } 1 \in B \\ A^c = \{1_A = 0\}, & \text{if } 0 \in B \text{ and } 1 \notin B \\ \Omega = \big\{1_A \in \{0,1\}\big\}, & \text{if } 0 \in B \text{ and } 1 \in B \\ \emptyset = \big\{1_A \notin \{0,1\}\big\}, & \text{if } 0 \notin B \text{ and } 1 \notin B. \end{cases}$$

If $A \in \mathscr{A}$, then all four inverse images are elements of $\mathscr{A}$, which follows from the definition of a σ-algebra (see Def. 1.4). Note that the set of these four inverse images is a σ-algebra on Ω (see again Exercise 1-2). ◁

Remark 2.7 [An Alternative Value Space of an Indicator] Also note that, instead of $(\mathbb{R}, \mathscr{B})$, we may also choose $\big(\{0,1\}, \mathscr{P}(\{0,1\})\big)$ as the value space of 1_A. The four possible inverse images are still A, A^c, Ω, and Ø. ◁

Remark 2.8 [Another Notation of an Indicator] If X is a measurable mapping on $(\Omega, \mathscr{A})$ and we consider the set $\{X{=}x\} = \{\omega \in \Omega\colon X(\omega) = x\}$, then, instead of $1_{\{X=x\}}$, we also use the notation $1_{X=x}$ for its indicator, that is,

$$1_{X=x} := 1_{\{X=x\}}. \tag{2.6}$$

In this definition, Remark 2.3 applies to X because the value space $(\Omega'_X, \mathscr{A}'_X)$ is irrelevant for introducing this notation. ◁

Example 2.9 [Joe and Ann With Randomized Assignment] In Equation (1.2) we specified the set $\Omega = \Omega_U \times \Omega_X \times \Omega_Y$ of possible outcomes of the random experiment considered. The third column of Table 1.2 displays the person variable U assigning an element of the set $\Omega_U = \{Joe, Ann\}$ to each possible outcome $\omega_i \in \Omega$. Hence, U is a mapping with domain Ω and co-domain Ω_U, that is, $U\colon \Omega \to \Omega_U$. If we choose $\mathscr{A} = \mathscr{P}(\Omega)$ to be the power set of Ω and the measurable space $(\Omega_U, \mathscr{A}_U)$ with $\mathscr{A}_U = \mathscr{P}(\Omega_U) = \big\{\{Joe\}, \{Ann\}, \Omega_U, \emptyset\big\}$, then the mapping $U\colon \Omega \to \Omega_U$ is a measurable mapping on $(\Omega, \mathscr{A})$ with value space $(\Omega_U, \mathscr{A}_U)$. The four inverse images of elements $A' \in \mathscr{A}_U$ are

$$\begin{aligned} U^{-1}(\{Joe\}) &= \{U{=}Joe\} = \{Joe\} \times \Omega_X \times \Omega_Y = \{\omega_1, \ldots, \omega_4\} \\ U^{-1}(\{Ann\}) &= \{U{=}Ann\} = \{U{\neq}Joe\} = \{Ann\} \times \Omega_X \times \Omega_Y = \{\omega_5, \ldots, \omega_8\} \\ U^{-1}(\Omega_U) &= \{U \in \Omega_U\} = \Omega = \{\omega_1, \ldots, \omega_8\} \\ U^{-1}(\emptyset) &= \{U \in \emptyset\} = \emptyset. \end{aligned}$$

Because we choose $\mathscr{A}$ to be the power set of Ω, all four inverse images $U^{-1}(A'), A' \in \mathscr{A}_U$, are necessarily elements of $\mathscr{A}$ [see Eq. (2.4)]. Note again that the set of these four inverse images is a σ-algebra on Ω, because $\{U{=}Ann\} = \{U{=}Joe\}^c$ (see again Exercise 1-2).

The fourth column of Table 1.2 displays the treatment variable X assigning an element of the set ℝ, namely the numbers 0 and 1, to each possible outcome $\omega_i \in \Omega$. If we choose $\mathscr{A} = \mathscr{P}(\Omega)$ to be the power set of Ω and the measurable space $(\mathbb{R}, \mathscr{B})$, then the mapping $X\colon \Omega \to \mathbb{R}$ is a measurable mapping on $(\Omega, \mathscr{A})$ and $(\mathbb{R}, \mathscr{B})$ its value space. ◁

Example 2.10 [Nonorthogonal Factors] In Example 1.3 the set

$$\Omega \;=\; \Omega_U \times \Omega_X \times \Omega_Y$$

of possible outcomes of this random experiment has been specified. The person variable $U\colon \Omega \to \Omega_U$ appearing in Table 1.3 is defined by

$$\forall\, \omega \in \Omega\colon \quad U(\omega) \;=\; \begin{cases} \mathit{Tom}, \text{ if } \omega \in \{\mathit{Tom}\} \times \Omega_X \times \Omega_Y \\ \mathit{Tim}, \text{ if } \omega \in \{\mathit{Tim}\} \times \Omega_X \times \Omega_Y \\ \quad\vdots \\ \mathit{Mia}, \text{ if } \omega \in \{\mathit{Mia}\} \times \Omega_X \times \Omega_Y. \end{cases} \tag{2.7}$$

Because we chose $\mathscr{A}$ to be the product of the σ-algebras $\mathscr{A}_U = \mathscr{P}(\Omega_U)$, $\mathscr{A}_X = \mathscr{P}(\Omega_X)$, and the Borel σ-algebra $\mathscr{B}$ (see Example 1.13), the definition of a product of σ-algebras (see SN-Def. 1.31) implies that all inverse images $\{U \in A'\} = U^{-1}(A')$, $A' \in \mathscr{A}_U$, are elements of $\mathscr{A}$ [see again Eq. (2.4)]. Hence, U is a measurable mapping on $(\Omega, \mathscr{A})$ and $(\Omega_U, \mathscr{A}_U)$ is its value space.

The treatment variable $X\colon \Omega \to \mathbb{R}$ appearing in Table 1.3 is defined by

$$\forall\, \omega \in \Omega\colon \quad X(\omega) \;=\; \begin{cases} 0, & \text{if } \omega \in \Omega_U \times \{\mathit{control}\} \times \Omega_Y \\ 1, & \text{if } \omega \in \Omega_U \times \{\mathit{treatment}\,1\} \times \Omega_Y \\ 2, & \text{if } \omega \in \Omega_U \times \{\mathit{treatment}\,2\} \times \Omega_Y. \end{cases} \tag{2.8}$$

Again, because we chose $\mathscr{A}$ to be the product of the σ-algebras $\mathscr{A}_U = \mathscr{P}(\Omega_U)$, $\mathscr{A}_X = \mathscr{P}(\Omega_X)$, and the Borel σ-algebra $\mathscr{B}$, the definition of a product of σ-algebras implies that all inverse images $\{X \in B\} = X^{-1}(B)$, $B \in \mathscr{B}$, are elements of $\mathscr{A}$ [see again Eq. (2.4)]. Hence, X is a measurable mapping on $(\Omega, \mathscr{A})$ and $(\mathbb{R}, \mathscr{B})$ is its value space (see also Exercise 2-1). The inverse images $X^{-1}(B)$, $B \in \mathscr{B}$, are

$$X^{-1}(B) \;=\; \begin{cases} \Omega_U \times \{\mathit{control}\} \times \Omega_Y \;=\; \{X{=}0\}, & \text{if } 0 \in B, 1,2 \notin B \\ \Omega_U \times \{\mathit{treatment}\,1\} \times \Omega_Y \;=\; \{X{=}1\}, & \text{if } 0,2 \notin B, 1 \in B \\ \Omega_U \times \{\mathit{treatment}\,2\} \times \Omega_Y \;=\; \{X{=}2\}, & \text{if } 0,1 \notin B, 2 \in B \\ \Omega_U \times \{\mathit{control},\, \mathit{treatment}\,1\} \times \Omega_Y \;=\; \big\{X \in \{0,1\}\big\}, & \text{if } 0,1 \in B, 2 \notin B \\ \Omega_U \times \{\mathit{control},\, \mathit{treatment}\,2\} \times \Omega_Y \;=\; \big\{X \in \{0,2\}\big\}, & \text{if } 0,2 \in B, 1 \notin B \\ \Omega_U \times \{\mathit{treatment}\,1,\, \mathit{treatment}\,2\} \times \Omega_Y \;=\; \big\{X \in \{1,2\}\big\}, & \text{if } 0 \notin B, 1,2 \in B \\ \Omega \;=\; \big\{X \in \{0,1,2\}\big\}, & \text{if } 0,1,2 \in B \\ \emptyset \;=\; \big\{X \in \mathbb{R} \setminus \{0,1,2\}\big\}, & \text{if } 0,1,2 \notin B. \end{cases} \tag{2.9}$$

Note that the set of these eight inverse images is a σ-algebra on Ω.

The (qualitative) covariate $Z\colon \Omega \to \Omega'_Z$ with codomain $\Omega'_Z := \{\mathit{low}, \mathit{med}, \mathit{hi}\}$, which appears in Table 1.3, is defined by

$$\forall\, \omega \in \Omega\colon \quad Z(\omega) \;=\; \begin{cases} \mathit{low}, & \text{if } \omega \in \{\mathit{Tim}, \mathit{Tom}\} \times \Omega_X \times \Omega_Y \\ \mathit{med}, & \text{if } \omega \in \{\mathit{Joe}, \ldots, \mathit{Eva}\} \times \Omega_X \times \Omega_Y \\ \mathit{hi}, & \text{if } \omega \in \{\mathit{Sue}, \mathit{Mia}\} \times \Omega_X \times \Omega_Y. \end{cases} \tag{2.10}$$

For this mapping we choose the value space $\big(\Omega'_Z, \mathscr{P}(\Omega'_Z)\big)$. Again, because we chose $\mathscr{A}$ to be the product of the σ-algebras $\mathscr{A}_U = \mathscr{P}(\Omega_U)$, $\mathscr{A}_X = \mathscr{P}(\Omega_X)$, and the Borel σ-algebra $\mathscr{B}$, the definition of a product of σ-algebras implies that all inverse images $Z^{-1}(A')$, $A' \in \mathscr{P}(\Omega'_Z)$, are elements of $\mathscr{A}$ [see again Eq. (2.4)].

The inverse images $Z^{-1}(A')$, $A' \in \mathscr{P}(\Omega'_Z)$, are

$$
\begin{aligned}
Z^{-1}(\{low\}) &= \{Z{=}low\} = \{Tim, Tom\} \times \Omega_X \times \Omega_Y \\
Z^{-1}(\{med\}) &= \{Z{=}med\} = \{Joe, Jim, Ann, Eva\} \times \Omega_X \times \Omega_Y \\
Z^{-1}(\{hi\}) &= \{Z{=}hi\} = \{Sue, Mia\} \times \Omega_X \times \Omega_Y \\
Z^{-1}(\{low, med\}) &= \big\{Z \in \{low, med\}\big\} = \{Tim, Tom, Joe, Jim, Ann, Eva\} \times \Omega_X \times \Omega_Y \\
Z^{-1}(\{low, hi\}) &= \big\{Z \in \{low, hi\}\big\} = \{Tim, Tom, Sue, Mia\} \times \Omega_X \times \Omega_Y \\
Z^{-1}(\{med, hi\}) &= \big\{Z \in \{med, hi\}\big\} = \{Joe, Jim, Ann, Eva, Sue, Mia\} \times \Omega_X \times \Omega_Y \\
Z^{-1}(\Omega'_Z) &= \{Z \in \Omega'_Z\} = \Omega \\
Z^{-1}(\emptyset) &= \{Z \in \emptyset\} = \emptyset.
\end{aligned}
$$

Note again that the set of these eight inverse images is a σ-algebra on Ω. As shown in Theorem 2.11 this is not a coincidence. ◁

2.1.3 σ-Algebra Generated by a Measurable Mapping

In Examples 2.6, 2.9, and 2.10 we already noted that the set of all inverse images under a measurable mapping is a σ-algebra on Ω. In a sense, such a σ-algebra carries the information of the measurable mapping Y; it contains all measurable sets that can be represented by Y. Hence, in the context of a probability space $(\Omega, \mathscr{A}, P)$, it contains all events represented by Y. In the following theorem we formulate the general proposition.

Theorem 2.11 **[σ-Algebra Generated by a Measurable Mapping]**
Let Y be a measurable mapping on $(\Omega, \mathscr{A})$ and $(\Omega'_Y, \mathscr{A}'_Y)$ its value space. Then

$$Y^{-1}(\mathscr{A}'_Y) := \big\{Y^{-1}(A') \colon A' \in \mathscr{A}'_Y\big\} \tag{2.11}$$

is a σ-algebra on Ω.

For a proof see Klenke (2020, Th. 1.78, p. 36).

The set $Y^{-1}(\mathscr{A}'_Y)$ contains all sets in $\mathscr{A}$ that can be represented by Y and the elements of $\mathscr{A}'_Y$. Because the set system $Y^{-1}(\mathscr{A}'_Y)$ is important, it has an own name and an alternative notation, which is often more convenient.

Definition 2.12 **[σ-Algebra Generated by a Measurable Mapping]**
The set $Y^{-1}(\mathscr{A}'_Y)$ defined by Equation (2.11) is called the σ-algebra generated by Y and $\mathscr{A}'_Y$. If there is no ambiguity about $\mathscr{A}'_Y$, then we also say that $Y^{-1}(\mathscr{A}'_Y)$ is generated by Y and use the notation

$$\sigma(Y) := Y^{-1}(\mathscr{A}'_Y). \tag{2.12}$$

Remark 2.13 [Measurability With Respect to a Measurable Mapping] Let Z be a measurable mapping on $(\Omega, \mathscr{A})$, let $(\Omega'_Z, \mathscr{A}'_Z)$ denote its value space, and assume

$$\sigma(Y) \subset \sigma(Z). \tag{2.13}$$

Then we say that Y is *Z-measurable* or *measurable with respect to Z*. ◁

Example 2.14 [Measurability of an Indicator] Let $X\colon (\Omega, \mathscr{A}) \to (\Omega', \mathscr{A}')$ be a measurable mapping, $x \in \Omega'$ be a value of X, let $\{x\} \in \mathscr{A}'$, and let $1_{X=x}$ denote the indicator of $\{X=x\}$. Then $\{X=x\} \in \mathscr{A}$ and

$$\sigma(1_{X=x}) \subset \sigma(X) \tag{2.14}$$

(see Exercise 2-2). ◁

2.1.4 Projection

In Examples 2.9 and 2.10, we used the mapping $U\colon \Omega \to \Omega_U$, where Ω_U is one of the factor sets of $\Omega = \Omega_U \times \Omega_X \times \Omega_Y$. According to the following definition, U is a *projection* (or *coordinate mapping*). For our purposes it suffices to consider only a finite Cartesian product set.

Definition 2.15 [Projection or Coordinate Mapping]
Let $\Omega = \bigtimes_{i=1}^{m} \Omega_i$. *Then* $\pi_i\colon \Omega \to \Omega_i$ *defined by*

$$\forall\, \omega \in \Omega\colon\; \pi_i(\omega) = \pi_i(\omega_1, \ldots, \omega_i, \ldots, \omega_m) = \omega_i \tag{2.15}$$

is called the ith projection or coordinate mapping.

Hence, in Examples 2.9 and 2.10, the mapping U is the first projection. In contrast, in Example 2.9, the treatment variable X is not a projection because the co-domain is $\Omega'_X = \{0, 1\}$ and not $\Omega_X = \{no, yes\}$.

Remark 2.16 [The Projection is a Measurable Mapping] If $(\Omega, \mathscr{A})$ is the product of the measurable spaces $(\Omega_i, \mathscr{A}_i)$, $i = 1, \ldots, m$ (see Def. 1.15), then each projection π_i, $i = 1, \ldots, m$, is an $(\mathscr{A}, \mathscr{A}_i)$-measurable mapping. Its value space is $(\Omega_i, \mathscr{A}_i)$. ◁

2.1.5 Multivariate Mapping

Now consider the product of the measurable spaces $(\Omega_i, \mathscr{A}_i)$, $i = 1, \ldots, m$, and note that the definitions of a measurable mapping and of a σ-algebra generated by a mapping also apply to *m-variate mappings* $Y\colon \Omega \to \Omega'_1 \times \ldots \times \Omega'_m$, and in particular to functions for which $\Omega'_1 \times \ldots \times \Omega'_m = \mathbb{R}^m$. Such an m-variate mapping is an m-tuple of mappings $Y_i\colon \to \Omega'_i$, that is, $Y = (Y_1, \ldots, Y_m)$.

Reading the following lemma, remember that $\bigotimes_{i=1}^{m} \mathscr{A}'_i$ denotes product of the σ-algebras $\mathscr{A}'_i$ (see Def. 1.15) and $Y_i^{-1}(\mathscr{A}'_i)$ the σ-algebra generated by Y_i (see Def. 2.12).

Lemma 2.17 **[σ-Algebra Generated by a Multivariate Mapping]**
Let Ω be a nonempty set, let $(\Omega_i', \mathscr{A}_i')$, $i = 1,\ldots,m$, $m \in \mathbb{N}$, be measurable spaces, and $Y\colon \Omega \to \Omega_1' \times \ldots \times \Omega_m'$ an m-variate mapping, that is, $Y = (Y_1,\ldots,Y_m)$ with $Y_i\colon \Omega \to \Omega_i'$, $i = 1,\ldots,m$. Then

$$\sigma(Y_1,\ldots,Y_m) := \sigma(Y) = Y^{-1}\left(\bigotimes_{i=1}^{m} \mathscr{A}_i'\right) = \sigma\left(\bigcup_{i=1}^{m} Y_i^{-1}(\mathscr{A}_i')\right) = \sigma\left(\bigcup_{i=1}^{m} \sigma(Y_i)\right). \tag{2.16}$$

Hence, the σ-algebra generated by an m-variate measurable mapping is the σ-algebra generated by the union of the σ-algebras $\sigma(Y_i) = Y_i^{-1}(\mathscr{A}_i')$ generated by the mappings Y_i. (For a proof of this lemma see SN-Lemma 2.37).

According to the following theorem, a multivariate mapping $Y = (Y_1,\ldots,Y_m)$ is measurable if and only if all its components Y_i are measurable. Reading this theorem, remember that $\left(\bigtimes_{i=1}^{m} \Omega_i', \bigotimes_{i=1}^{m} \mathscr{A}_i'\right)$ denotes the product of the measurable spaces $(\Omega_i, \mathscr{A}_i)$, $i = 1,\ldots,m$ (see Def. 1.15.) (For a proof see SN-Th. 2.38).

Theorem 2.18 **[Measurability of a Multivariate Mapping]**
Under the assumptions of Lemma 2.17, the following two propositions are equivalent to each other:

(a) $Y = (Y_1,\ldots,Y_m)\colon (\Omega,\mathscr{A}) \to \left(\bigtimes_{i=1}^{m} \Omega_i', \bigotimes_{i=1}^{m} \mathscr{A}_i'\right)$ is a measurable mapping.
(b) Each $Y_i\colon (\Omega,\mathscr{A}) \to (\Omega_i', \mathscr{A}_i')$, $i = 1,\ldots,m$, is a measurable mapping.

Remark 2.19 **[σ-Algebra Generated by a Family of Mappings]** Let I be a (finite, countable, or uncountable) index set and let $(Y_i, i \in I)$ be a family of measurable mappings $Y_i\colon (\Omega,\mathscr{A}) \to (\Omega_i', \mathscr{A}_i')$. The σ-algebra generated by this family is defined as

$$\sigma(Y_i, i \in I) := \sigma\Big(\bigcup_{i\in I} \sigma(Y_i)\Big). \tag{2.17}$$

◁

Example 2.20 **[Joe and Ann With Randomized Assignment]** In Table 1.2 we specified the function $X\colon \Omega \to \mathbb{R}$ indicating with its values 1 and 0 whether or not the drawn person is treated and the function $Y\colon \Omega \to \mathbb{R}$ indicating with its values 1 and 0 whether or not the drawn person is successful. If we specify the σ-algebra $\mathscr{A}$ on Ω such that X and Y are both $(\mathscr{A},\mathscr{B})$-measurable, then the bivariate function $(X,Y)\colon \Omega \to \mathbb{R}^2$ is $(\mathscr{A},\mathscr{B}_2)$-measurable, where $\mathscr{B}_2$ denotes the *Borel σ-algebra on* $\mathbb{R}^2$, which is the σ-algebra generated by the set system $\mathscr{I}_2$ of all half-open rectangles in $\mathbb{R}^2$, that is, by the set system

$$\mathscr{I}_2 = \big\{(x_1,x_2) \in \mathbb{R}^2\colon a_1 < x_1 \le b_1,\ a_2 < x_2 \le b_2\big\}, \quad a_1,b_1,a_2,b_2 \in \mathbb{R}. \tag{2.18}$$

And vice versa, if we specify the σ-algebra $\mathscr{A}$ on Ω such that the bivariate function $(X,Y)\colon \Omega \to \mathbb{R}^2$ is $(\mathscr{A},\mathscr{B}_2)$-measurable then X and Y are both $(\mathscr{A},\mathscr{B})$-measurable. In this example X, Y, and (X,Y) are measurable with respect to $\mathscr{A}$ whenever the two inverse images $X^{-1}(\{1\})$ and $Y^{-1}(\{1\})$ are elements of $\mathscr{A}$ (see SN-Exercise 2-10). ◁

Remark 2.21 [Lower Dimensional Multivariate Mappings] Lemma 2.17 and Remark 1.10 imply

$$\forall J \subset I = \{1, \ldots, m\}: \quad \sigma(Y_i, i \in J) \subset Y^{-1}\left(\bigotimes_{i=1}^{m} \mathscr{A}_i'\right).$$

Furthermore, Theorem 2.18 implies: If

$$Y = (Y_1, \ldots, Y_m): (\Omega, \mathscr{A}) \to \left(\bigtimes_{i=1}^{m} \Omega_i', \bigotimes_{i=1}^{m} \mathscr{A}_i'\right)$$

is a measurable mapping and $J = \{i_1, \ldots, i_k\} \subset I$, $k \le m$, then

$$Y_J := (Y_{i_1}, \ldots, Y_{i_k}): (\Omega, \mathscr{A}) \to \left(\bigtimes_{j=1}^{k} \Omega_{i_j}', \bigotimes_{j=1}^{k} \mathscr{A}_{i_j}'\right)$$

is measurable as well. ◁

Remark 2.22 [An Implication of Z-Measurability of Y] Let Y and Z be measurable mappings on $(\Omega, \mathscr{A})$. If Y is Z-measurable, then the *bivariate measurable mapping* (Y, Z) is Z-measurable as well, that is,

$$\sigma(Y) \subset \sigma(Z) \quad \Rightarrow \quad \sigma(Y, Z) = \sigma(Z). \tag{2.19}$$

Furthermore,

$$\sigma(Y) \subset \sigma(Y, Z). \tag{2.20}$$

These propositions immediately follow from Lemma 2.17. ◁

2.1.6 Composition

The following lemmas will be useful whenever we deal with compositions of measurable mappings. Let $X: \Omega \to \Omega'$ and $g: \Omega' \to \Omega''$ be mappings. Then a mapping $Z: \Omega \to \Omega''$ is called the *composition of X and g*, denoted $g \circ X$ or $g(X)$ if

$$Z(\omega) = g\big(X(\omega)\big), \quad \forall\, \omega \in \Omega. \tag{2.21}$$

Lemma 2.23 [Measurability of a Composition]
If $X: (\Omega, \mathscr{A}) \to (\Omega', \mathscr{A}')$ and $g: (\Omega', \mathscr{A}') \to (\Omega'', \mathscr{A}'')$ are measurable mappings and $Z = g \circ X$, then Z is $\mathscr{A}$-measurable, that is, then $Z^{-1}(\mathscr{A}'') \subset \mathscr{A}$.

According to this lemma, measurability is preserved by the composition of mappings. (For a proof see SN-Th. 2.49.) Note that no requirements are necessary for the three measurable spaces involved.

In contrast, in the following lemma, we presume that the mapping g is numerical. In this case, X-measurability of Z is equivalent to the existence of a function g such that $g^{-1}(\overline{\mathscr{B}}) \subset \mathscr{A}'$ and $Z = g \circ X$. (For a proof see Klenke, 2020, Cor. 1.97, pp. 42.)

Lemma 2.24 [Factorization Lemma of Measurable Functions]
Let $X: \Omega \to \Omega'$ be a mapping, let $(\Omega', \mathscr{A}')$ be a measurable space, and let $Z: \Omega \to \overline{\mathbb{R}}$ be a function. Then Z is measurable with respect to X, that is, $Z^{-1}(\overline{\mathscr{B}}) \subset X^{-1}(\mathscr{A}')$, if and only if there is a measurable function $g: (\Omega', \mathscr{A}') \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ such that $Z = g \circ X$ is the composition of X and g. We call g a factorization of Z with respect to X.

If, instead of $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$ we consider a measurable space $\big(\Omega'', \mathscr{P}(\Omega'')\big)$, where Ω'' is finite or countable and $\mathscr{P}(\Omega'')$ denotes the power set of Ω'', then the elements $\omega'' \in \Omega''$ can be renamed by real numbers such as $1, 2$, etc. Renaming is a one-to-one measurable function, because the σ-algebra on Ω'' is the power set of Ω''. Hence, Lemma 2.24 implies the following corollary:

Corollary 2.25 [Factorization Lemma of a Mapping Into a Finite or Countable Set]
Assume that $X: (\Omega, \mathscr{A}) \to (\Omega', \mathscr{A}')$ is a measurable mapping, and let $Z: \Omega \to \Omega''$ be a mapping, where Ω'' is finite or countable. Then

$$Z^{-1}\big(\mathscr{P}(\Omega'')\big) \subset X^{-1}(\mathscr{A}') \quad \Leftrightarrow \quad \exists\, g: (\Omega', \mathscr{A}') \to \big(\Omega'', \mathscr{P}(\Omega'')\big) \text{ such that } Z = g \circ X.$$

Example 2.26 [Joe and Ann With Randomized Assignment] Consider again the mappings $X, Y: \Omega \to \{0,1\}$ displayed in Table 1.2. As a value space of X and Y we can choose $\big(\{0,1\}, \mathscr{P}(\{0,1\})\big)$. Furthermore, we choose $\big(\{0,1\} \times \{0,1\}, \mathscr{P}(\{0,1\} \times \{0,1\})\big)$ as a value space of the bivariate random variable (X, Y). Finally, define the random variable $(X \cdot Y): \Omega \to \overline{\mathbb{R}}$ with value space $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$ by

$$\forall\, \omega \in \Omega: \quad (X \cdot Y)(\omega) \;=\; X(\omega) \cdot Y(\omega) \;=\; \begin{cases} 1, \text{ if } X(\omega) = Y(\omega) = 1 \\ 0, \text{ otherwise,} \end{cases} \tag{2.22}$$

and consider the function $g: \{0,1\} \times \{0,1\} \to \overline{\mathbb{R}}$ defined by

$$\forall\, (x, y) \in \{0,1\} \times \{0,1\}: \quad g(x, y) \;=\; x \cdot y \;=\; \begin{cases} 1, \text{ if } x = y = 1 \\ 0, \text{ otherwise.} \end{cases} \tag{2.23}$$

According to Lemma 2.24, the mapping $X \cdot Y$ is (X, Y)-measurable (see Exercise 2-3). Furthermore, $X \cdot Y$ and $g(X, Y)$ are identical, because

$$\forall\, \omega \in \Omega: \quad (X \cdot Y)(\omega) \;=\; g(X, Y)(\omega) \;=\; g\big(X(\omega), Y(\omega)\big). \tag{2.24}$$

◁

Many more useful details related to measurability of mappings are found in SN-chapter 2, for example, on measurability of functions of measurable functions such as their weighted sums or their products.

2.2 Distribution of a Random Variable

In Definition 2.2 we defined a random variable Y on a probability space $(\Omega, \mathscr{A}, P)$ as a measurable mapping $Y: (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$. There, we required that the inverse images $Y^{-1}(A') = \{Y \in A'\}$ of all sets $A' \in \mathscr{A}'_Y$ under Y are elements of the σ-algebra $\mathscr{A}$ on Ω. Because the probability measure $P: \mathscr{A} \to [0,1]$ assigns a probability $P(A)$ to *all elements* of $\mathscr{A}$, this property allows us to define the *distribution* of a random variable as follows:

Definition 2.27 [Distribution of a Random Variable]
Let Y be a random variable on $(\Omega, \mathscr{A}, P)$ and let $(\Omega'_Y, \mathscr{A}'_Y)$ be its value space. Then the function $P_Y: \mathscr{A}'_Y \to [0,1]$ defined by

$$P_Y(A') = P\big(Y^{-1}(A')\big), \quad \forall\, A' \in \mathscr{A}'_Y, \tag{2.25}$$

is called the distribution of Y (with respect to P).

Remark 2.28 [A New Probability Space] Definition 2.27 implies that *every* random variable Y on a probability space $(\Omega, \mathscr{A}, P)$ has a distribution P_Y. Furthermore, $P_Y: \mathscr{A}'_Y \to [0,1]$ is also a measure, the *image measure* of P under Y (see SN-Th. 2.79 and SN-Def. 2.80). Because $P_Y(\Omega'_Y) = P(\Omega) = 1$, we can conclude that P_Y is a probability measure and that $(\Omega'_Y, \mathscr{A}'_Y, P_Y)$ is a probability space. ◁

Example 2.29 [Joe and Ann With Randomized Assignment] In Example 2.9 we showed that U is a random variable on $(\Omega, \mathscr{A}, P)$ with value space $(\Omega_U, \mathscr{A}_U)$, where $\Omega_U = \{Joe, Ann\}$ and $\mathscr{A}_U = \big\{\{Joe\}, \{Ann\}, \Omega_U, \emptyset\big\}$. The distribution of U is

$$\begin{aligned}
P_U(\{Joe\}) &= P\big(U^{-1}(\{Joe\})\big) = P(\{\omega_1, \ldots, \omega_4\}) = P(\{\omega_1\}) + \ldots + P(\{\omega_4\}) = .5,\\
P_U(\{Ann\}) &= P\big(U^{-1}(\{Ann\})\big) = P(\{\omega_5, \ldots, \omega_8\}) = P(\{\omega_5\}) + \ldots + P(\{\omega_8\}) = .5,\\
P_U(\{\Omega_U\}) &= P\big(U^{-1}(\Omega_U)\big) = P(\{\omega_1, \ldots, \omega_8\}) = P(\{\omega_1\}) + \ldots + P(\{\omega_8\}) = 1,\\
P_U(\{\emptyset\}) &= P\big(U^{-1}(\emptyset)\big) = P(\emptyset) = 0.
\end{aligned}$$

In Example 2.9 we also specified the random variable X on $(\Omega, \mathscr{A}, P)$ with value space $(\mathbb{R}, \mathscr{B})$, where $\mathscr{B}$ denotes the Borel σ-algebra. The distribution of X is as follows:

$$\forall B \in \mathscr{B}: \quad P_X(B) = P\big(X^{-1}(B)\big) = \begin{cases} P(\{\omega_1, \omega_2, \omega_5, \omega_6\}) = .6, & \text{if } 0 \in B \text{ and } 1 \notin B\\ P(\{\omega_3, \omega_4, \omega_7, \omega_8\}) = .4, & \text{if } 0 \notin B \text{ and } 1 \in B\\ P(\Omega) = 1, & \text{if } 0 \in B \text{ and } 1 \in B\\ P(\emptyset) = 0, & \text{if } 0 \notin B \text{ and } 1 \notin B. \end{cases}$$

Note that in this equation we assign a probability to *all* elements $B \in \mathscr{B}$, not only to four elements of $\mathscr{B}$. For example, according to the assignment rule displayed above, $P_X([0, 0.5]) = P_X([-10, 0.5]) = P_X([0, 0.9]) = P_X(\{0\}) = .6$. ◁

Example 2.30 [Joe and Ann With Randomized Assignment] Now we consider the distribution of the bivariate random variable (X, Y) in the random experiment displayed in Table 1.2. A value space of (X, Y) is $\big(\{0,1\}\times\{0,1\}, \mathscr{P}(\{0,1\}\times\{0,1\})\big)$. The probabilities $P_{X,Y}(\{x, y\})$ of the singletons of the power set $\mathscr{P}(\{0,1\}\times\{0,1\})$ can be specified as follows:

$$P_{X,Y}(B) \;=\; P\big((X,Y)^{-1}(B)\big) \;=\; \begin{cases} P(\{\omega_1,\omega_5\}) \;=\; .33, & \text{if } B=\{(0,0)\} \\ P(\{\omega_2,\omega_6\}) \;=\; .27, & \text{if } B=\{(0,1)\} \\ P(\{\omega_3,\omega_7\}) \;=\; .16, & \text{if } B=\{(1,0)\} \\ P(\{\omega_4,\omega_8\}) \;=\; .24, & \text{if } B=\{(1,1)\}. \end{cases}$$

The probabilities of all other elements of $\mathscr{P}(\{0,1\}\times\{0,1\})$ can be computed from these four probabilities via Rule (x) of Box 1.1 (see Exercise 2-6). ◁

Example 2.31 [Nonorthogonal Factors] In Example 2.10 we defined the random variable U with value space $(\Omega_U, \mathscr{A}_U)$, where $\mathscr{A}_U = \mathscr{P}(\Omega_U)$. The distribution of U can be specified as follows:

$$\forall\,\{u\}\in\mathscr{A}_U\colon \quad P_U(\{u\}) \;=\; P\big(U^{-1}(\{u\})\big) \;=\; \frac{1}{8}. \tag{2.26}$$

The singletons are disjunct and, except for the empty set, all $A'\in\mathscr{A}_U$ are unions of the singletons $\{u\}$. Therefore, the probabilities of all elements $A'\in\mathscr{A}_U$ can be computed as a sum of the probabilities of these singletons. For example, the probability of sampling a male person is

$$\begin{aligned} P_U(\{Tom, Tim, Joe, Jim\}) \;&=\; P\big(U^{-1}(\{Tom, Tim, Joe, Jim\})\big) \\ &=\; P_U(\{Tom\}) + P_U(\{Tim\}) + P_U(\{Joe\}) + P_U(\{Jim\}) \\ &=\; \frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8} \;=\; \frac{1}{2} \end{aligned}$$

[see Box 1.1 (ii)]. Note that the distribution of U, that is, the function $P_U\colon \mathscr{A}_U \to [0,1]$, assigns a probability $P_U(A')$ to $2^8 = 256$ sets $A'\in\mathscr{A}_U$.

In Example 2.10 we also specified the random variable X with value space $(\mathbb{R},\mathscr{B})$. The distribution of X can be specified as follows:

$$\begin{aligned} \forall B\in\mathscr{B}\colon \quad P_X(B) \;&=\; P\big(X^{-1}(B)\big) \\ &=\; \begin{cases} P(\Omega_U \times \{control\} \times \Omega_Y) \;=\; 1/3, & \text{if } 0\in B, 1,2\notin B \\ P(\Omega_U \times \{treatment\,1\} \times \Omega_Y) \;=\; 1/3, & \text{if } 0,2\notin B, 1\in B \\ P(\Omega_U \times \{treatment\,2\} \times \Omega_Y) \;=\; 1/3, & \text{if } 0,1\notin B, 2\in B \\ P(\Omega_U \times \{control,\ treatment\,1\} \times \Omega_Y) \;=\; 2/3, & \text{if } 0,1\in B, 2\notin B \\ P(\Omega_U \times \{control,\ treatment\,2\} \times \Omega_Y) \;=\; 2/3, & \text{if } 0,2\in B, 1\notin B \\ P(\Omega_U \times \{treatment\,1,\ treatment\,2\} \times \Omega_Y) \;=\; 2/3, & \text{if } 0\notin B, 1,2\in B \\ P(\Omega) \;=\; 1, & \text{if } 0,1,2\in B \\ P(\emptyset) \;=\; 0, & \text{if } 0,1,2\notin B. \end{cases} \end{aligned}$$

Although there are only eight elements in $\sigma(X)$, this equation assigns a probability to *all* elements $B\in\mathscr{B}$, the number of which is uncountable. ◁

2.3 P-Equivalent Random Variables

Two random variables X and Y on a probability space $(\Omega, \mathscr{A}, P)$ can be *identical* or *equivalent with respect to a probability measure*. If we consider the probability measure P, then related terms are *X and Y are almost surely identical with respect to P* and *the values of X and Y are identical for P-almost all* $\omega \in \Omega$.

2.3.1 Identical Random Variables

Remark 2.32 [Identical Random Variables] Let X and Y be random variables on a probability space $(\Omega, \mathscr{A}, P)$ and let $(\Omega'_X, \mathscr{A}'_X)$ and $(\Omega'_Y, \mathscr{A}'_Y)$ denote their value spaces, respectively. Then X and Y are called *identical*, denoted $X = Y$, if

$$\forall\, \omega \in \Omega\colon\; X(\omega) = Y(\omega). \tag{2.27}$$

If X and Y are identical, then the images $X(\Omega) = \{X(\omega)\colon \omega \in \Omega\}$ and $Y(\Omega) = \{Y(\omega)\colon \omega \in \Omega\}$ are identical, whereas the co-domains Ω'_X and Ω'_Y may differ from each other. ◁

Example 2.33 [Joe and Ann With Randomized Assignment] In Example 2.9 we specified the treatment variable X as a random variable on the probability space $(\Omega, \mathscr{A}, P)$ and its values space $(\mathbb{R}, \mathscr{B})$. Hence, we specified the treatment variable as a measurable mapping $X\colon (\Omega, \mathscr{A}) \to (\mathbb{R}, \mathscr{B})$. In this example, we may also consider the random variable X^* defined by

$$\forall\, \omega \in \Omega\colon\; X^*(\omega) = \begin{cases} 0, \text{ if } \omega \in \{\omega_1, \omega_2, \omega_5, \omega_6\} \\ 1, \text{ if } \omega \in \{\omega_3, \omega_4, \omega_7, \omega_8\}, \end{cases}$$

and, instead of $(\mathbb{R}, \mathscr{B})$, we may choose the value space $\big(\{0,1\}, \mathscr{P}(\{0,1\})\big)$. Then X and X^* are identical although they have different value spaces $(\mathbb{R}, \mathscr{B})$ and $\big(\{0,1\}, \mathscr{P}(\{0,1\})\big)$, respectively. However, $X(\Omega) = X^*(\Omega) = \{0,1\}$, that is, the images of X and X^* are identical.

Note that the distributions P_X and P_{X^*} are not identical, because the domain of P_X is $\mathscr{B}$, whereas the domain of P_{X^*} is $\mathscr{P}(\{0,1\})$. In Example 2.29, we specified the distribution $P_X\colon \mathscr{B} \to [0,1]$ of X. In contrast, the distribution $P_{X^*}\colon \mathscr{P}(\{0,1\}) \to [0,1]$ of X^* is specified by

$$P_{X^*}(A') \;=\; P\big(X^{*-1}(A')\big) \;=\; \begin{cases} P(\{\omega_1, \omega_2, \omega_5, \omega_6\}) \;=\; .6, & \text{if } A' = \{0\} \\ P(\{\omega_3, \omega_4, \omega_7, \omega_8\}) \;=\; .4, & \text{if } A' = \{1\} \\ P(\Omega) \;=\; 1, & \text{if } A' = \{0,1\} \\ P(\emptyset) \;=\; 0, & \text{if } A' = \emptyset \end{cases}$$

for all elements A' of the power set $\mathscr{P}(\{0,1\})$. ◁

Remark 2.34 [More About Identical Random Variables] Hence, if X and Y are identical random variables on $(\Omega, \mathscr{A}, P)$ and $(\Omega'_X, \mathscr{A}'_X)$, $(\Omega'_Y, \mathscr{A}'_Y)$ denote their value spaces, then X and Y have identical values for all $\omega \in \Omega$ [see Eq. (2.27)] and their images $X(\Omega)$ and $Y(\Omega)$ are identical. However, their value spaces and their distributions may differ from each other. Their distributions are identical only if their value spaces are identical. Nevertheless, if $X = Y$, then $P_X(A') = P_Y(A')$, if $A' \in \mathscr{A}'_X$ and $A' \in \mathscr{A}'_Y$. ◁

2.3.2 P-Equivalent Random Variables

Even if random variables on a probability space $(\Omega, \mathscr{A}, P)$ are not identical they still can have identical distributions. This can be the case, for example, if they are *P-equivalent* or *identical almost surely* with respect to P.

Definition 2.35 [P-Equivalent Random Variables]
Let X and Y be random variables on a probability space $(\Omega, \mathscr{A}, P)$ with value spaces $(\Omega'_X, \mathscr{A}'_X)$ and $(\Omega'_Y, \mathscr{A}'_Y)$, respectively. Then X and Y are called P-equivalent or P-almost surely identical, denoted $X \underset{P}{=} Y$, if

$$\exists\, A \in \mathscr{A}\colon\ \big(\forall\, \omega \in \Omega \setminus A\colon\ X(\omega) = Y(\omega) \wedge P(A) = 0\big). \tag{2.28}$$

Obviously, if $X = Y$, then $X \underset{P}{=} Y$ and the set A occurring in Proposition 2.28 is the empty set. In contrast, $X \underset{P}{=} Y$ does not imply $X = Y$ (see Example 2.37).

Remark 2.36 [Propositions Equivalent to P-Equivalence] Under the assumptions of Definition 2.35, the following propositions are equivalent to $X \underset{P}{=} Y$:

$$\exists\, B \in \mathscr{A}\colon\ \big(\forall\, \omega \in B\colon\ X(\omega) = Y(\omega) \wedge P(B) = 1\big). \tag{2.29}$$

$$\neg \exists\, C \in \mathscr{A}\colon\ \big(\forall\, \omega \in C\colon\ X(\omega) \neq Y(\omega) \wedge P(C) > 0\big). \tag{2.30}$$

$\lhd$

Example 2.37 [No Treatment for Joe] Table 2.1 displays another random experiment with Joe and Ann. We consider the columns headed by $P(Y{=}1|X,U)$ and $P(Y{=}1|X,U)^*$, two versions of the (X,U)-conditional probability of the event $\{Y{=}1\}$. The term 'conditional probability of an event given a random variable' will only be introduced in Remark 4.10. In the present context it suffices to know that $P(Y{=}1|X,U)$ and $P(Y{=}1|X,U)^*$ are two random variables and that the rows of these two columns in Table 2.1 show their values for all $\omega \in \Omega$. The two random variables $P(Y{=}1|X,U)$ and $P(Y{=}1|X,U)^*$ are not identical but they are P-equivalent. Inspection of this table shows that Equation (2.27) does not hold, because

$$P(Y{=}1|X,U)(\omega_3) \ \neq\ P(Y{=}1|X,U)^*(\omega_3),$$

for instance. In contrast, Proposition (2.28) does hold, because, for $A = \{\omega_3, \omega_4\}$,

$$\forall\, \omega \in \Omega \setminus A\colon\ P(Y{=}1|X,U)(\omega) = P(Y{=}1|X,U)^*(\omega) \wedge P(A) = 0.$$

Hence, the values of $P(Y{=}1|X,U)$ and $P(Y{=}1|X,U)^*$ are identical for all $\omega \in \Omega \setminus A$. That is, they are identical for all ω except for $\omega \in A$, which has the probability $P(A) = 0$. We call such a set a *null set with respect to P*. $\lhd$

Example 2.38 [No Treatment for Joe] In the same random experiment displayed in Table 2.1 we may also consider the functions $P^{X=1}(Y{=}1|U)$ and $P^{X=1}(Y{=}1|U)^*$ whose values are specified in the last two columns of the table. These two functions are not only random variables on the probability space $(\Omega, \mathscr{A}, P)$, but also on the probability space $(\Omega, \mathscr{A}, P^{X=1})$, whose probability measure $P^{X=1}$ has been specified in the column headed $P^{X=1}(\{\omega_i\})$. The random variables $P^{X=1}(Y{=}1|U)$ and $P^{X=1}(Y{=}1|U)^*$ are not P-equivalent because their values differ for ω_1 to ω_4 and $P(\{\omega_1, \ldots, \omega_4\}) > 0$. However, these two random variables are $P^{X=1}$-equivalent, because their values only differ for ω_1 to ω_4 and $P^{X=1}(\{\omega_1, \ldots, \omega_4\}) = 0$. The set $\{\omega_1, \ldots, \omega_4\}$ is a nullset with respect to $P^{X=1}$. $\lhd$

Table 2.1. No treatment for Joe

Outcomes ω_i (Person, Treatment, Success)	$P(\{\omega_i\})$	$P^{X=0}(\{\omega_i\})$	$P^{X=1}(\{\omega_i\})$	Observables: Person variable U	Treatment variable X	Outcome variable Y	Conditional probabilities: $P(Y=1 \mid X, U)$	$P(Y=1 \mid X, U)^*$	$P(Y=1 \mid X)$	$P(X=1 \mid U)$	$P^{X=0}(Y=1 \mid U)$	$P^{X=1}(Y=1 \mid U)$	$P^{X=1}(Y=1 \mid U)^*$
$\omega_1 = (Joe, no, -)$	.15	.1875	0	*Joe*	0	0	.7	.7	.512	0	.7	99	.8
$\omega_2 = (Joe, no, +)$	.35	.4375	0	*Joe*	0	1	.7	.7	.512	0	.7	99	.8
$\omega_3 = (Joe, yes, -)$	0	0	0	*Joe*	1	0	99	.8	.4	0	.7	99	.8
$\omega_4 = (Joe, yes, +)$	0	0	0	*Joe*	1	1	99	.8	.4	0	.7	99	.8
$\omega_5 = (Ann, no, -)$	.24	.3	0	*Ann*	0	0	.2	.2	.512	.5	.2	.4	.4
$\omega_6 = (Ann, no, +)$	.06	.075	0	*Ann*	0	1	.2	.2	.512	.5	.2	.4	.4
$\omega_7 = (Ann, yes, -)$	.12	0	.6	*Ann*	1	0	.4	.4	.4	.5	.2	.4	.4
$\omega_8 = (Ann, yes, +)$	.08	0	.4	*Ann*	1	1	.4	.4	.4	.5	.2	.4	.4

Remark 2.39 [An Alternative Term] If the random variables X and Y on the probability space $(\Omega, \mathscr{A}, P)$ are P-equivalent, then we also say that *the values of X and Y are identical for P-almost all* $\omega \in \Omega$, abbreviated P-a.a. $\omega \in \Omega$. That is,

$$X \underset{P}{=} Y \quad \Leftrightarrow \quad \big(X(\omega) = Y(\omega), \quad \text{for } P\text{-a.a. } \omega \in \Omega\big). \tag{2.31}$$

◁

Remark 2.40 [Singleton With a Positive Probability] Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$ with value spaces $(\Omega'_X, \mathscr{A}'_X)$ and $(\Omega'_Y, \mathscr{A}'_Y)$, respectively. Furthermore, assume that $\omega \in \Omega$, $\{\omega\} \in \mathscr{A}$, and $P(\{\omega\}) > 0$. Then

$$X \underset{P}{=} Y \quad \Rightarrow \quad X(\omega) = Y(\omega) \tag{2.32}$$

(see SN-Rem. 2.71). ◁

SN-Theorem 2.85 on the equivalence of image measures immediately implies the following corollary on the equivalence of the distributions of two P-equivalent random variables:

Corollary 2.41 [P**-Equivalence Implies Equal Distributions**]
Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$ with the same value space $(\Omega', \mathscr{A}')$ and let P_X and P_Y denote their distributions. Then

$$X \underset{P}{=} Y \quad \Rightarrow \quad P_X = P_Y. \tag{2.33}$$

In other words, if X and Y are P-equivalent and they share the same value space, then their distributions are identical. In chapter 3 we will see that also the expectations, variances, and covariance of X and Y are identical if X and Y are P-equivalent, provided that their expectations, variances, and covariance exist [see Box 3.1 (vi), Box 3.3 (v), and Box 3.4 (xi)]. Note that identical distributions of X and Y do not imply that X and Y are P-equivalent.

Example 2.42 [Flipping Two Coins] Consider the random experiment of flipping two fair coins. The set of possible outcomes is $\Omega = \{(h,h),(h,t),(t,h),(t,t)\}$. As the set of possible events $\mathscr{A}$ we can choose the power set $\mathscr{P}(\Omega)$ and the probability measure on $\mathscr{A}$ is specified by $P(\{\omega\}) = 1/4$ for all $\omega \in \Omega$. The probabilities of all other $2^4 - 4 = 12$ events can be computed from these four probabilities.

Now consider the projections π_1 and π_2 (see Def. 2.15) and the random variables X_1 and X_2 on the probability space $(\Omega, \mathscr{A}, P)$ defined by

$$X_i(\omega) = \begin{cases} 1, \text{ if } \pi_i(\omega) = h \\ 0, \text{ otherwise,} \end{cases}$$

for $i = 1,2$. The random variables X_1 and X_2 indicate (with 1) whether or not we flip *heads* in the first and second flip, respectively.

Now, X_1 and X_2 have identical distributions, However, neither $X_1 = X_2$ nor $X_1 \underset{P}{=} X_2$. If we consider $\big(\{0,1\}, \mathscr{P}(\{0,1\})\big)$ to be the value space of X_1 and X_2, then their distribution is

$$P_{X_i}(A') = \begin{cases} 1/2, & \text{if } A' = \{0\} \\ 1/2, & \text{if } A' = \{1\} \\ 1, & \text{if } A' = \{0,1\} \\ 0, & \text{if } A' = \emptyset, \end{cases}$$

for $i = 1,2$. Similarly, π_1 and π_2 have identical distributions, but neither $\pi_1 = \pi_2$ nor $\pi_1 \underset{P}{=} \pi_2$. If we choose $\big(\{h,t\}, \mathscr{P}(\{h,t\})\big)$ as the value space of π_1 and π_2, then the distribution of these two random variables is

$$P_{\pi_i}(A') = \begin{cases} 1/2, & \text{if } A' = \{t\} \\ 1/2, & \text{if } A' = \{h\} \\ 1, & \text{if } A' = \{h,t\} \\ 0, & \text{if } A' = \emptyset, \end{cases}$$

for $i = 1,2$. ◁

The following theorem provides an equivalent condition for P-equivalence of two compositions $f(X)$ and $g(X)$. (For a proof see SN-Th. 2.86.) Reading this theorem, note that the measurable functions $f, g: (\Omega'_X, \mathscr{A}'_X) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ are random variables on the probability space $(\Omega'_X, \mathscr{A}'_X, P_X)$.

Theorem 2.43 [P-Equivalence of Compositions]
Let X be a random variable on $(\Omega, \mathscr{A}, P)$, let $(\Omega'_X, \mathscr{A}'_X)$ denote its value space and P_X its distribution. Furthermore, let $f, g : (\Omega'_X, \mathscr{A}'_X) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ be measurable functions. Then:

$$f(X) \underset{P}{=} g(X) \quad \Leftrightarrow \quad f \underset{P_X}{=} g. \tag{2.34}$$

Remark 2.44 [Singleton of a Composition With a Positive Probability] Let the assumptions of Theorem 2.43 hold. If $\{X{=}x\} \in \mathscr{A}$ and $P(X{=}x) > 0$, then

$$f(X) \underset{P}{=} g(X) \quad \Rightarrow \quad f(x) \ = \ g(x). \tag{2.35}$$

◁

Remark 2.45 [A Sufficient Condition For P-Equivalence of Two Compositions] Let the assumptions of Theorem 2.43 hold and let $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ denote the *image of X*. Then:

$$\forall\, x \in X(\Omega): \ f(x) = g(x) \quad \Rightarrow \quad f \underset{P_X}{=} g \ \wedge \ f(X) \underset{P}{=} g(X)\,. \tag{2.36}$$

(see Exercise 2-4). ◁

Remark 2.46 [P-Equivalence and P^B-Equivalence] Let X and Y be random variables on a probability space $(\Omega, \mathscr{A}, P)$, let $B \in \mathscr{A}$, and $P(B) > 0$. Then

$$X \cdot 1_B \underset{P}{=} Y \cdot 1_B \quad \Leftrightarrow \quad X \cdot 1_B \underset{P^B}{=} Y \cdot 1_B \quad \Leftrightarrow \quad X \underset{P^B}{=} Y \tag{2.37}$$

(see Exercise 2-5). ◁

Example 2.47 [Indicator of a Null Set] Let $(\Omega, \mathscr{A}, P)$ be a probability space and $A \in \mathscr{A}$. If $P(A) = 0$, then

$$1_A \underset{P}{=} 0 \quad \text{and} \quad 1_{A^c} \ = \ 1 - 1_A \underset{P}{=} 1, \tag{2.38}$$

because $\forall\, \omega \in \Omega \setminus A$: $1_A(\omega) = 0$ and $P(A) = 0$ [see Prop. (2.28)]. ◁

2.4 Independence of Random Variables

Using the concept of a σ-algebra generated by a random variable (see Def. 2.12) we define independence of two random variables as follows:

Definition 2.48 [Independence of Two Random Variables]
Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$. Then independence of X and Y, denoted $X \perp\!\!\!\perp Y$, is defined by

$$P(A \cap B) \ = \ P(A) \cdot P(B), \quad \forall (A, B) \in \sigma(X) \times \sigma(Y)\,. \tag{2.39}$$

Box 2.1 Independence of Random Variables

Let X, Y, and Z be random variables on a probability space $(\Omega, \mathscr{A}, P)$ and let $(\Omega'_X, \mathscr{A}'_X)$, $(\Omega'_Y, \mathscr{A}'_Y)$, and $(\Omega'_Z, \mathscr{A}'_Z)$ denote their value spaces. Furthermore, let $X \perp\!\!\!\perp Y$ be defined by Equation (2.39) and $X \perp\!\!\!\perp Y \perp\!\!\!\perp Z$ by Equation (2.40). Then:

$$X \perp\!\!\!\perp Y \;\Leftrightarrow\; Y \perp\!\!\!\perp X \tag{i}$$

$$X \perp\!\!\!\perp Y \perp\!\!\!\perp Z \;\Rightarrow\; \big(X \perp\!\!\!\perp Y \wedge X \perp\!\!\!\perp Z \wedge Y \perp\!\!\!\perp Z\big) \tag{ii}$$

$$X \underset{P}{=} x,\; x \in \Omega'_X \;\Rightarrow\; X \perp\!\!\!\perp Y \tag{iii}$$

$$\big(X \perp\!\!\!\perp Y \wedge \sigma(Z) \subset \sigma(Y)\big) \;\Rightarrow\; \big(X \perp\!\!\!\perp Z \wedge X \perp\!\!\!\perp (Y, Z)\big) \tag{iv}$$

$$X \perp\!\!\!\perp (Y, Z) \;\Rightarrow\; \big(X \perp\!\!\!\perp Z \wedge X \perp\!\!\!\perp Y\big) \tag{v}$$

$$\big(X \perp\!\!\!\perp Y \wedge X \underset{P}{=} Z\big) \;\Rightarrow\; Z \perp\!\!\!\perp Y. \tag{vi}$$

Obviously, independence of two random variables is a symmetric concept, that is, $X \perp\!\!\!\perp Y$ is equivalent to $Y \perp\!\!\!\perp X$. This and other useful properties of independence of random variables are gathered in Box 2.1. They are special cases of the properties of conditional independence displayed in Box 6.1. (Proofs are found in the solution to SN-Exercise 16-3.)

Now we turn to independence of severable random variables. Again we use the notation $\sigma(X_i)$ for a σ-algebra generated by a random variable X_i (see Def. 2.12).

Definition 2.49 [Independence of Several Random Variables]
Let $X_1, \ldots, X_m$ be random variables on $(\Omega, \mathscr{A}, P)$. Then independence of $X_1, \ldots, X_m$, denoted $X_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_m$, is defined by

$$P(A_1 \cap \ldots \cap A_m) = P(A_1) \cdot \ldots \cdot P(A_m), \quad \forall (A_1, \ldots, A_m) \in \sigma(X_1) \times \ldots \times \sigma(X_m). \tag{2.40}$$

Remark 2.50 [An Implication of Independence of $X_1, \ldots, X_m$] If $(m - n) \geq 2$, then

$$X_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_m \;\Rightarrow\; X_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_{m-n}. \tag{2.41}$$

For example, *tripelwise independence* of random variables implies their *pairwise independence*. More precisely,

$$X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3 \;\Rightarrow\; X_1 \perp\!\!\!\perp X_2 \wedge X_1 \perp\!\!\!\perp X_3 \wedge X_2 \perp\!\!\!\perp X_3 \tag{2.42}$$

(see Exercise 2-7). Remember, the corresponding property does not hold for independence of events. That is, if $(\Omega, \mathscr{A}, P)$ is a probability space and $A_1, A_2, A_3 \in \mathscr{A}$, then

$$A_1 \perp\!\!\!\perp A_2 \perp\!\!\!\perp A_3 \;\not\Rightarrow\; A_1 \perp\!\!\!\perp A_2 \wedge A_1 \perp\!\!\!\perp A_3 \wedge A_2 \perp\!\!\!\perp A_3$$

(see Rem. 1.40). Finally, note that pairwise independence of X_1, X_2, X_3 does not imply their tripelwise independence, that is,

$$X_1 \perp\!\!\!\perp X_2 \wedge X_1 \perp\!\!\!\perp X_3 \wedge X_2 \perp\!\!\!\perp X_3 \;\not\Rightarrow\; X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3. \tag{2.43}$$

◁

2.5 Probability Function, Distribution Function and Density

In this section we introduce three functions each of which determines the distribution of a random variable, the *probability function*, the *distribution function*, and the *density*. The distribution of a random variable is defined for any random variable. The probability function presumes that the random variable is discrete. In this section, we confine ourselves to define the distribution function and a density only for a (uni-variate) *real-valued* random variable. (For the multivariate cases see SN-sections 5.7.2 and 5.7.4.) We start with the concept of the *probability function of a discrete random variable*, then turn to a real-valued random variable and its *distribution function*, and finally we introduce the *density of a continuous real-valued random variable*.

2.5.1 Probability Function of a Discrete Random Variable

The most important property of a discrete random variable is that its distribution can be specified by its *probability function*. This concept is introduced together with the concept of a *discrete random variable* in Definition 2.51. Reading this definition, remember, if Y is a random variable on $(\Omega, \mathscr{A}, P)$ with value space $(\Omega'_Y, \mathscr{A}'_Y)$, then the distribution P_Y of Y is a probability measure on $(\Omega'_Y, \mathscr{A}'_Y)$.

Definition 2.51 [Discrete Random Variable and its Probability Function] *Let Y be a random variable on $(\Omega, \mathscr{A}, P)$ with value space $(\Omega'_Y, \mathscr{A}'_Y)$ and assume that $\Omega'_0 \subset \Omega'_Y$ is finite or countable with, $\{y\} \in \mathscr{A}'_Y$ for all $y \in \Omega'_0$, and $P_Y(\Omega'_0) = 1$. Then Y and its distribution P_Y are called* discrete, *and the function $p_Y : \Omega'_Y \to [0,1]$ defined by*

$$p_Y(y) = \begin{cases} P_Y(\{y\}), & \text{if } y \in \Omega'_0, \\ 0, & \text{if } y \in \Omega'_Y \setminus \Omega'_0, \end{cases} \tag{2.44}$$

is called the probability function *of Y.*

Remark 2.52 [Probability Function vs. Distribution] The *distribution* P_Y is defined for *every* random variable, whereas the concept of a *probability function* p_Y only applies to a *discrete* random variable. While P_Y assigns probabilities to *subsets* of the co-domain Ω'_Y of Y, the probability function p_Y assigns a probability to each *element* y in Ω'_Y. Note that p_Y is a discrete real-valued random variable on the probability space $(\Omega'_Y, \mathscr{A}'_Y, P_Y)$. ◁

Example 2.53 [Discrete Random Variable and Image of Y] Let Y be a random variable on $(\Omega, \mathscr{A}, P)$ and assume that the image $Y(\Omega)$ of Y is finite or countable. Then, according to Definition 2.27,

$$P_Y\big(Y(\Omega)\big) = P\big(\{Y \in Y(\Omega)\}\big) = P\big(Y^{-1}(Y(\Omega))\big) = P(\Omega) = 1. \tag{2.45}$$

If $\{y\} \in \mathscr{A}'_Y$ for all $y \in Y(\Omega)$ holds as well, then, according to Definition 2.51, the random variable Y is discrete and $Y(\Omega)$ can play the role of Ω'_0. Furthermore, if we assume that the image $Y(\Omega)$ of Y is finite or countable, then $\big(Y(\Omega), \mathscr{P}(Y(\Omega))\big)$ can take the role of $(\Omega'_Y, \mathscr{A}'_Y)$ in Definition 2.51. ◁

Remark 2.54 [The Probability Function Determines the Distribution] Note that σ-additivity of the probability measure P_Y [see Def. 1.18 (c)] implies that P_Y is uniquely determined by the probability function p_Y [see Rule (x) in Box 1.1]. Vice versa, according to Definition 2.51, P_Y specifies p_Y. Hence, if X and Y are discrete random variables on $(\Omega, \mathscr{A}, P)$, both with value space $(\Omega', \mathscr{A}')$, then

$$p_X = p_Y \quad \Leftrightarrow \quad P_X = P_Y. \tag{2.46}$$

◁

2.5.2 *Distribution Function of a Real-Valued Random Variable*

If we consider a univariate *real-valued* random variable Y, that is, a random variable Y with value space $(\mathbb{R}, \mathscr{B})$, then the *distribution function* F_Y assigns to each $y \in \mathbb{R}$ the probability $P(Y \le y)$ of the event

$$\{Y \le y\} \; = \; \{\omega \in \Omega \colon Y(\omega) \le y\} \; = \; Y^{-1}((-\infty, y]), \tag{2.47}$$

that is, the event that Y takes on a value *smaller or equal* y. In this equation, $(-\infty, y]$ denotes a *half open interval*, namely the set of all real numbers that are smaller than or equal to y. Also note that $\{Y \le y\} \subset \Omega$, whereas $(-\infty, y] \in \mathscr{B}$.

Definition 2.55 [Distribution Function of a Real-Valued Random Variable]
Let Y denote a real-valued random variable on the probability space $(\Omega, \mathscr{A}, P)$. Then the (cumulative) distribution function $F_Y \colon \mathbb{R} \to [0, 1]$ of Y is defined by:

$$\forall\, y \in \mathbb{R}\colon \quad F_Y(y) \; := \; P_Y\big((-\infty, y]\big). \tag{2.48}$$

According to Definition 2.27 and because of Equation (2.47),

$$P_Y\big((-\infty, y]\big) \; = \; P\big(Y^{-1}((-\infty, y])\big) \; = \; P(Y \le y). \tag{2.49}$$

Hence, we can also write

$$\forall\, y \in \mathbb{R}\colon \quad F_Y(y) \; = \; P_Y\big((-\infty, y]\big) \; = \; P(Y \le y). \tag{2.50}$$

Example 2.56 [Distribution Function of an Indicator] Consider the probability space $(\Omega, \mathscr{A}, P)$, let $A \in \mathscr{A}$, 1_A its indicator, and $P(A) = .6$. Then $P(1_A = 0) = 1 - P(1_A = 1) = 1 - .6 = .4$ and

$$F_{1_A}(y) \; = \; \begin{cases} 0, & \text{if } y < 0 \\ .4, & \text{if } 0 \le y < 1 \\ 1, & \text{if } y \ge 1 \end{cases} \tag{2.51}$$

(see Fig. 2.1). Note that the distribution function F_{1_A} assigns a value to *all* real numbers $y \in \mathbb{R}$. ◁

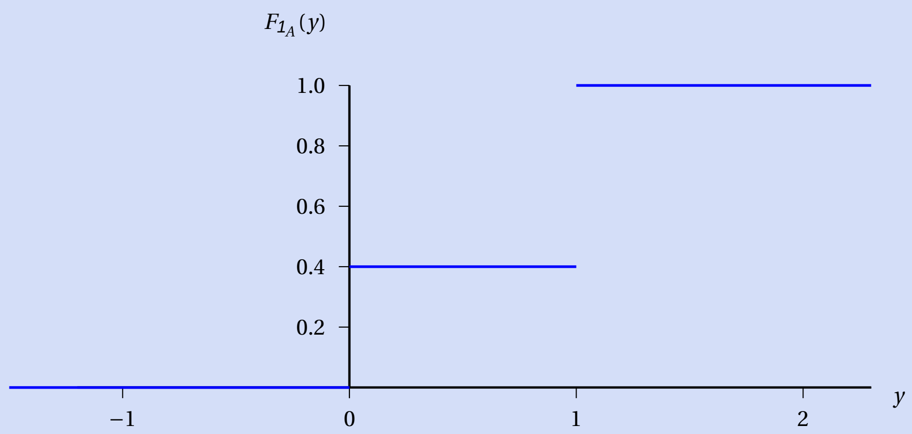

Figure 2.1. Distribution function F_{1_A} of the indicator 1_A of an event A with $P(A) = .6$.

Remark 2.57 [Probabilities of Intervals] Equation (2.50) implies that we can compute the probability $P(a < Y \leq b)$ of Y taking a value in the half open interval $(a, b]$ by

$$P(a < Y \leq b) \;=\; F_Y(b) - F_Y(a), \quad \text{if } a < b, \tag{2.52}$$

because

$$P(a < Y \leq b) \;=\; P_Y\big((-\infty, b] \setminus (-\infty, a]\big) \;=\; P_Y\big((-\infty, b]\big) \;-\; P_Y\big((-\infty, a]\big)$$

[see Box 1.1 (vi)]. ◁

Remark 2.58 [The Distribution Function Determines the Distribution] Every random variable Y has a distribution P_Y. In contrast, the distribution function F_Y is defined only for uni- and multivariate real-valued random variables (see SN-Def. 5.89). The distribution function uniquely determines the distribution P_Y of a real-valued random variable (see SN-Rem. 5.83). ◁

2.5.3 Density of a Continuous Real-Valued Random Variable

In this section we introduce the concept of a density f_Y of a continuous univariate real-valued random variable. In Definition 2.59 we refer to the *Lebesgue measure* λ on the measurable space $(\mathbb{R}, \mathscr{B})$, which is uniquely defined by

$$\lambda\big((a, b]\big) \;=\; b - a, \quad \forall\, a, b \in \mathbb{R}, \quad \text{with } a < b \tag{2.53}$$

(see SN-sect. 1.4.2). The number $\lambda\big((a, b]\big)$ may be called the *length* of the half open interval $(a, b]$.

Definition 2.59 [Continuous Real-Valued Random Variable and Its Density]
Let $Y: (\Omega, \mathscr{A}) \to (\mathbb{R}, \mathscr{B})$ *be a real-valued random variable on the probability space* $(\Omega, \mathscr{A}, P)$ *with distribution* P_Y *and let* λ *denote the Lebesgue measure on* $(\mathbb{R}, \mathscr{B})$. *We call Y continuous, if there is a nonnegative measurable function* $f_Y : \mathbb{R} \to \mathbb{R}$ *such that*

$$P_Y(B) = \int_B f_Y \, d\lambda, \quad \forall B \in \mathscr{B}. \tag{2.54}$$

A function f_Y *satisfying Equation (2.54) is called a (probability) density of Y (with respect to* λ*).*

Remark 2.60 [λ-Equivalence of Two Densities] A density f_Y is not uniquely defined. However, two functions satisfying (2.54) are *λ-equivalent*. That is, if f_Y and f_Y^* satisfy Equation (2.54), then there is a $B \in \mathscr{B}$ such that $\lambda(B) = 0$ and $f_Y(y) = f_Y^*(y)$ for all $y \in \mathbb{R} \setminus B$ (see SN-Def. 2.68). ◁

Remark 2.61 [Distribution Function and Density] Also note that Equation (2.54) is equivalent to

$$F_Y(y) = \int_{(-\infty, y]} f_Y \, d\lambda, \quad \forall\, y \in \mathbb{R}, \tag{2.55}$$

because $(-\infty, y] \in \mathscr{B}$ and

$$F_Y(y) = P_Y\big((-\infty, y]\big) = \int 1_{(-\infty, y]} \, dP_Y = \int_{(-\infty, y]} f_Y \, d\lambda, \quad \forall\, y \in \mathbb{R}, \tag{2.56}$$

(see SN-Th. 3.67). ◁

The following corollary shows the relationship between the distribution function F_Y of Y and a density of Y in terms of the Riemann integral.

Corollary 2.62 [Riemann Integral of the Density]
If f_Y *is a density of a real-valued random variable Y on* $(\Omega, \mathscr{A}, P)$ *and* f_Y *is Riemann integrable, then*

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t) \, dt, \quad \forall\, y \in \mathbb{R}. \tag{2.57}$$

Remark 2.63 [Interpretation of Densities] Note that the term $f_Y(t)$ in Equation (2.57) is not a probability. Instead it is a value of the density for $t \in \mathbb{R}$. Figure 2.2 shows the relationship between the density of a standard normal distribution and its distribution function. ◁

Example 2.64 [Density of a Standard Normal Distribution] The function $f_Y : \mathbb{R} \to \mathbb{R}$ defined by

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{y^2}{2}\right), \quad \forall\, y \in \mathbb{R}, \tag{2.58}$$

is a density. It is a density of the *standard normal distribution*. Its graph is depicted in Figure 2.2, together with its distribution function F_Y. ◁

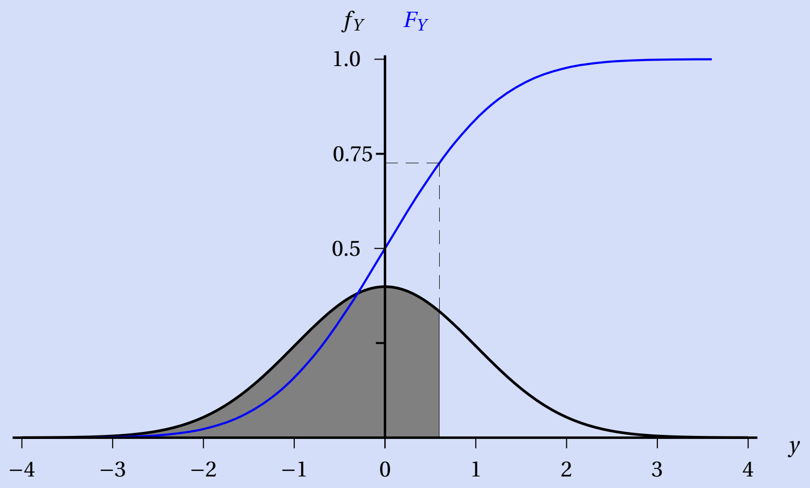

Figure 2.2. Density and distribution function of the standard normal distribution

Remark 2.65 [Probability $P(a<Y\le b)$] The probability that Y takes on a value in the interval $(a,b]$ can be computed by

$$P(a<Y\le b) \;=\; F_Y(b)-F_Y(a) \;=\; \int_a^b f_Y(y)\,dy, \quad \text{if } a<b, \tag{2.59}$$

using the density f_Y, provided that f_Y exists and that it is Riemann integrable. This probability can be represented by the area between the density and the y-axis above the half open interval $(a,b]$ (see Fig. 2.3). ◁

Remark 2.66 [Continuity of Y Implies $P(Y=y)=0$] Consider a continuous random variable Y. Because $\lambda(\{y\})=0$ [see Eq. (2.53)], Definition 2.59 implies

$$P(Y=y) \;=\; 0, \quad \forall y\in\mathbb{R}. \tag{2.60}$$

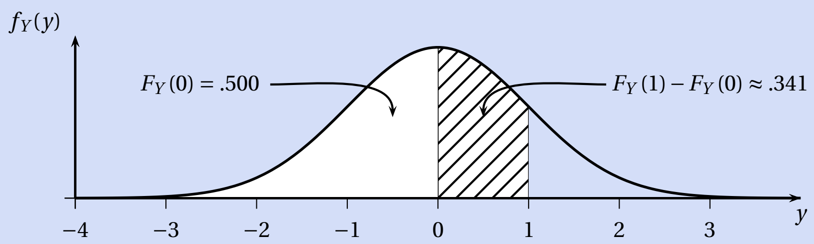

Figure 2.3. Integral of a density for the intervals $(-\infty,0]$ and $(0,1]$.

Hence, if Y is continuous, then additivity of P yields, for all $a, b \in \mathbb{R}$, $a < b$,

$$P(a < Y \le b) = P(a \le Y \le b) = P(a \le Y < b) = P(a < Y < b). \tag{2.61}$$

$\triangleleft$

Some examples of continuous random variables and their densities, such as the densities of normal distributions, central χ^2-distributions, central t-distributions, and central F-distributions are found in SN-section 8.2.

2.6 Summary and Conclusions

In this chapter we introduced the concept of a measurable mapping Y on a measurable space $(\Omega, \mathscr{A})$ with value space $(\Omega'_Y, \mathscr{A}'_Y)$. A random variable on a probability space $(\Omega, \mathscr{A}, P)$ is defined as a measurable mapping on $(\Omega, \mathscr{A})$. The only distinction between the two concepts is that there is a probability measure P on $\mathscr{A}$ if Y is a random variable.

The crucial property of a measurable mapping and a random variable Y is that all inverse images $Y^{-1}(A')$, $A' \in \mathscr{A}'_Y$, are elements of $\mathscr{A}$. If Y is a random variable, then this guarantees that all these inverse images have probabilities assigned by P, which is used to define the *distribution* P_Y of the random variable Y. Note that random variables do not have to be numerical, that is, their values do not have to be real numbers or one of the values ∞ or $-\infty$. Instead, they can take on values in any set Ω'_Y, provided that there is a σ-algebra $\mathscr{A}'_Y$ on this set and $Y^{-1}(A') \in \mathscr{A}$, for all $A' \in \mathscr{A}'_Y$.

We also introduced special mappings such as an *indicator* of an event, a *projection*, a *multivariate mapping*, and a *composition*. The most important properties of these special mappings related to measurability were treated. Box 2.2 provides a glossary of these and other concepts related to random variables such as *P-equivalence of two random variables*, the *probability function of a disrete random variable*, the *distribution function of a real-valued random variable*, and the *density of a real-valued continuous random variable*.

2.7 Exercises

▷ **Exercise 2-1** Consider the random variable X in Example 2.10. Instead of $(\mathbb{R}, \mathscr{B})$, choose $\big(\{0,1,2\}, \mathscr{P}(\{0,1,2\})\big)$ as the value space of X, where $\mathscr{P}(\{0,1,2\})\big)$ denotes the power set of $\{0,1,2\}$. Write down all inverse images $X^{-1}(B)$, $B \in \mathscr{P}(\{0,1,2\})$.

▷ **Exercise 2-2** Prove the propositions of Example 2.14.

▷ **Exercise 2-3** Consider Example 2.26 and, using Lemma 2.24, show that the random variable $X \cdot Y$ is (X, Y)-measurable.

▷ **Exercise 2-4** Prove the proposition of Remark 2.45.

▷ **Exercise 2-5** Prove the proposition of Remark 2.46.

▷ **Exercise 2-6** Compute the probability $P_{X,Y}(B)$ of $B = \{(0,0),(0,1)\}$ via Rule (x) of Box 1.1, using the probabilities computed in Example 2.30.

▷ **Exercise 2-7** Assume that X_1, X_2, X_3 are random variables on the probability space $(\Omega, \mathscr{A}, P)$ and prove $X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3 \;\Rightarrow\; X_1 \perp\!\!\!\perp X_2$.

Box 2.2 Glossary of new concepts

$Y^{-1}(A')$	*The inverse image of A' under Y.* It is defined by $$Y^{-1}(A') := \{\omega \in \Omega : Y(\omega) \in A'\}.$$
$\{Y \in A'\}$	An alternative notation of $Y^{-1}(A')$.
$Y: (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$	*Measurable mapping.* It is a mapping on Ω with values in the set Ω'_Y satisfying $\{Y \in A'\} \in \mathscr{A}$, for all $A' \in \mathscr{A}'_Y$.
Y	*Random variable on* $(\Omega, \mathscr{A}, P)$ *with value space* $(\Omega'_Y, \mathscr{A}'_Y)$. It is defined as a measurable mapping $Y: (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$. If $(\Omega'_Y, \mathscr{A}'_Y) = (\mathbb{R}, \mathscr{B})$, then Y is called *real-valued*. If $(\Omega'_Y, \mathscr{A}'_Y) = (\overline{\mathbb{R}}, \overline{\mathscr{B}})$, then Y is called *numerical*.
$Y^{-1}(\mathscr{A}'_Y)$	σ-*algebra generated by* $Y: (\Omega, \mathscr{A}) \to (\Omega'_Y, \mathscr{A}'_Y)$. It is the set of all events $\{Y \in A'\}$, $A' \in \mathscr{A}'_Y$.
$\sigma(Y)$	An alternative notation of $Y^{-1}(\mathscr{A}'_Y)$.
$X \underset{P}{=} Y$	*X and Y are P-equivalent.* It is defined for random variables X, Y on $(\Omega, \mathscr{A}, P)$ by the existence of an $A \in \mathscr{A}$ such that $P(A) = 0$ and $X(\omega) = Y(\omega)$, $\forall \omega \in \Omega \setminus A$.
$X_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_m$	*Independence of random variables $X_1, \ldots, X_m$ on* $(\Omega, \mathscr{A}, P)$. It is defined by $$P\big(\bigcap_{i=1}^m A_i\big) = \prod_{i=1}^m P(A_i), \quad \forall (A_1, \ldots, A_m) \in \sigma(X_1) \times \ldots \times \sigma(X_m).$$
P_Y	*Distribution* of a random variable Y on $(\Omega, \mathscr{A}, P)$ with value space $(\Omega'_Y, \mathscr{A}'_Y)$. It is a probability measure on $(\Omega'_Y, \mathscr{A}'_Y)$ defined by $P_Y(A') = P(\{Y \in A'\})$, $\forall A' \in \mathscr{A}'_Y$.
p_Y	*Probability function of a discrete random variable* Y on a probability space $(\Omega, \mathscr{A}, P)$ with value space $(\Omega'_Y, \mathscr{A}'_Y)$. Assume that $\Omega'_0 \subset \Omega'_Y$ is finite or countable with $P_Y(\Omega'_0) = 1$ and $\{y\} \in \mathscr{A}'_Y$ for all $y \in \Omega'_0$. Then Y is called *discrete* and its probability function $p_Y : \Omega'_Y \to [0,1]$ is defined by $$p_Y(y) = \begin{cases} P_Y(\{y\}), & \text{if } y \in \Omega'_0, \\ 0, & \text{if } y \in \Omega'_Y \setminus \Omega'_0. \end{cases}$$
F_Y	*Distribution function* of a real-valued random variable Y on $(\Omega, \mathscr{A}, P)$. It is a function on the set $\mathbb{R}$ of real numbers defined by: $$F_Y(y) := P_Y\big((\infty, y]\big) = P(Y \le y), \quad \forall y \in \mathbb{R}.$$
λ	*Lebesgue measure.* It is a function on the Borel σ-algebra $\mathscr{B}$ satisfying $$\lambda\big((a,b]\big) = b - a, \quad \forall a, b \in \mathbb{R}, \quad \text{with } a < b.$$
f_Y	*Density* of a continuous real-valued random variable Y on $(\Omega, \mathscr{A}, P)$. If there is a nonnegative function on $\mathbb{R}$ that is integrable with respect to λ and satisfies $$P_Y(B) = \int_B f_Y \, d\lambda, \quad \forall B \in \mathscr{B},$$ then Y is called *continuous* and f_Y a (probability) *density of* Y.

Solutions

▷ **Solution 2-1** These inverse images are identical to those listed in Equation (2.9).

▷ **Solution 2-2** The set $\{X{=}x\}$ is an element of $\mathscr{A}$ because we assumed $\{x\} \in \mathscr{A}'$, $\{X{=}x\} = X^{-1}(\{x\})$ and a measurable mapping is defined such that all inverse images $X^{-1}(A')$, $A' \in \mathscr{A}'$, are elements of $\mathscr{A}$. Furthermore, by definition, $\sigma(X) = \{X^{-1}(A') : A' \in \mathscr{A}'\}$. Because we assume $\{x\} \in \mathscr{A}'$ and $\mathscr{A}'$ is a σ-algebra, this implies that all inverse images

$$\begin{aligned} 1_{X=x}^{-1}(\{1\}) &= X^{-1}(\{x\}) = \{\omega \in \Omega\colon X(\omega) \in \{x\}\}, \\ 1_{X=x}^{-1}(\{0\}) &= X^{-1}(\{x\}^c) = \{\omega \in \Omega\colon X(\omega) \in \{x\}^c\}, \\ 1_{X=x}^{-1}(\{0,1\}) &= X^{-1}(\Omega') = \{\omega \in \Omega\colon X(\omega) \in \Omega'\}, \\ 1_{X=x}^{-1}(\emptyset) &= X^{-1}(\emptyset) = \{\omega \in \Omega\colon X(\omega) \in \emptyset\} \end{aligned}$$

are elements of $\sigma(X)$. Hence, $\sigma(1_{X=x}) \subset \sigma(X)$, if $\{x\} \in \mathscr{A}'$. No other assumption about the σ-algebra $\mathscr{A}'$ is necessary for proving this proposition.

▷ **Solution 2-3** We have to show that the composition $X \cdot Y = g(X,Y)$ is (X,Y)-measurable. In Example 2.26 we already specified the function $g\colon \{0,1\}\times\{0,1\} \to \overline{\mathbb{R}}$ [see Eq. (2.23)], chose the value space $\big(\{0,1\}, \mathscr{P}(\{0,1\})\big)$ of X and Y and the value space $\big(\{0,1\}\times\{0,1\}, \mathscr{P}(\{0,1\}\times\{0,1\})\big)$ of the bivariate random variable (X,Y). According to Lemma 2.24, we have to show that g is $\big(\mathscr{P}(\{0,1\}\times\{0,1\}), \overline{\mathscr{B}}\big)$-measurable. However, this is trivial, because all inverse images $g^{-1}(B)$, $B \in \overline{\mathscr{B}}$, must be elements of the power set $\mathscr{P}(\{0,1\}\times\{0,1\})$, which, by its definition, contains all subsets of $\{0,1\}\times\{0,1\}$ as elements.

▷ **Solution 2-4**

$$\begin{aligned} & \forall\, x \in X(\Omega)\colon\ f(x) = g(x) \\ \Rightarrow\ & \forall\, x \in X(\Omega)\colon\ f(x) = g(x) \,\wedge\, P_X\big(\Omega'_X \setminus X(\Omega)\big) = 0 \\ \Rightarrow\ & \exists\, A' \in \mathscr{A}'_X\colon\ \big(\forall\, x \in \Omega'_X \setminus A'\colon\ f(x) = g(x) \,\wedge\, P(A') = 0\big) && [A' = \Omega'_X \setminus X(\Omega)] \\ \Rightarrow\ & f(X) \underset{P}{=} g(X) && [\text{Def. 2.35}] \\ \Rightarrow\ & f \underset{P_X}{=} g. && [\text{Th. 2.43}] \end{aligned}$$

▷ **Solution 2-5** We show $X\cdot 1_B \underset{P}{=} Y\cdot 1_B \;\Rightarrow\; X\cdot 1_B \underset{P^B}{=} Y\cdot 1_B$. Define $A := \{\omega \in \Omega\colon X\cdot 1_B(\omega) \neq Y\cdot 1_B(\omega)\}$.

$$\begin{aligned} X\cdot 1_B \underset{P}{=} Y\cdot 1_B \quad &\Rightarrow\quad P(A) = 0 && [(2.28)] \\ &\Rightarrow\quad P(A\cap B) = 0 && [\text{Box 1.1 (v)}] \\ &\Rightarrow\quad P(A\cap B)/P(B) = 0 && [P(B) > 0] \\ &\Rightarrow\quad P(A\,|\,B) = 0 && [(1.10)] \\ &\Rightarrow\quad P^B(A) = 0 && [(1.21)] \\ &\Rightarrow\quad X\cdot 1_B \underset{P^B}{=} Y\cdot 1_B. && [(2.28)] \end{aligned}$$

Now we show $X\cdot 1_B \underset{P^B}{=} Y\cdot 1_B \;\Rightarrow\; X \underset{P^B}{=} Y$. Suppose $X \underset{P^B}{=} Y$ does not hold. Then there is an element $C \in \mathscr{A}$ such that $P^B(C) > 0$, where $C := \{\omega \in \Omega\colon X(\omega) \neq Y(\omega)\}$. This implies

$$\begin{aligned} P^B(B\cap C) &= P(B\cap C\,|\,B) && [(1.21)] \\ &= \frac{P(B\cap C\cap B)}{P(B)} = \frac{P(B\cap C)}{P(B)} && [(1.10),\ B\cap C\cap B = B\cap C] \\ &= P(C\,|\,B) = P^B(C) && [(1.21), (1.10)] \\ &> 0. && [P^B(C) > 0] \end{aligned}$$

Now

$$B \cap C = \big\{\omega \in \Omega\colon X(\omega) \neq Y(\omega) \wedge \omega \in B\big\} = \big\{\omega \in \Omega\colon X \cdot 1_B(\omega) \neq Y \cdot 1_B(\omega)\big\}.$$

Because $(B \cap C) \in \mathscr{A}$ and $P^B(B \cap C) > 0$, this is a contradiction to $X \cdot 1_B \underset{P^B}{=} Y \cdot 1_B$ [see Prop. (2.30)].

We complete the proof showing $X \underset{P^B}{=} Y \;\Rightarrow\; X \cdot 1_B \underset{P}{=} Y \cdot 1_B$. Again the proof is by contraposition.

$$\begin{aligned}
\neg(X \cdot 1_B \underset{P}{=} Y \cdot 1_B) \quad &\Rightarrow \quad P(B \cap C) > 0 && [\text{def. of } B \cap C,\ (2.28)]\\
&\Rightarrow \quad \frac{P(B \cap C)}{P(B)} > 0 && [P(B) > 0]\\
&\Rightarrow \quad P(C \,|\, B) > 0 && [(1.10)]\\
&\Rightarrow \quad P^B(C) > 0 && [(1.21)]\\
&\Rightarrow \quad \neg(X \underset{P^B}{=} Y).
\end{aligned}$$

▷ **Solution 2-6** We consider the probability space $\big(\Omega', \mathscr{P}(\Omega'), P_{X,Y}\big)$, where $\Omega' = \{0,1\} \times \{0,1\} = \big\{(0,0),(0,1),(1,0),(1,1)\big\}$ satisfies the requirements of the set Ω_0 made in Rule (x) of Box 1.1. We also consider the intersection $B \cap \Omega' = B = \{(0,0),(0,1)\}$, its probability $P_{X,Y}(B)$, and the four singletons $\{\omega'\} \in \mathscr{P}(\Omega')$, namely $\{(0,0)\}$, $\{(0,1)\}$, $\{(1,0)\}$, and $\{(1,1)\}$. Hence, according to Rule (x) of Box 1.1,

$$\begin{aligned}
P_{X,Y}(B) &= P_{X,Y}\big(\{(0,0),(0,1)\}\big)\\
&= \sum_{\omega' \in B \cap \Omega'} P_{X,Y}(\{\omega'\}) = \sum_{\omega' \in B} P_{X,Y}(\{\omega'\})\\
&= P_{X,Y}\big(\{(0,0)\}\big) + P_{X,Y}\big(\{(0,1)\}\big)\\
&= .33 + .27 = .6.
\end{aligned}$$

▷ **Solution 2-7**

$$\begin{aligned}
& X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp X_3\\
\Leftrightarrow\quad & P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3), \quad \forall (A_1, A_2, A_3) \in \sigma(X_1) \times \sigma(X_2) \times \sigma(X_3) && [\text{Def. } 2.49]\\
\Rightarrow\quad & P(A_1 \cap A_2 \cap \Omega) = P(A_1) \cdot P(A_2) \cdot P(\Omega), \quad \forall (A_1, A_2) \in \sigma(X_1) \times \sigma(X_2) && [\Omega \in \sigma(X_3)]\\
\Leftrightarrow\quad & P(A_1 \cap A_2) = P(A_1) \cdot P(A_2), \quad \forall (A_1, A_2) \in \sigma(X_1) \times \sigma(X_2) && [A_1 \cap A_2 \cap \Omega = A_1 \cap A_2,\ P(\Omega) = 1]\\
\Leftrightarrow\quad & X_1 \perp\!\!\!\perp X_2\,. && [\text{Def. } 2.49]
\end{aligned}$$

Chapter 3
Expectation, Variance, Covariance, and Correlation

This chapter is devoted to the concepts of *expectation, variance, covariance, correlation,* and *linear quasi-regression,* which is the kind of dependence that is quantified by a covariance and a correlation. We will see that *variances, covariances,* and *correlations* are expectations of special random variables. All these quantities describe important properties of numerical random variables, although, in general, they do not determine the complete distribution.

3.1 Expectation

3.1.1 General Definition

In the following definition we use the concept of the *integral of a numerical random variable with respect to P*, which is well-known from measure theory (for an introduction, see SN-ch. 3). We presume that the numerical random variable Y on the probability space $(\Omega, \mathscr{A}, P)$ is *quasi-integrable* with respect to the measure P. That is, we presume that $\int Y^+ dP$ or $\int Y^- dP$ are finite, where

$$Y^+ := \max\big(Y(\omega), 0\big), \quad \forall\, \omega \in \Omega \tag{3.1}$$

and

$$Y^- := -\min\big(Y(\omega), 0\big), \quad \forall\, \omega \in \Omega \tag{3.2}$$

denote the *positive* and *negative parts of Y*, respectively (see SN-Rem. 2.62 and SN-Def. 3.28).

Definition 3.1 [Expectation of a Numerical Random Variable]
Let Y be a numerical random variable on $(\Omega, \mathscr{A}, P)$ *that is quasi-integrable with respect to P. Then we define*

$$E(Y) := \int Y\, dP, \tag{3.3}$$

call it the expectation of Y (with respect to P), and say that it exists.

Remark 3.2 [Existence of the Expectation] Note that $E(Y)$ can be infinite. Furthermore, if $E(Y)$ *exists*, then we also say that Y is a random variable *with expectation* $E(Y)$. If Y is not quasi-integrable with respect to P and therefore also not P-integrable, then we say that the expectation of Y with respect to P does *not exist*. ◁

Remark 3.3 [Notation and Synonymous Terms] Synonymous terms for *expectation* are *expectation value*, *theoretical mean*, *true mean*, and *population mean*. The reference to the measure P is usually omitted if the context is unambiguous. ◁

Remark 3.4 [An Alternative Notation] An alternative notation for the integral $\int Y\,dP$ is $\int Y(\omega)\,P(d\omega)$, which explicitly uses the values $Y(\omega)$ of Y (see SN-Rem. 3.32). Hence,

$$E(Y) = \int Y\,dP = \int Y(\omega)\,P(d\omega). \tag{3.4}$$

This notation conveys the idea that the values $Y(\omega)$ of Y are weighted by the probability of $d\omega$, symbolizing the length of an infinitesimal interval between two elements in $\mathbb{R}$. ◁

Remark 3.5 [Numerical Random Variable With a Finite Number of Values] Assume that Y has only a finite number of different values $y_1, \dots, y_n \in \mathbb{R}$, that is, assume that the *image* $Y(\Omega) = \{Y(\omega) \colon \omega \in \Omega\}$ *of* Y is the set $\{y_1, \dots, y_n\}$. Then the expectation $E(Y)$ exists and

$$E(Y) = \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i), \tag{3.5}$$

using the notation $P(Y{=}y_i) = P(\{Y{=}y\})$ [see Eq. (2.3)]. ◁

Remark 3.6 [Numerical Random Variable With a Countable Number of Values] Assume that Y has a countable number of different values $y_1, y_2, \ldots \in \mathbb{R}$, that is, assume that the image $Y(\Omega) = \{Y(\omega) \colon \omega \in \Omega\}$ of Y is the set $\{y_1, y_2, \ldots\}$. Then the expectation $E(Y)$ exists and

$$E(Y) = \sum_{i=1}^{\infty} y_i \cdot P(Y{=}y_i) := \lim_{n\to\infty} \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i). \tag{3.6}$$

Equations (3.5) and (3.6) can be written in a single equation as

$$E(Y) = \sum_{y \in Y(\Omega)} y \cdot P(Y{=}y). \tag{3.7}$$

If $Y(\Omega) = \{y_1, \dots, y_n\}$ is finite, then Equation (3.5) applies. If $Y(\Omega) = \{y_1, y_2, \dots\}$ is not finite but countable, then Equation (3.6) applies. ◁

Example 3.7 [Expectation of an Indicator] If $(\Omega, \mathscr{A}, P)$ is a probability space and 1_A is the indicator of $A \in \mathscr{A}$, then Equations (2.5) and (3.5) yield

$$E(1_A) = 0 \cdot P(1_A{=}0) + 1 \cdot P(1_A{=}1) = P(1_A{=}1) = P(A), \tag{3.8}$$

because $\{1_A{=}1\} = \{\omega \in \Omega \colon 1_A(\omega){=}1\} = A$. Considering the event $\{Y{=}y\} = \{\omega \in \Omega \colon Y(\omega){=}y\}$ and using the notation $1_{Y=y} := 1_{\{Y=y\}}$, this implies

$$E(1_{Y=y}) = P(Y{=}y). \tag{3.9}$$

Hence, the expectation of the indicator $1_{Y=y}$ is the probability $P(Y{=}y)$. ◁

Example 3.8 [Joe and Ann With Randomized Assignment] Consider the random experiment displayed in Table 1.2 and the random variables X and Y. Applying Equation (3.8) to (the indicator variable) X yields:

$$
\begin{aligned}
E(X) &= P(X{=}1) \\
&= P\big(\{(Joe, yes, -), (Joe, yes, +), (Ann, yes, -), (Ann, yes, +)\}\big) \\
&= P\big(\{(Joe, yes, -)\}\big) + P\big(\{(Joe, yes, +)\}\big) + P\big(\{(Ann, yes, -)\}\big) + P\big(\{(Ann, yes, +)\}\big) \\
&= .04 + .16 + .12 + .08 \;=\; .4.
\end{aligned}
$$

Similarly, for Y we obtain

$$
\begin{aligned}
E(Y) &= P(Y{=}1) \\
&= P\big(\{(Joe, no, +), (Joe, yes, +), (Ann, no, +), (Ann, yes, +)\}\big) \\
&= P\big(\{(Joe, no, +)\}\big) + P\big(\{(Joe, yes, +)\}\big) + P\big(\{(Ann, no, +)\}\big) + P\big(\{(Ann, yes, +)\}\big) \\
&= .21 + .16 + .06 + .08 \;=\; .51.
\end{aligned}
$$

◁

3.1.2 Computing the Expectation Using a Density

According to the following theorem, the expectation of a continuous real-valued random variable Y can also be computed using the *Riemann integral* and a density f_Y of Y with respect to the Lebesgue measure [see Eq. (2.53)]. (For a proof see SN-Th. 6.11.)

Theorem 3.9 [Expectation of a Continuous Real-Valued Random Variable]
Let Y be a continuous real-valued random variable on $(\Omega, \mathscr{A}, P)$ with expectation $E(Y)$ and a density f_Y that is Riemann integrable. Then

$$E(Y) \;=\; \int_{-\infty}^{\infty} y\ f_Y(y)\, dy. \tag{3.10}$$

Example 3.10 [Expectation of a Random Variable With a Standard Normal Distribution]
Assume that Y has a standard normal distribution, that is, assume that the function f_Y specified by Equation (2.58) is a density of Y. Then $E(Y) = \int_{-\infty}^{\infty} y\ f_Y(y)\, dy = 0$. (For a proof see SN-Exercise 8-4.) ◁

3.1.3 Transformation Theorem

Let Y be a random variable on $(\Omega, \mathscr{A}, P)$ with value space $(\Omega'_Y, \mathscr{A}'_Y)$ and let g be a random variable on $(\Omega'_Y, \mathscr{A}'_Y, P_Y)$ with value space $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$. Remember, the distribution P_Y of Y is a probability measure on $(\Omega'_Y, \mathscr{A}'_Y)$ (see Rem. 2.28). In the sequel, $E_Y(g)$ denotes the *expectation of g with respect to the probability measure P_Y*.

Theorem 3.11 applies when we consider the expectation of a *composition* $g \circ Y = g(Y)$ of a mapping Y [see Eq. (2.21)] and a numerical function g, or if we consider the expectation of g with respect to the distribution P_Y (see Def. 2.27). (For a proof see SN-Theorem 3.57.)

Theorem 3.11 [Transformation Theorem]
Let Y be a random variable on $(\Omega, \mathscr{A}, P)$ with value space $(\Omega'_Y, \mathscr{A}'_Y)$ and let $g\colon \Omega'_Y \to \overline{\mathbb{R}}$ be a random variable on $(\Omega'_Y, \mathscr{A}'_Y, P_Y)$ with value space $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$. If g is nonnegative or has a finite expectation $E_Y(g)$, then

$$E\big(g(Y)\big) = \int g(Y)\, dP = \int g\, dP_Y = E_Y(g). \tag{3.11}$$

The virtue of Equation (3.11) is that we do not have to know the distribution of $g(Y)$. Instead, the distribution of Y suffices in order to compute the expectation of the composition $g(Y)$.

Remark 3.12 [An Alternative Notation] Using the alternative notation $\int g(y)\, P_Y(dy)$ for the integral $\int g\, dP_Y$ (see Rem. 3.3),

$$E_Y(g) = \int g\, dP_Y = \int g(y)\, P_Y(dy). \tag{3.12}$$

◁

Remark 3.13 [Finite Number of Values] If Y takes on only a finite number of different values $y_1, \dots, y_n$, then Equation (3.11) simplifies to

$$E\big(g(Y)\big) = \sum_{i=1}^{n} g(y_i) \cdot P(Y = y_i) = \sum_{i=1}^{n} g(y_i) \cdot P_Y(\{y_i\}) = E_Y(g), \tag{3.13}$$

where $P(Y = y_i) = P_Y(\{y_i\})$, $i = 1, \dots, n$.

If we consider a bivariate random variable (X, Y) that has a finite number of different values (x_i, y_j), $i = 1, \dots, n$, $j = 1, \dots, m$, then Equation (3.11) simplifies to

$$\begin{aligned} E\big(g(X,Y)\big) &= \sum_{i=1}^{n} \sum_{j=1}^{m} g(x_i, y_j) \cdot P(X = x_i, Y = y_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} g(x_i, y_j) \cdot P_{X,Y}(\{x_i, y_j\}) \\ &= E_{X,Y}(g), \end{aligned} \tag{3.14}$$

where $E_{(X,Y)}(g)$ denotes the expectation of g with respect to the joint distribution $P_{X,Y}$ of X and Y. ◁

Remark 3.14 [A Special Case] If we consider the special case in which Y is numerical and g is the identity function $id\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, defined by $id(y) = y$, for all $y \in \overline{\mathbb{R}}$, then $id(Y) = Y$ and Equations (3.11) yield

$$E(Y) = \int Y\, dP = \int y\, dP_Y. \tag{3.15}$$

If Y is discrete with a finite number of different values $y_1, \dots, y_n \in \mathbb{R}$, then this equation simplifies to

Box 3.1 Rules of Computation for Expectations

Let Y be a numerical random variable on $(\Omega, \mathscr{A}, P)$ and let the expectation $E(Y)$ exist. Then:

$$E(Y) = \alpha, \quad \text{if } Y \underset{P}{=} \alpha,\ \alpha \in \mathbb{R} \tag{i}$$

$$E(\alpha + Y) = \alpha + E(Y), \quad \text{if } \alpha \in \mathbb{R} \tag{ii}$$

$$E(\alpha \cdot Y) = \alpha \cdot E(Y), \quad \text{if } \alpha \in \mathbb{R}. \tag{iii}$$

Additionally, let $A \in \mathscr{A}$. Then:

$$E(1_A \cdot Y) = 0, \quad \text{if } P(A) = 0. \tag{iv}$$

Let Y_1 and Y_2 be numerical random variables on $(\Omega, \mathscr{A}, P)$ that are nonnegative or with finite expectations. Then:

$$Y_1 \underset{P}{=} Y_2 \quad \Rightarrow \quad E(Y_1) = E(Y_2). \tag{v}$$

$$Y_1 \underset{P}{=} Y_2 \quad \Leftrightarrow \quad \forall A \in \mathscr{A}:\ E(1_A\, Y_1) = E(1_A\, Y_2). \tag{vi}$$

For $i = 1, \dots, m$, let Y_i be numerical random variables on $(\Omega, \mathscr{A}, P)$ that are nonnegative or real-valued with finite expectations $E(Y_i)$. Then:

$$E\Big(\sum_{i=1}^{m} Y_i\Big) = \sum_{i=1}^{m} E(Y_i). \tag{vii}$$

$$Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_m \quad \Rightarrow \quad E\Big(\prod_{i=1}^{m} Y_i\Big) = \prod_{i=1}^{m} E(Y_i). \tag{viii}$$

For $i = 1, \dots, m$, let $\alpha_i \in \mathbb{R}$ and Y_i be real-valued random variables on $(\Omega, \mathscr{A}, P)$ with finite expectations $E(Y_i)$. Then:

$$E\Big(\sum_{i=1}^{m} \alpha_i \cdot Y_i\Big) = \sum_{i=1}^{m} \alpha_i \cdot E(Y_i). \tag{ix}$$

$$E(Y) = \sum_{i=1}^{n} y_i \cdot P(Y = y_i) = \sum_{i=1}^{n} y_i \cdot P_Y(\{y_i\}) \tag{3.16}$$

[cf. Eq. (3.5)]. Hence, in this case the expectation of Y is simply the weighted sum of its values, each one weighted by its probability $P(Y = y_i) = P_Y(\{y_i\})$. ◁

3.1.4 Rules of Computation for Expectations

Some rules of computation for expectations are gathered in Box 3.1. These rules are proved in SN-Exercise 6-1. Note that the assumptions in these rules are always chosen to avoid terms such as $\infty - \infty$, because the difference between infinity and infinity is not defined. In contrast, $\infty + \infty = \infty$.

Example 3.15 [Joe and Ann With Randomized Assignment] Consider again the random variables X and Y displayed in Table 1.2. In Example 2.26 we defined the new random variable $X \cdot Y$ and specified the function $g\colon \{0,1\} \times \{0,1\} \to \overline{\mathbb{R}}$ such that $X \cdot Y$ is the composition of (X, Y) and g, that is, $X \cdot Y = g(X, Y)$. Hence, applying Equation (3.14) and using the probabilities displayed in Table 1.2 yields

$$\begin{aligned}
E(X \cdot Y) &= E\big(g(X,Y)\big) && [X \cdot Y = g(X,Y)] \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} g(x_i, y_j) \cdot P(X{=}x_i, Y{=}y_j) && [(3.14)] \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} x_i \cdot y_j \cdot P(X{=}x_i, Y{=}y_j) && [(2.23)] \\
&= 1 \cdot 1 \cdot P(X{=}1, Y{=}1) && [0 \cdot \alpha = 0, \alpha \in \mathbb{R}] \\
&= 1 \cdot (.16 + .08) = .24. && [\text{Table } 1.2]
\end{aligned}$$

◁

3.1.5 Conditional Expectation Value Given an Event

If we consider the expectation with respect to the conditional-probability measure P^B [see Eq. (1.21)], then we adapt the notation as follows:

$$E^B(Y) := \int Y \, dP^B, \tag{3.17}$$

presuming that Y is quasi-integrable with respect to P^B (see sect. 3.1.1).

Definition 3.16 [Conditional Expectation Value Given the Event B]
Let Y be a numerical random variable on $(\Omega, \mathscr{A}, P)$, assume $B \in \mathscr{A}$, $P(B) > 0$, and that Y is quasi-integrable with respect to the measure P^B. Then the expectation $E^B(Y)$ of Y with respect to the B-conditional-probability measure P^B is also called the B-conditional expectation value of Y or the conditional expectation value of Y given the event B. We also use the notation

$$E(Y \mid B) := E^B(Y). \tag{3.18}$$

Remark 3.17 [Discrete Y] If $P(B) > 0$ and $Y(\Omega) = \{y_1, \dots, y_n\} \subset \mathbb{R}$, then

$$E(Y \mid B) = E^B(Y) = \sum_{i=1}^{n} y_i \cdot P^B(Y{=}y_i) = \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i \mid B). \tag{3.19}$$

◁

Remark 3.18 [Conditional Expectation Value of an Indicator] If $P(B) > 0$, then considering the indicator 1_A of an event $A \in \mathscr{A}$, Equations (2.5) and (3.19) yield

$$E(1_A | B) = 0 \cdot P(1_A = 0 | B) + 1 \cdot P(1_A = 1 | B) = P(1_A = 1 | B) = P(A | B), \tag{3.20}$$

because $\{1_A = 1\} = A$. ◁

Remark 3.19 [$(X=x)$-Conditional-Probability Measure] If the assumptions of Definition 3.16 hold, X is a random variable on $(\Omega, \mathscr{A}, P)$, and

$$B = \{X=x\} = \{\omega \in \Omega \colon X(\omega) = x\},$$

then we also use the notation

$$P^{X=x} := P^B \tag{3.21}$$

for the conditional-probability measure P^B defined by Equation (1.21). ◁

Remark 3.20 [$(X=x)$-Conditional Expectation Value] If $B = \{X=x\}$, then we also write the expectation of Y with respect to P^B as follows:

$$E^{X=x}(Y) := E^B(Y). \tag{3.22}$$

Furthermore, we define

$$E(Y | X=x) := E(Y | B), \tag{3.23}$$

and call it the *$(X=x)$-conditional expectation value of Y*. Hence,

$$E(Y | X=x) = E^{X=x}(Y). \tag{3.24}$$

Note that in this definition of $E(Y|X=x)$ we presume $P(X=x) > 0$ (see Th. 1.33), which implies that $E(Y|X=x)$ is a uniquely defined number. [In contrast, in Definition 4.17 we will introduce a more general definition of $E(Y|X=x)$ that applies without assuming $P(X=x) > 0$. This will allow to state propositions about $E(Y|X=x)$ for P_X*-almost all* $x \in \Omega'_X$ (see Rem. 2.39). However, if $P(X=x) = 0$, then $E(Y|X=x)$ is not a uniquely defined number.] ◁

Remark 3.21 [Special Case if X is a Constant P-Almost Surely] Note that

$$X \underset{P}{=} x \quad \Rightarrow \quad E(Y | X=x) = E^{X=x}(Y) = E(Y). \tag{3.25}$$

This proposition follows from Equations (3.22) to (3.24), Equation (3.17), and Equation (1.24). ◁

Remark 3.22 [Conditional Probability $P(Y=y | X=x)$] Let $P(X=x) > 0$ and $1_{Y=y} = 1_{\{Y=y\}}$. Then we introduce the notation

$$P(Y=y | X=x) := E(1_{Y=y} | X=x) = E^{X=x}(1_{Y=y}) = P^{X=x}(Y=y), \tag{3.26}$$

and call this number the *$(X=x)$-conditional probability of (the event) $\{Y=y\}$*. ◁

Remark 3.23 [Conditional Expectation Value if Y is Discrete] If we assume $P(X{=}x) > 0$ and $Y(\Omega) = \{y_1, \ldots, y_n\} \subset \mathbb{R}$, then using the notation introduced in Equations (3.22) to (3.26), we can rewrite Equation (3.19) as

$$E(Y|X{=}x) = E^{X=x}(Y) = \sum_{i=1}^{n} y_i \cdot P^{X=x}(Y{=}y_i) = \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i|X{=}x). \tag{3.27}$$

◁

Remark 3.24 [Conditional Expectation Value if $g(Y)$ is Discrete] Equation (3.27) can be generalized considering, instead of Y, a composition $g(Y)$ of Y and a measurable function $g\colon (\Omega'_Y, \mathscr{A}'_Y) \to (\mathbb{R}, \mathscr{B})$. Under the assumptions of Remark 3.23,

$$E\big(g(Y)\big|X{=}x\big) = E^{X=x}\big(g(Y)\big) = \sum_{i=1}^{n} g(y_i) \cdot P^{X=x}(Y{=}y_i) = \sum_{i=1}^{n} g(y_i) \cdot P(Y{=}y_i|X{=}x). \tag{3.28}$$

◁

Equations (3.28) will now be generalized using the $(X{=}x)$-conditional distribution of Y, which is defined as follows:

Definition 3.25 [$(X{=}x)$-**Conditional Distribution of** Y]
Let X and Y be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let $(\Omega'_X, \mathscr{A}'_X)$ and $(\Omega'_Y, \mathscr{A}'_Y)$ denote their value spaces, let $x \in \Omega'_X$, $\{x\} \in \mathscr{A}'_X$, and $P(X{=}x) > 0$. Then we define the function $P_{Y|X=x}\colon \mathscr{A}'_Y \to [0,1]$ by

$$\forall\, A' \in \mathscr{A}'_Y\colon\; P_{Y|X=x}(A') = P\big(Y^{-1}(A')\big|X{=}x\big), \tag{3.29}$$

and call it the $(X{=}x)$-conditional distribution of Y.

(For a more general definition see SN-sect. 17.2). Equations (3.26) and (2.27) yield

$$\begin{aligned}\forall\, A' \in \mathscr{A}'_Y\colon\; P_{Y|X=x}(A') &= P\big(Y^{-1}(A')\big|X{=}x\big)\\ &= P^{X=x}\big(Y^{-1}(A')\big)\\ &= P_Y^{X=x}(A').\end{aligned} \tag{3.30}$$

Hence, $P_{Y|X=x}$, the $(X{=}x)$-conditional distribution of Y is identical to $P_Y^{X=x}$, the *distribution of Y with respect to the measure $P^{X=x}$*. That is,

$$P_{Y|X=x} = P_Y^{X=x}. \tag{3.31}$$

The transformation theorem (cf. Th. 3.11) for the expectation $E_Y^{X=x}(g)$ of g with respect to the distribution $P_Y^{X=x}$ of Y with respect to the measure $P^{X=x}$ can now be formulated as follows:

Theorem 3.26 [Transformation Theorem For $E_Y^{X=x}(g)$]
Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$, let $(\Omega'_X, \mathscr{A}'_X)$ and $(\Omega'_Y, \mathscr{A}'_Y)$ denote their value spaces, and let $x \in \Omega'_X$, $\{x\} \in \mathscr{A}'_X$, and $P(X{=}x) > 0$. Furthermore, assume that $g\colon (\Omega'_Y, \mathscr{A}'_Y) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ is a measurable function. If g is nonnegative or $E_Y^{X=x}(g)$ is finite, then

$$\begin{aligned} E_Y^{X=x}(g) &= \int g\, dP_Y^{X=x} = \int g(Y)\, dP^{X=x} = E^{X=x}\big(g(Y)\big) \\ &= \int g\, dP_{Y|X=x} = E\big(g(Y)\,\big|\,X{=}x\big). \end{aligned} \tag{3.32}$$

Remark 3.27 [An Alternative Notation] Using the alternative notation for the integral (see Rem. 3.3), we can also write

$$E_Y^{X=x}(g) = \int g(y)\, P_Y^{X=x}(dy) = \int g(y)\, P_{Y|X=x}(dy). \tag{3.33}$$

◁

There are two important points in Equations (3.32) and (3.33). *First*, these equations show the relationship between integrals of the composition $g(Y)$ with respect to the probability measure $P^{X=x}$ on $(\Omega, \mathscr{A})$ on one side, and the $(X{=}x)$-conditional distribution $P_{Y|X=x}$ of Y on the other side. *Second*, $E\big(g(Y)\,\big|\,X{=}x\big)$ is identical to the expectation of g with respect to the $(X{=}x)$-conditional distribution $P_{Y|X=x}$ of Y. Thus, in order to compute $E\big(g(Y)\,\big|\,X{=}x\big)$, using the $(X{=}x)$-conditional distribution $P_{g(Y)|X=x}$ of $g(Y)$ is not necessary.

3.1.6 Rules of Computation for Conditional Expectation Values

Remark 3.28 [Using Box 3.1 for Conditional Expectation Values] The rules of computation for expectations also apply if we replace the measure P by P^B or $P^{X=x}$ and the expectation $E(\,)$ by $E^B(\,)$ or $E^{X=x}(\,)$ (for more details see SN-Box 9.1). For example, replacing $E(\,)$ by $E^{X=x}(\,)$, Rule (ix) of Box 3.1 yields

$$E^{X=x}\Big(\sum_{i=1}^{n} \alpha_i \cdot Y_i\Big) = \sum_{i=1}^{n} \alpha_i \cdot E^{X=x}(Y_i), \tag{3.34}$$

and applying Equation (3.27),

$$E\Big(\sum_{i=1}^{n} \alpha_i \cdot Y_i\,\Big|\,X{=}x\Big) = \sum_{i=1}^{n} \alpha_i \cdot E(Y_i\,|\,X{=}x), \tag{3.35}$$

provided that $P(X{=}x) > 0$ and, for all $i = 1, \dots, n$, the random variables Y_i are real-valued and $\alpha_i \in \mathbb{R}$.

However, there are additional properties some of which are summarized in Box 3.2 (for a proof see SN-Exercise 9-2). Reading Rule (ii), note that (X, Z) is a (bivariate) random variable on $(\Omega, \mathscr{A}, P)$ with values $(x, z) \in \Omega'_X \times \Omega'_Z$ (see, e. g., sect. 2.1.5 for more details on multivariate random variables). Rule (ii) shows how the $(X{=}x)$-conditional expectation value $E(Y\,|\,X{=}x)$ can be computed from the conditional expectation values $E(Y\,|\,X{=}x, Z{=}z)$ and

Box 3.2 Rules of Computation for $(X{=}x)$-Conditional Expectation Values

Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$ with value spaces $(\Omega'_X, \mathscr{A}'_X)$ and $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$, respectively. Furthermore, let $x \in \Omega'_X$ with $\{x\} \in \mathscr{A}'_X$ and $P(X{=}x) > 0$. If $E(Y|X{=}x)$ exists, $f\colon (\Omega'_X, \mathscr{A}'_X) \to (\mathbb{R}, \mathscr{B})$ is a measurable function, and $E(Y^2)$, $E\big(f(X)^2\big) < \infty$, then

$$E\big(f(X)\cdot Y \,\big|\, X{=}x\big) = f(x)\cdot E(Y|X{=}x) = E\big(f(x)\cdot Y \,\big|\, X{=}x\big). \tag{i}$$

Additionally, let Z be a random variable on $(\Omega, \mathscr{A}, P)$ with value space $(\Omega'_Z, \mathscr{A}'_Z)$, assume that the image $Z(\Omega)$ of Z is finite or countable, and that $P_Z\big(Z(\Omega)\big) = 1$. If, for all $z \in Z(\Omega)$, $\{z\} \in \mathscr{A}'_Z$, and $P(X{=}x, Z{=}z) > 0$, then

$$E(Y|X{=}x) = \sum_{z \in Z(\Omega)} E(Y|X{=}x, Z{=}z)\cdot P(Z{=}z\,|\,X{=}x). \tag{ii}$$

the conditional probabilities $P(Z{=}z\,|\,X{=}x)$. Hence, considering Equation (3.27) and Rule (ii) in Box 3.2 shows that we have two different equations for computing the conditional expectation value $E(Y|X{=}x)$. ◁

Remark 3.29 [Special Case if X is a Constant P-Almost Surely] According to Proposition (3.25), $X \underset{P}{=} x$ implies $E(Y|X{=}x) = E(Y)$. Therefore, Rule (i) of Box 3.2 yields

$$X \underset{P}{=} x,\ x \in \mathbb{R} \quad \Rightarrow \quad E(X\cdot Y) = X\cdot E(Y) = x\cdot E(Y) = E(x\cdot Y), \tag{3.36}$$

if we consider the identity function $f(X) = id(X) = X$. This proposition generalizes Rule (iii) of Box 3.1. Furthermore, because

$$X \underset{P}{=} x \quad \Rightarrow \quad E(Y|X{=}x, Z{=}z) = E(Y|Z{=}z) \tag{3.37}$$

(see Exercise 3-8), Rule (ii) of Box 3.2 yields

$$E(Y) = \sum_{z \in Z(\Omega)} E(Y|Z{=}z)\cdot P(Z{=}z) \tag{3.38}$$

as a special case (see SN-Exercise 9-3). According to this equation, we can also compute the expectation of Y from the conditional expectations $E(Y|Z{=}z)$ and the probabilities $P(Z{=}z)$. Hence, if these parameters are available, then there is no need anymore to know or use the values y of Y in order to compute $E(Y)$ [cf. Eq. (3.7)]. ◁

Example 3.30 [Joe and Ann With Randomized Assignment] Consider the random experiment displayed in Table 1.2 and define the event

$$B = \big\{(Joe, yes, -), (Joe, yes, +), (Ann, yes, -), (Ann, yes, +)\big\} = \Omega_U \times \{yes\} \times \Omega_Y,$$

that the *drawn person is treated* (irrespective of whether or not the person is treated and success occurs), and the event

$$C = \big\{(Joe, no, +), (Joe, yes, +), (Ann, no, +), (Ann, yes, +)\big\} = \Omega_U \times \Omega_X \times \{+\}.$$

that *success occurs* (irrespective of which person is drawn and whether or not the person is treated). In Table 1.2 we assigned probabilities to each elementary event $\{\omega_i\}$, $\omega_i \in \Omega$, and defined the treatment variable $X := 1_B$ as well as the outcome variable $Y := 1_C$. In Example 3.8 we already computed the expectations $E(X) = .4$ and $E(Y) = .51$. Equations (2.5) and (3.19) yield

$$\begin{aligned}
E(Y|X{=}1) &= E(1_C|B) = P(C|B) = \frac{P(B\cap C)}{P(B)} \\
&= \frac{P\big(\{(Joe, yes, +), (Ann, yes, +)\}\big)}{P(B)} = \frac{P\big(\{(Joe, yes, +)\}\big) + P\big(\{(Ann, yes, +)\}\big)}{P(B)} \\
&= \frac{.16 + .08}{.4} = .6.
\end{aligned}$$

Hence, the conditional probability $P(Y{=}1|X{=}1) = E(Y|X{=}1)$ [see Eq. (3.26)] of success given treatment is .6. In contrast,

$$\begin{aligned}
E(Y|X{=}0) &= E(1_C|B^c) = P(C|B^c) = \frac{P(B^c\cap C)}{P(B^c)} \\
&= \frac{P\big(\{(Joe, no, +), (Ann, no, +)\}\big)}{P(B^c)} = \frac{P\big(\{(Joe, no, +)\}\big) + P\big(\{(Ann, no, +)\}\big)}{P(B^c)} \\
&= \frac{.21 + .06}{.6} = .45
\end{aligned}$$

(see also Exercise 3-4).

Note that, in this example, in which the treatment variable X and the person variable U are *independent*,

$$E(Y|X{=}1) - E(Y|X{=}0) = P(Y{=}1|X{=}1) - P(Y{=}1|X{=}0) = .15,$$

which is identical to the average of the two individual treatment effects,

$$\begin{aligned}
&E(Y|X{=}1, U{=}Joe) - E(Y|X{=}0, U{=}Joe) \\
&= P(Y{=}1|X{=}1, U{=}Joe) - P(Y{=}1|X{=}0, U{=}Joe) = .8 - .7 = .1
\end{aligned}$$

for Joe, and

$$\begin{aligned}
&E(Y|X{=}1, U{=}Ann) - E(Y|X{=}0, U{=}Ann) \\
&= P(Y{=}1|X{=}1, U{=}Ann) - P(Y{=}1|X{=}0, U{=}Ann) = .4 - .2 = .2
\end{aligned}$$

for Ann (see Example 1.24). ◁

Example 3.31 [Joe and Ann With Self-Selection] If we compute the difference between the corresponding conditional expectation values for the random experiment presented in Table 1.4, then we receive

$$E(Y|X{=}1) - E(Y|X{=}0) = P(Y{=}1|X{=}1) - P(Y{=}1|X{=}0) = .42 - .6 = -.18,$$

while the two individual treatment effects for Joe and Ann remain unchanged if compared to the corresponding values obtained in Example 3.30. Hence, in the example of Table 1.4, the difference $E(Y|X{=}1) - E(Y|X{=}0)$ is completely misleading if interpreted as the average effect of the treatment. Note that, in the random experiment presented in Table 1.4, the treatment variable X and the person variable U are *not independent* [see the conditional probabilities $P(X{=}1|U{=}Joe) = .04$ and $P(X{=}1|U{=}Ann) = .76$ in that table]. ◁

3.2 Variance, Covariance, and Correlation

Variances, covariances, and correlations are expectations of special random variables. They describe important properties of numerical random variables, although, in general, they do not determine the complete distribution.

3.2.1 Variance and Standard Deviation

Variance and standard deviation are the most important parameters quantifying the *variability* of a numerical random variable. We start with the definition of the variance and its square root, the standard deviation, assuming that the expectation of Y^2 is finite, which implies that $E(Y)$ and $Var(Y)$ are finite as well (see SN-Remarks 6.25 and 6.27).

Definition 3.32 [Variance and Standard Deviation]
Let Y be a numerical random variable on the probability space $(\Omega, \mathscr{A}, P)$ and assume that $E(Y^2)$ is finite. Then the variance of Y is defined by

$$Var(Y) := E\Big(\big(Y - E(Y)\big)^2\Big), \tag{3.39}$$

and the standard deviation of Y by the positive square root of the variance, that is,

$$SD(Y) := \sqrt{Var(Y)}. \tag{3.40}$$

According to this definition, $Var(Y)$ is the expectation of the squared *mean-centered random variable $Y - E(Y)$*. Note that variances and standard deviations are nonnegative. The variance of Y is also denoted by σ_Y^2 and the standard deviation by σ_Y.

Remark 3.33 [Variance of an Indicator] Let $(\Omega, \mathscr{A}, P)$ be a probability space and $A \in \mathscr{A}$. Then Rule (i) of Box 3.3 and Equation (3.8) yield

$$Var(1_A) = P(A) \cdot \big(1 - P(A)\big) = P(A) \cdot P(A)^2 \tag{3.41}$$

(see Exercise 3-1). Hence, the variance of the indicator of an event A does not contain any information that is not already contained in its expectation $E(1_A) = P(A)$. Other important properties of variances are summarized in Box 3.3. Note that a crucial assumption for Rule (vi) is independence of the random variables $Y_1, \dots, Y_m$ [see Def. (2.49)]. If this assumption does not apply, then we have to use Rule (ix) of Box 3.4, which involves the covariances of the random variables $Y_1, \dots, Y_m$. ◁

3.2.2 Covariance

While the variance quantifies the variability of a numerical random variable, the covariance quantifies the degree of co-variation of two numerical random variables, that is, the degree to which the two variables vary together in the following sense: If one variable takes on a large value (i. e., a large positive deviation from its expectation), then the other one

Box 3.3 Rules of Computation for Variances

Let X and Y be numerical random variables on $(\Omega, \mathscr{A}, P)$ with finite second moments. Then:

$$Var(Y) = E(Y^2) - E(Y)^2 \tag{i}$$

$$Var(\alpha + Y) = Var(Y), \quad \text{if } \alpha \in \mathbb{R} \tag{ii}$$

$$Var(\alpha \cdot Y) = \alpha^2 \cdot Var(Y), \quad \text{if } \alpha \in \mathbb{R} \tag{iii}$$

$$\exists\, y \in \mathbb{R} : Y \underset{P}{=} y \quad \Leftrightarrow \quad Var(Y) = 0 \tag{iv}$$

$$X \underset{P}{=} Y \quad \Rightarrow \quad Var(X) = Var(Y). \tag{v}$$

For $i = 1, \dots, m$, assume that the random variables Y_i on $(\Omega, \mathscr{A}, P)$ are real-valued, independent, and with finite second moments, and let $\alpha_i \in \mathbb{R}$. Then

$$Var\Big(\sum_{i=1}^{m} \alpha_i \cdot Y_i\Big) = \sum_{i=1}^{m} \alpha_i^2 \cdot Var(Y_i). \tag{vi}$$

tends to take on a large value as well. Furthermore, if one variable takes on a small value (i. e., a large negative deviation from its expectation), then the other one tends to take on a small value, too. In this case the covariance will be positive. However, the covariance may also be a negative real number. In this case, the two random variables co-vary in the following sense: If one variable takes on a large value, then the other one tends to take on a small value. Furthermore, if one variable takes on a small value, then the other one tends to take on a large value.

Reading Definition 3.34, note that assuming finiteness of $E(X^2)$ and $E(Y^2)$ implies that $E(X)$, $E(Y)$, $Var(X)$, $Var(Y)$, and $E(X \cdot Y)$ are finite as well (see SN-Rem. 7.1), which in turn implies that $Cov(X, Y)$ is finite.

Definition 3.34 [Covariance]
Let X and Y be two numerical random variables on the probability space $(\Omega, \mathscr{A}, P)$ such that $E(X^2)$ and $E(Y^2)$ are finite. Then the covariance of X and Y is defined by

$$Cov(X,Y) := E\Big(\big(X - E(X)\big) \cdot \big(Y - E(Y)\big)\Big). \tag{3.42}$$

Comparing Equations (3.39) and (3.42) to each other shows that the variance is the covariance of a numerical random variable with itself.

Remark 3.35 [Correlated Numerical Random Variables] According to this definition, the covariance of X and Y is the expectation of the product of the mean centered variables $X - E(X)$ and $Y - E(Y)$. Hence, a covariance can be negative, zero, or positive. If the covariance is different from zero, then we say that X and Y are *correlated*; if the covariance is zero, then we say that they are *uncorrelated*. ◁

Box 3.4 Rules of Computation for Covariances

Let X and Y be numerical random variables on the probability space $(\Omega, \mathscr{A}, P)$ with $E(X^2), E(Y^2) < \infty$ Then:

$$
\begin{aligned}
Cov(X,Y) &= E(X \cdot Y) - E(X) \cdot E(Y) && \text{(i)}\\
Cov(\alpha + X, \beta + Y) &= Cov(X,Y), \quad \text{if } \alpha, \beta \in \mathbb{R} && \text{(ii)}\\
Cov(\alpha X, \beta Y) &= \alpha\beta\, Cov(X,Y), \quad \text{if } \alpha, \beta \in \mathbb{R} && \text{(iii)}\\
Cov(X,X) &= Var(X) && \text{(iv)}\\
Cov(X,Y) &= Cov(Y,X) && \text{(v)}\\
X \perp\!\!\!\perp Y \quad &\Rightarrow \quad Cov(X,Y) = 0 && \text{(vi)}\\
\exists x \in \mathbb{R}: X \underset{P}{=} x \quad &\Rightarrow \quad Cov(X,Y) = 0. && \text{(vii)}
\end{aligned}
$$

If, for $i = 1, \ldots, m$, $\alpha_i \in \mathbb{R}$ and Y_i are real-valued random variables on the probability space $(\Omega, \mathscr{A}, P)$ with $E(Y_i^2) < \infty$, then

$$
\begin{aligned}
Var\Big(\sum_{i=1}^{m} \alpha_i Y_i\Big) &= \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j Cov(Y_i, Y_j) && \text{(viii)}\\
&= \sum_{i=1}^{m} \alpha_i^2 Var(Y_i) + \sum_{i=1}^{m} \sum_{j=1, i \neq j}^{m} \alpha_i \alpha_j Cov(Y_i, Y_j). && \text{(ix)}
\end{aligned}
$$

If, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, $\alpha_i, \beta_j \in \mathbb{R}$ and X_i, Y_j are real-valued random variables on the probability space $(\Omega, \mathscr{A}, P)$ with $E(X_i^2), E(Y_j^2) < \infty$, then

$$
Cov\Big(\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{m} \beta_j Y_j\Big) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j Cov(X_i, Y_j). \qquad \text{(x)}
$$

If $E(Y^2), E(X_1^2), E(X_2^2) < \infty$, then

$$
X_1 \underset{P}{=} X_2 \quad \Rightarrow \quad Cov(Y, X_1) = Cov(Y, X_2). \qquad \text{(xi)}
$$

Remark 3.36 [Covariance of Indicators] Let $(\Omega, \mathscr{A}, P)$ be a probability space and $A, B \in \mathscr{A}$. Then Rule (i) of Box 3.4 and Equation (3.8) yield

$$
Cov(1_A, 1_B) = E(1_A \cdot 1_B) - E(1_A) \cdot E(1_B) = P(A \cap B) - P(A) \cdot P(B), \qquad (3.43)
$$

because $1_A \cdot 1_B = 1_{A \cap B}$. The right-hand side of this equation shows that the covariance of the indicators of two events quantifies the degree of their dependence [cf. Eq. (1.27)]. ◁

The most important rules of computation for covariances are summarized in Box 3.4. Proofs of these rules are found in SN-Exercise 6-3 (see also Exercises 3-2 and 3-3).

3.2.3 Correlation

The covariance is not invariant under multiplication with constants [scale transformations; see Box 3.4 (iii)] of the random variables involved. In contrast, the correlation, which quantifies the strength of the same kind of dependence *is invariant* under scale transformations and under translations with additive constants (see Rem. 3.41).

Definition 3.37 [Correlation]
Let X and Y be numerical random variables on the probability space $(\Omega, \mathscr{A}, P)$ such that $E(X^2)$ and $E(Y^2)$ are finite. Then the correlation of X and Y is defined by

$$Corr\,(X,Y) := \begin{cases} \dfrac{Cov\,(X,Y)}{SD(X) \cdot SD(Y)}, & \text{if } SD(X),\, SD(Y) > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.44}$$

Remark 3.38 [Correlation of a Random Variable With Itself] Assume that $Var\,(X) > 0$. Because $Cov\,(X,X) = Var\,(X) = SD(X) \cdot SD(X)$, Equation (3.44) implies that $Corr\,(X,X) = 1$. Similarly, because $Cov\,(X,-X) = -Var\,(X) = -SD(X) \cdot SD(X)$, Equation (3.44) implies that $Corr\,(X,-X) = -1$. ◁

Remark 3.39 [Range of the Correlation] Note that

$$-1 \le Corr\,(X,Y) \le 1, \tag{3.45}$$

provided that $Corr\,(X,Y)$ that the assumptions hold under which the correlation is defined (see SN-Exercise 7-5). ◁

Remark 3.40 [Correlation and Z-Transformed Variables] If the standard deviations of X and Y are positive [i. e., if $SD(X), SD(Y) > 0$], then the correlation is also the expectation of the product of the *Z-transformed variables* $\big(X - E(X)\big)/SD(X)$ and $\big(Y - E(Y)\big)/SD(Y)$, respectively, that is,

$$Corr\,(X,Y) = E\left(\frac{X - E(X)}{SD(X)} \cdot \frac{Y - E(Y)}{SD(Y)}\right) \tag{3.46}$$

(see SN-Exercise 7-6). ◁

Remark 3.41 [An Invariance Property of the Correlation] The correlation of linear transformations of X and Y is

$$Corr\,(a_0 + a_1 X\,,\; b_0 + b_1 Y\,) = \begin{cases} Corr\,(X,Y), & \text{if } a_1 \cdot b_1 > 0, \\ -Corr\,(X,Y), & \text{if } a_1 \cdot b_1 < 0, \\ 0, & \text{if } a_1 \cdot b_1 = 0, \end{cases} \tag{3.47}$$

where $a_0, a_1, b_0, b_1 \in \mathbb{R}$ (see SN-Exercise 7-7). This equation implies that the correlation is invariant (up to change of signs) under linear transformations, which include *translations* ($a_1 = b_1 = 1$ and $a_0 \ne 0$ and/or $b_0 \ne 0$) and *scale transformations* ($a_0 = b_0 = 0$ and $1 \ne a_1 \ne 0$ and/or $1 \ne b_1 \ne 0$). ◁

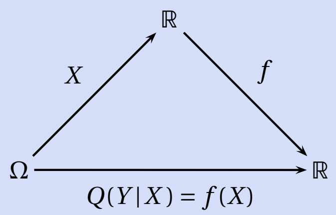

Figure 3.1. The regressor X, the linear quasi-regression f and their composition $Q(Y|X) = f(X)$.

3.3 Linear Quasi-Regression

3.3.1 *Simple Linear Quasi-Regression*

The covariance of two numerical random variables and their correlation quantify the strength of their dependence that can be described by a *linear quasi-regression*, sometimes also referred to as the *ordinary least-squares regression*.

Remark 3.42 [Implications of Finite Second Moments] Reading the following definition note that $E(X^2), E(Y^2) < \infty$ implies that $E(X)$, $E(Y)$, and $E(X \cdot Y)$ are finite (see SN-Rem. 7.1). ◁

Definition 3.43 [Simple Linear Quasi-Regression of Y on X]
Let X and Y be real-valued random variables on the probability space $(\Omega, \mathscr{A}, P)$ such that $E(X^2), E(Y^2) < \infty$ and $Var(X) > 0$, and define the function $MSE: \mathbb{R}^2 \to \mathbb{R}$ by

$$MSE(a_0, a_1) = E\Big(\big(Y - (a_0 + a_1 X)\big)^2\Big), \quad \forall (a_0, a_1) \in \mathbb{R}^2. \tag{3.48}$$

(i) *If $(\alpha_0, \alpha_1) \in \mathbb{R}^2$ minimizes MSE, then the function $f: \mathbb{R} \to \mathbb{R}$ defined by*

$$f(x) = \alpha_0 + \alpha_1 x, \quad \forall x \in \mathbb{R}, \tag{3.49}$$

is called the simple linear quasi-regression of Y on X.

(ii) *The composition of X and f is denoted by $Q(Y|X)$, that is,*

$$Q(Y|X) := f(X) = \alpha_0 + \alpha_1 X. \tag{3.50}$$

In this context the random variable X is called the *regressor* and Y the *regressand*. As suggested by the term linear quasi-regression, there is also a (true) linear regression (see Def. 4.19).

Remark 3.44 [Distinguishing Between f and $f(X)$] According to Equation (3.49), the simple linear quasi-regression f assigns a real number *to all real numbers* [see Eq. (3.49)

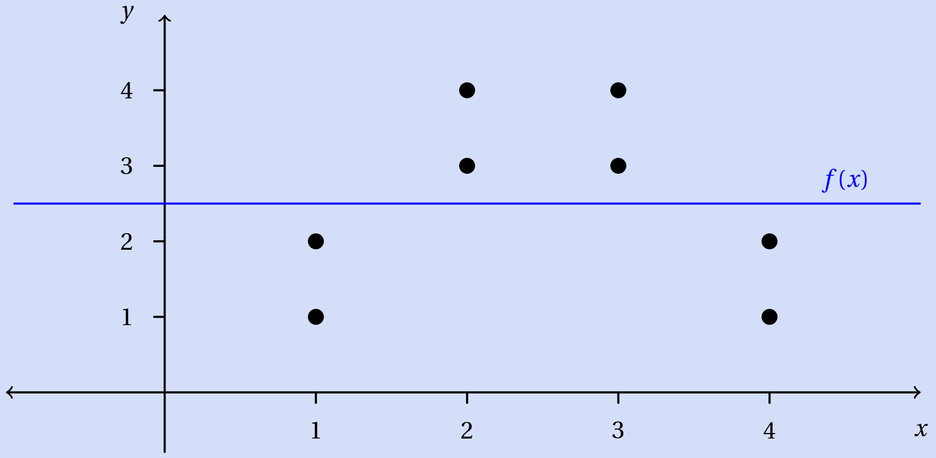

Figure 3.2. Linear quasi-regression of Y on X.

and Fig. 3.1)]. This applies even if X takes on only two different real values. Also note that $Q(Y|X)$ is an X-measurable random variable on $(\Omega, \mathscr{A}, P)$, whereas the simple linear quasi-regression f is a $\mathscr{B}$-measurable random variable on $(\mathbb{R}, \mathscr{B}, P_X)$ (see Lemma 2.24). $\lhd$

Theorem 3.45 [Two Characterizations of the Simple Linear Quasi-Regression]
Let the assumptions of Definition 3.43 hold, let $\alpha_0, \alpha_1 \in \mathbb{R}$, and consider a function $f: \mathbb{R} \to \mathbb{R}$ satisfying

$$f(x) = \alpha_0 + \alpha_1 x, \quad \forall\, x \in \mathbb{R}\,. \tag{3.51}$$

Then f is the simple linear quasi-regression of Y on X if and only if one of the following two conditions hold:

(i) $\alpha_0 = E(Y) - \alpha_1 E(X)$ *and* $\alpha_1 = \dfrac{Cov(X,Y)}{Var(X)}$.

(ii) The coefficients α_0, α_1 are such that

$$E(\epsilon) = 0 = Cov(X, \epsilon) \tag{3.52}$$

holds for $\epsilon := Y - (\alpha_0 + \alpha_1 X)$.

Hence, according to (i), a linear function f defined by Equation (3.51) whose coefficients α_0 and α_1 satisfy

$$\alpha_0 = E(Y) - \alpha_1 E(X) \tag{3.53}$$

and

$$\alpha_1 = \frac{Cov(X,Y)}{Var(X)}, \tag{3.54}$$

respectively, is the simple linear quasi-regression of Y on X.

Furthermore, according to (ii), a linear function f defined by Equation (3.51) whose residual $\epsilon = Y - (\alpha_0 + \alpha_1 X)$ satisfies Equations (3.52) is the simple linear quasi-regression of Y on X. Hence, each of conditions (i) and (ii) may replace the least-MSE criterion in the definition of a simple linear quasi-regression.

Example 3.46 [Discrete Regressor With Four Values] Let X and Y be real-valued random variables on $(\Omega, \mathscr{A}, P)$, both with values $1, 2, 3$ and 4. Furthermore, let their distribution be specified by

$$\begin{aligned} &P(X{=}1, Y{=}1) = 1/8, && P(X{=}1, Y{=}2) = 1/8,\\ &P(X{=}2, Y{=}3) = 1/8, && P(X{=}2, Y{=}4) = 1/8,\\ &P(X{=}3, Y{=}3) = 1/8, && P(X{=}3, Y{=}4) = 1/8,\\ &P(X{=}4, Y{=}1) = 1/8, && P(X{=}4, Y{=}2) = 1/8. \end{aligned}$$

Then the linear quasi-regression $f: \mathbb{R} \to \mathbb{R}$ is specified by

$$f(x) = \alpha_0 + \alpha_1 \cdot x = 2.5 + 0 \cdot x = 2.5, \quad \forall\, x \in \mathbb{R},$$

and the composition of X and f is

$$Q(Y|X) = \alpha_0 + \alpha_1 \cdot X = 2.5 + 0 \cdot X = 2.5$$

(see Exercise 3-6). The black points in Figure 3.2 represent the eight pairs of values of X and Y. All values of the linear quasi-regression are on the horizontal line, which, in this example, is parallel to the x-axis because its slope is 0. In contrast, in this example, $P_X(\{x\}) = 0$, for all $x \in \mathbb{R} \setminus \{1, 2, 3, 4\}$. Nevertheless, as mentioned before, a linear quasi-regression f is a function assigning a real number *to all real numbers*. In this specific example, the real number assigned is the same for all $x \in \mathbb{R}$, namely 2.5. ◁

3.3.2 Multiple Linear Quasi-Regression

In the following definition we consider more than a single regressor introducing the concept of the *multiple linear regression of* Y *on* $X_1, \ldots, X_m$. In this definition we refer to the variance-covariance matrix

$$\boldsymbol{\Sigma}_{\boldsymbol{xx}} = \begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \cdots & \sigma_{X_1X_m} \\ \sigma_{X_2X_1} & \sigma^2_{X_2} & \cdots & \sigma_{X_2X_m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_mX_1} & \sigma_{X_mX_2} & \cdots & \sigma^2_{X_m} \end{bmatrix} \tag{3.55}$$

of the regressors $X_1, \ldots, X_m$. The diagonal components of $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ are the variances of the variables $X_1, \ldots, X_m$, because $\sigma_{X_iX_i} := Cov(X_i, X_i) = Var(X_i) = \sigma^2_{X_i}$, $i = 1, \ldots, m$.

Definition 3.47 [Multiple Linear Quasi-Regression of Y on $X_1, \ldots, X_m$]
Let $X_1, \ldots, X_m$ and Y be real-valued random variables on the probability space $(\Omega, \mathscr{A}, P)$ with finite expectations $E(Y^2), E(X_i^2)$, $i = 1, \ldots, m$, and invertible covariance matrix of $X = (X_1, \ldots, X_m)$. Furthermore, define the function $MSE: \mathbb{R}^{m+1} \to \mathbb{R}$ by

$$MSE(a_0, a_1, \ldots, a_m) \;=\; E\Big(\big(Y - (a_0 + \sum_{i=1}^{m} a_i X_i)\big)^2\Big), \quad \forall\, (a_0, a_1, \ldots, a_m) \in \mathbb{R}^{m+1}. \tag{3.56}$$

(i) If $(\alpha_0, \alpha_1, \ldots, \alpha_m) \in \mathbb{R}^{m+1}$ minimizes MSE, then the function $f: \mathbb{R}^m \to \mathbb{R}$ defined by

$$f(x_1, \ldots, x_m) \;=\; \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_m x_m \quad \forall\, (x_1, \ldots, x_m) \in \mathbb{R}^m, \tag{3.57}$$

is called the (multiple) linear quasi-regression of Y on $X_1, \ldots, X_m$.

(ii) The composition of $X = (X_1, \ldots, X_m)$ and f is denoted $Q(Y \mid X_1, \ldots, X_m)$, that is,

$$Q(Y \mid X_1, \ldots, X_m) \;:=\; f(X) \;=\; \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_m X_m. \tag{3.58}$$

Theorem 3.45 can be generalized for the multiple linear quasi-regression, presenting two conditions each of which is equivalent to the condition used in Definition 3.47 of the multiple linear regression of Y on $X_1, \ldots, X_m$. In Theorem 3.48 we also refer to the covariance column vector of the regressors $X_1, \ldots, X_m$ with Y, that is,

$$\boldsymbol{\Sigma}_{\boldsymbol{xy}} \;=\; \begin{bmatrix} \sigma_{X_1 Y} \\ \vdots \\ \sigma_{X_m Y} \end{bmatrix},$$

where $\sigma_{X_i Y} := Cov(X_i, Y)$, $i = 1, \ldots, m$. Finally, we use the column vector

$$\boldsymbol{\alpha} \;=\; \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}.$$

Theorem 3.48 [Two Characterizations of the Multiple Linear Quasi-Regression]
Let the assumptions of Definition 3.47 hold, let $\alpha_0, \alpha_1, \ldots, \alpha_m \in \mathbb{R}$, and consider a function $f: \mathbb{R}^m \to \mathbb{R}$ satisfying

$$f(x_1, \ldots, x_m) \;=\; \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_m x_m, \quad \forall\, (x_1, \ldots, x_m) \in \mathbb{R}^m. \tag{3.59}$$

Then f is the multiple linear quasi-regression of Y on X if and only if one of the following two conditions hold:

(i) $\alpha_0 = E(Y) - \big(\alpha_1 E(X_1) + \ldots + \alpha_m E(X_m)\big)$ and $\boldsymbol{\alpha} = \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{xy}}$, where $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_m)$, $\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}$ denotes the inverse of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ of $(X_1, \ldots, X_m)$, and $\boldsymbol{\Sigma}_{\boldsymbol{xy}}$ the covariance column vector of $(X_1, \ldots, X_m)$ and Y.

(ii) The $\alpha_0, \alpha_1, \ldots, \alpha_m$ *are such that* $\epsilon := Y - (\alpha_0 + \alpha_1 X_1 + \ldots + \alpha_m X_m)$ *satisfies*

$$E(\epsilon) = 0 = Cov(X_i, \epsilon), \quad i = 1, \ldots, m. \tag{3.60}$$

For a proof of Theorem 3.48 see SN-Theorem 7.30. According to this theorem, the coefficients of the multiple linear quasi-regression satisfy the equations

$$\alpha_0 = E(Y) - \big(\alpha_1 E(X_1) + \ldots + \alpha_m E(X_m)\big) \tag{3.61}$$

and

$$\boldsymbol{\alpha} = \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{xy}}. \tag{3.62}$$

In fact, these equations are equivalent to postulating that f minimizes the function *MSE* defined in Equation (3.56). Furthermore, the residual ϵ of the multiple linear quasi-regression satisfies Equations (3.60), and again, these equations are equivalent to postulating that f minimizes the function *MSE*.

3.4 Summary and Conclusions

In this chapter we introduced the *expectation* of a numerical random variable. This concept has then been used to define the *variance* of a numerical random variable, the *covariance* of two numerical random variables, and their *correlation*. The chapter ended with the concept of a *linear quasi-regression* that can be used to describe how a real-valued random variable Y can best be approximated by linear function of X, where 'best' means minimizing the mean squared error function. Box 3.5 provides a glossary of these terms. The linear quasi-regression describes that kind of dependence of Y on X that is quantified by the covariance $Cov(X, Y)$ and the correlation $Corr(X, Y)$. However, this kind of dependence is blind against dependencies that can be described by nonlinear functions of X (see Example 3.46). Therefore, the next chapter deals with a concept that is much better suited for describing how Y depends on X, or more precisely, how the conditional expectation values $E(Y|X{=}x)$ of Y depend on the values x of X.

3.5 Exercises

$$X \underset{P}{=} x \quad \Rightarrow \quad E(Y|X{=}x, Z{=}z) = E(Y|Z{=}z) \tag{3.63}$$

▷ **Exercise 3-1** Prove Equation (3.41).

▷ **Exercise 3-2** Prove Box 3.4 (i) using the rules of computation for expectations gathered in Box 3.1.

▷ **Exercise 3-3** Prove rule (iii) of Box 3.4 using rule (i) of Box 3.4 and the rules of computation for expectations gathered in Box 3.1.

Box 3.5 Glossary of new concepts

$E(Y)$ — *Expectation* of a numerical random variable Y on $(\Omega, \mathscr{A}, P)$. If Y is quasi-integrable, that is, if $\int Y^+ dP < \infty$ or $\int Y^- dP < \infty$ holds for the positive and negative parts of Y, then it is defined by $E(Y) := \int Y\,dP$. If Y is discrete with a finite number of real values $y_1, \ldots, y_n$, then

$$E(Y) = \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i).$$

$E(Y|X{=}x)$ — $(X{=}x)$-*conditional expectation value* of a numerical random variable Y on $(\Omega, \mathscr{A}, P)$. If Y is quasi-integrable with respect to $P^{X=x}$, then it is defined by $E(Y|X{=}x) := E^{X=x}(Y) = \int Y\,dP^{X=x}$. If Y is discrete with a finite number of real values $y_1, \ldots, y_n$, then

$$E(Y|X{=}x) = \sum_{i=1}^{n} y_i \cdot P(Y{=}y_i|X{=}x).$$

$Var(Y)$ — *Variance* of a numerical random variable Y on $(\Omega, \mathscr{A}, P)$. If $E(Y^2) < \infty$, then $Var(Y) := E\big([Y - E(Y)]^2\big)$.

$SD(Y)$ — *Standard deviation of* Y. If $E(Y^2) < \infty$, then $SD(Y) := \sqrt{(Var(Y)}$.

$Cov(X,Y)$ — *Covariance* of numerical random variables X and Y on $(\Omega, \mathscr{A}, P)$. If $E(X^2), E(Y^2) < \infty$, then $Cov(X,Y) := E\big([X - E(X)] \cdot [Y - E(Y)]\big)$.

$Corr(X,Y)$ — *Correlation* of numerical random variables X and Y on $(\Omega, \mathscr{A}, P)$. If $E(X^2), E(Y^2) < \infty$, then $Corr(X,Y) := Cov(X,Y)/\big(SD(X) \cdot SD(Y)\big)$, if $SD(X), SD(Y) > 0$. Otherwise, we define $Corr(X,Y) := 0$.

$Q(Y|X)$ — The composition $f(X)$ of a real-valued random variable X on $(\Omega, \mathscr{A}, P)$ and the *simple linear quasi-regression* $f: \mathbb{R} \to \mathbb{R}$ defined by

$$\forall x \in \mathbb{R}: \; f(x) = \alpha_0 + \alpha_1 x, \quad \alpha_0, \alpha_1 \in \mathbb{R},$$

where α_0, α_1 are such that $E(\epsilon) = Cov(X, \epsilon) = 0$ holds for $\epsilon = Y - f(X)$.

$Q(Y|X_1, \ldots, X_m)$ — The composition $f(X_1, \ldots, X_m)$ of an m-variate real-valued random variable $X = (X_1, \ldots, X_m)$ on $(\Omega, \mathscr{A}, P)$ and the *multiple linear quasi-regression* $f: \mathbb{R}^m \to \mathbb{R}$ defined by

$$\forall (x_1, \ldots, x_m) \in \mathbb{R}^m: \; f(x_1, \ldots, x_m) = \alpha_0 + \sum_{i=1}^{m} \alpha_i x_i, \quad \alpha_0, \alpha_1, \ldots \alpha_m \in \mathbb{R},$$

where $\alpha_0, \alpha_1, \ldots \alpha_m$ are such that $E(\epsilon) = Cov(X_i, \epsilon) = 0$, for all $i = 1, \ldots, m$, holds for $\epsilon = Y - f(X_1, \ldots, X_m)$.

▷ **Exercise 3-4** Download *IP-Book Table 1.2.sav* from *www.causal-effects.de*. This data set has been generated from Table 1.2 for a sample of size $N = 10,000$. Estimate the conditional expectation values $E(Y \mid X{=}0) = P(Y{=}1 \mid X{=}0)$ and $E(Y \mid X{=}1) = P(Y{=}1 \mid X{=}1)$.

▷ **Exercise 3-5** Consider Table 1.2. Compute the variances, the covariance, and the correlation of the random variables X and Y.

▷ **Exercise 3-6** Consider Example 3.46 and compute the coefficients α_0 and α_1 of the linear quasi-regression of Y on X.

▷ **Exercise 3-7** Consider Example 3.46 and compute the coefficients of linear quasi-regression of Y on (X, X^2). For $x{=}1$ to $x{=}4$, compute its values and compare them to the values of the linear quasi-regression $Q(Y \mid X)$ in Example 3.46.

▷ **Exercise 3-8** Prove Proposition (3.63).

Solutions

▷ **Solution 3-1**

$$\begin{aligned}
Var(1_A) &= E(1_A{}^2) - E(1_A)^2 && \text{[Box 3.3 (i)]}\\
&= E(1_A) - E(1_A)^2 && [1_A{}^2 = 1_A]\\
&= P(A) - P(A)^2 && [(3.8)]\\
&= P(A) \cdot \big(1 - P(A)\big).
\end{aligned}$$

▷ **Solution 3-2**

$$\begin{aligned}
Cov(X,Y) &= E\Big(\big(X - E(X)\big) \cdot \big(Y - E(Y)\big)\Big) && [(3.42)]\\
&= E\big(X \cdot Y - X \cdot E(Y) - E(X) \cdot Y + E(X) \cdot E(Y)\big)\\
&= E(X \cdot Y) - E\big(X \cdot E(Y)\big) - E\big(E(X) \cdot Y\big) + E\big(E(X) \cdot E(Y)\big) && \text{[Box 3.1 (ix)]}\\
&= E(X \cdot Y) - E(X) \cdot E(Y) - E(X) \cdot E(Y) + E(X) \cdot E(Y) && \text{[Box 3.1 (i), (iii)]}\\
&= E(X \cdot Y) - E(X) \cdot E(Y).
\end{aligned}$$

▷ **Solution 3-3**

$$\begin{aligned}
Cov(\alpha X, \beta Y) &= E(\alpha X \cdot \beta Y) - E(\alpha X) \cdot E(\beta Y) && \text{[Box 3.4 (i)]}\\
&= \alpha\beta\, E(X \cdot Y) - \alpha\beta\, E(X) \cdot E(Y) && \text{[Box 3.1 (iii)]}\\
&= \alpha\beta\big(E(X \cdot Y) - E(X) \cdot E(Y)\big)\\
&= \alpha\beta\, Cov(X,Y). && \text{[Box 3.4 (i)]}
\end{aligned}$$

▷ **Solution 3-4** No solution provided. Just compare your results to the true parameters presented in Example 3.30.

▷ **Solution 3-5** Using Equation (3.41) for X in Table 1.2 and $P(X{=}1) = .4$ (see Example 3.8) yields

$$Var(X) = P(X{=}1) \cdot \big(1 - P(X{=}1)\big) = .4 \cdot (1 - .4) = .4 - .16 = .24.$$

Using the same equation for Y in that table and $P(Y{=}1) = .51$ (see again Example 3.8) yields

$$Var(Y) = P(Y{=}1) \cdot \big(1 - P(Y{=}1)\big) = .51 \cdot (1 - .51) = .51 - .2601 = .2499.$$

Using Equation (3.43) for X and Y in Table 1.2 and $P(X{=}1, Y{=}1) = .24$ (see Example 3.15) yields

$$Cov(X,Y) = P(X=1,Y=1) - P(X=1) \cdot P(Y=1) = .24 - .4 \cdot .51 = .036.$$

Finally, using Equation (3.44) yields

$$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X) \cdot SD(X)} = \frac{.036}{\sqrt{.24} \cdot \sqrt{.2499}} = \frac{.036}{.2449}. = .1469988.$$

Note that $Corr(X,Y)$ is not necessarily the best parameter to quantify the dependence between two binary random variables. However, it is defined also in this case.

▷ **Solution 3-6** We start computing the expectations of X and Y, the variance of X, and the covariance $Cov(X,Y)$. Using Equation (3.5) yields

$$\begin{aligned} E(X) &= \sum_{i=1}^{4} x_i \cdot P(X=x_i) = 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + 4 \cdot P(X=4) \\ &= 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 2.5. \end{aligned}$$

and

$$\begin{aligned} E(Y) &= \sum_{i=1}^{4} y_i \cdot P(Y=y_i) = 1 \cdot P(Y=1) + 2 \cdot P(Y=2) + 3 \cdot P(Y=3) + 4 \cdot P(Y=4) \\ &= 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 2.5. \end{aligned}$$

Using Equation (3.13) yields,

$$\begin{aligned} Var(X) &= E\Big(\big(X - E(X)\big)^2\Big) \\ &= \sum_{i=1}^{4} \big(x_i - E(X)\big)^2 \cdot P(X=x_i) \\ &= (1-2.5)^2 \cdot \frac{1}{4} + (2-2.5)^2 \cdot \frac{1}{4} + (3-2.5)^2 \cdot \frac{1}{4} + (4-2.5)^2 \cdot \frac{1}{4} \\ &= \frac{1}{4} \cdot \big(2.25 + .25 + .25 + 2.25\big) = 1.25. \end{aligned}$$

and

$$\begin{aligned} & Cov(X,Y) \\ &= E\Big(\big(X - E(X)\big) \cdot \big(Y - E(Y)\big)\Big) && [(3.42)] \\ &= E(X \cdot Y) - E(X) \cdot E(Y) && [\text{Box } 3.4\ (i)] \\ &= \sum_{(x,y)} (x \cdot y) \cdot P(X=x, Y=y) - E(X) \cdot E(Y) && [(3.13)] \\ &= \sum_{i=1}^{4} \sum_{j=1}^{4} (x_i \cdot y_j) \cdot P(X=x_i, Y=y_j) - E(X) \cdot E(Y) \\ &= 1 \cdot 1 \cdot \frac{1}{8} + 1 \cdot 2 \cdot \frac{1}{8} + 2 \cdot 3 \cdot \frac{1}{8} + 2 \cdot 4 \cdot \frac{1}{8} + 3 \cdot 3 \cdot \frac{1}{8} + 3 \cdot 4 \cdot \frac{1}{8} + 4 \cdot 1 \cdot \frac{1}{8} + 4 \cdot 2 \cdot \frac{1}{8} - 2.5 \cdot 2.5 \\ &= \frac{1}{8} \cdot \big(1 + 2 + 6 + 8 + 9 + 12 + 4 + 8\big) = \frac{50}{8} - 6.25 = 0. \end{aligned}$$

The formulas of Theorem 3.45 (i) now yield

$$\alpha_1 = \frac{Cov(X,Y)}{Var(X)} = \frac{0}{1.25} = 0 \quad \text{and} \quad \alpha_0 = E(Y) - \alpha_1 \cdot E(X) = 2.5 - 0 \cdot 2.5 = 2.5.$$

▷ **Solution 3-7** We use Equations (3.61) and (3.62) to compute α_0 and the vector $\boldsymbol{\alpha} = \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{xy}}$. In Exercise 3-6 we already computed the $E(X)$, $E(Y)$, $Var(X)$, and $Cov(X,Y)$. Additionally, we need $E(X^2)$, $Var(X^2)$, $Cov(X,X^2)$, and $Cov(X^2,Y)$. Using Equation (3.13) yields

$$E(X^2) = \sum_{i=1}^{4} x_i^2 \cdot P(X{=}x_i) = 1\cdot P(X{=}1) + 4\cdot P(X{=}2) + 9\cdot P(X{=}3) + 16\cdot P(X{=}4)$$
$$= 1\cdot\frac{1}{4} + 4\cdot\frac{1}{4} + 9\cdot\frac{1}{4} + 16\cdot\frac{1}{4} = 7.5.$$

Box 3.3 (i) and Equation (3.13) yield

$$\begin{aligned} Var(X^2) &= E(X^4) - E(X^2)^2 && \text{[Box 3.3 (i)]}\\ &= 1\cdot P(X{=}1) + 16\cdot P(X{=}2) + 81\cdot P(X{=}3) + 256\cdot P(X{=}4) - E(X^2)^2\\ &= (1+16+81+256)\cdot\frac{1}{4} - 7.5^2 = 88.5 - 56.25 = 32.25. \end{aligned}$$

Similarly, Box 3.4 (i) and Equation (3.13) yield

$$\begin{aligned} Cov(X,X^2) &= E(X\cdot X^2) - E(X)\cdot E(X^2) && \text{[Box 3.4 (i)]}\\ &= \sum_{i=1}^{4}(x_i\cdot x_i^2)\cdot P(X{=}x_i) - E(X)\cdot E(X^2)\\ &= 1\cdot\frac{1}{4} + 8\cdot\frac{1}{4} + 27\cdot\frac{1}{4} + 64\cdot 4\cdot\frac{1}{4} - 2.5\cdot 7.5\\ &= 100\cdot\frac{1}{4} - 18.75 = 25 - 18.75 = 6.25, \end{aligned}$$

and

$$\begin{aligned} &Cov(X^2,Y)\\ &= E(X^2\cdot Y) - E(X^2)\cdot E(Y) && \text{[Box 3.4 (i)]}\\ &= \sum_{i=1}^{4}\sum_{j=1}^{4}(x_i^2\cdot y_j)\cdot P(X{=}x_i, Y{=}y_j) - E(X^2)\cdot E(Y)\\ &= 1\cdot 1\cdot\frac{1}{8} + 1\cdot 2\cdot\frac{1}{8} + 4\cdot 3\cdot\frac{1}{8} + 4\cdot 4\cdot\frac{1}{8} + 9\cdot 3\cdot\frac{1}{8} + 9\cdot 4\cdot\frac{1}{8} + 16\cdot 1\cdot\frac{1}{8} + 16\cdot 2\cdot\frac{1}{8} - 7.5\cdot 2.5\\ &= \frac{1}{8}\cdot\big(1 + 2 + 12 + 16 + 27 + 36 + 16 + 32\big) - 18.75 = 17.75 - 18.75 = -1, \end{aligned}$$

and Equation (3.62) yields

$$\begin{aligned} \boldsymbol{\alpha} = \begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} &= \begin{bmatrix} Var(X) & Cov(X,X^2)\\ Cov(X^2,X) & Var(X^2)\end{bmatrix}^{-1} \begin{bmatrix} Cov(X,Y)\\ Cov(X^2,Y)\end{bmatrix}\\ &= \begin{bmatrix}1.25 & 6.25\\ 6.25 & 32.25\end{bmatrix}^{-1}\begin{bmatrix}0\\ -1\end{bmatrix} = \begin{bmatrix}25.8 & -5\\ -5 & 1\end{bmatrix}\begin{bmatrix}0\\ -1\end{bmatrix} = \begin{bmatrix}5\\ -1\end{bmatrix}. \end{aligned}$$

[The inverse of $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ is easily obtained in R by `solve(matrix(c(1.25,6.25, 6.25,32.25),byrow=TRUE,nrow=2))`.]
Inserting these results for α_1 and α_2 into Equation (3.61) yields

$$\alpha_0 = E(Y) - \big(\alpha_1\cdot E(X) + \alpha_2\cdot E(X^2)\big) = 2.5 - (5\cdot 2.5 - 1\cdot 7.5) = -2.5.$$

Inserting the results for α_0, α_1, and α_2, we obtain the following four values of the linear quasi-regression of Y on (X, X^2):

$$f(1) = -2.5 + 5\cdot 1 - 1\cdot 1 = 1.5, \qquad f(2) = -2.5 + 5\cdot 2 - 1\cdot 2^2 = 3.5,$$

$$f(3) = -2.5 + 5\cdot 3 - 1\cdot 3^2 = 3.5, \qquad f(4) = -2.5 + 5\cdot 4 - 1\cdot 4^2 = 1.5.$$

These values are identical to the conditional expectation values $E(Y|X{=}1)$ to $E(Y|X{=}4)$ (see Fig. 3.2). [The same four values would also be obtained if we would compute the values of the multiple linear quasi-regression of Y on $(1_{X=2}, 1_{X=3}, 1_{X=4})$.]

▷ **Solution 3-8**

$$
\begin{aligned}
X \underset{P}{=} x \quad &\Leftrightarrow \quad P\big(\{\omega \in \Omega\colon X(\omega) = x\}\big) = 1 && [(2.29)]\\
&\Leftrightarrow \quad P(X{=}x) = 1 && [(2.3)]\\
&\Rightarrow \quad P(X{=}x, Z{=}z) = P(Z{=}z) && [\text{Box 1.1 (viii)}]\\
&\Rightarrow \quad P(X{=}x, Z{=}z)/P(Z{=}z) = 1 && \\
&\Rightarrow \quad P(X{=}x \,|\, Z{=}z) = 1 && [(1.10)]\\
&\Rightarrow \quad E(Y \,|\, Z{=}z) = E(Y \,|\, X{=}x, Z{=}z). && [\text{Box 3.2 (ii)}]
\end{aligned}
$$

Chapter 4
Conditional Expectation

In this chapter we introduce *conditional expectations* and some related concepts. The concept of a conditional expectation has been introduced by Kolmogorov (1933/1977), together with the axioms of probability. (For an English translation see Kolmogorov, 1956). A conditional expectation $E(Y|X)$ can be used to describe how the conditional expectation values $E(Y|X{=}x)$ of a numerical random variable Y depend on the values x of a (not necessarily numerical and possibly multivariate) random variable X. In many cases conditional expectations and their values are what we try to estimate in statistical modeling. Conditional expectations are also used to define *conditional independence of random variables given a random variable* (see ch. 6).

4.1 Discrete Conditional Expectation

We start with the a discrete conditional expectation, building on the concept of a conditional expectation value $E(Y|X{=}x)$, which has been introduced in Definition 3.17 and Remark 3.20 assuming $P(X{=}x) > 0$. Note that this assumption does not hold if X is a continuous real-valued random variable (see Rem. 2.66). This is the case, for example, if X has a normal distribution. Also note that the value space $(\Omega'_X, \mathscr{A}'_X)$ can consists of any set Ω'_X and a σ-algebra on this set. In particular, Ω'_X does not have to be (a subset of) the set of real numbers.

Definition 4.1 [X**-Conditional Expectation if** X **is Discrete**]
Let X, Y be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and nonnegative or has a finite expectation. If the image $X(\Omega) = \{X(\omega)\colon \omega \in \Omega\}$ of X is finite or countable and, for all $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$ and $P(X{=}x) > 0$, then we call

$$E(Y|X) = \sum_{x \in X(\Omega)} E(Y|X{=}x) \cdot 1_{X=x} \tag{4.1}$$

the discrete X-conditional expectation of Y.

Remark 4.2 [Alternative Notation for the Sum] Let the assumptions of Definition 4.1 hold. If $X(\Omega)$ is *finite* with elements $x_1, \ldots, x_n$, then we may also write

$$E(Y|X) = \sum_{i=1}^{n} E(Y|X{=}x_i) \cdot 1_{X=x_i}. \tag{4.2}$$

If $X(\Omega)$ is *countable* with elements $x_1, x_2, \ldots$, then

$$E(Y|X) = \sum_{i=1}^{\infty} E(Y|X=x_i) \cdot 1_{X=x_i}, \tag{4.3}$$

which is a convenient notation for

$$E(Y|X) = \lim_{n\to\infty} \sum_{i=1}^{n} E(Y|X=x_i) \cdot 1_{X=x_i}. \tag{4.4}$$

◁

Remark 4.3 [Values of $E(Y|X)$] Because the indicators $1_{X=x}$ are random variables on the probability space $(\Omega, \mathscr{A}, P)$ (see Example 2.6), Equation (4.1) shows that $E(Y|X)$ is a random variable on $(\Omega, \mathscr{A}, P)$ as well (see SN-Example 2.61) and that the values of $E(Y|X)$ are the conditional expectation values $E(Y|X=x)$. More precisely,

$$\forall\, x \in X(\Omega), \forall\, \omega \in \Omega: \quad E(Y|X)(\omega) = E(Y|X=x), \quad \text{if } \omega \in \{X=x\}. \tag{4.5}$$

◁

4.2 Conditional Expectation

In the following definition of a conditional expectation $E(Y|X)$ we drop the assumption $P(X=x) > 0$ for all $x \in X(\Omega)$. Hence, this definition also applies if X is continuous. In this definition, we use $\sigma(X)$, the σ-algebra generated by the random variable X (see Def. 2.12) and the concept of measurability of a random variable with respect to another one (see Rem. 2.13). Note again that the value space $(\Omega'_X, \mathscr{A}'_X)$ of X can be any measurable space. In particular, in contrast to the co-domain of Y, we do not have to assume that Ω'_X is a subset of $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$.

Definition 4.4 [Conditional Expectation]
Let X, Y be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and nonnegative or with finite expectation $E(Y)$. Then a numerical random variable V on $(\Omega, \mathscr{A}, P)$ is called a version of the X-conditional expectation of Y (with respect to P), if the following two conditions hold:

(a) $\sigma(V) \subset \sigma(X)$
(b) $E(1_C \cdot V) = E(1_C \cdot Y), \quad \forall\, C \in \sigma(X).$

If V satisfies (a) and (b), then we also use the notation $E(Y|X) := V$.

A version $E(Y|X)$ of the X-conditional expectation of Y is a random variable on $(\Omega, \mathscr{A}, P)$, and according to condition (a) of Definition 4.4, $E(Y|X)$ is measurable with respect to X (cf. Rem. 2.13).

Remark 4.5 [The Set $\mathscr{E}(Y|X)$] Note that there can be several random variables satisfying conditions (a) and (b). Therefore, we define $\mathscr{E}(Y|X)$ to be the set of all random variables satisfying conditions (a) and (b) of Definition 4.4. Hence, $\mathscr{E}(Y|X)$ denotes the set of all versions of the X-conditional expectation of Y with respect to the measure P. ◁

Remark 4.6 [The Discrete Conditonal Expectation is a Conditonal Expectation] The discrete conditional expectation $E(Y|X)$ introduced in Definition 4.1 is an element of $\mathscr{E}(Y|X)$, that is, it is a version of the conditional expectation of Y given X as introduced in Definition 4.4 (see Exercise 4-1). ◁

Remark 4.7 [P-Uniqueness] According to SN-Remark 10.15,

$$V, V^* \in \mathscr{E}(Y|X) \quad \Rightarrow \quad V \underset{P}{=} V^*, \tag{4.6}$$

where $V \underset{P}{=} V^*$ is a shortcut for

$$P\big(\{\omega \in \Omega\colon V(\omega) = V^*(\omega)\}\big) = 1 \tag{4.7}$$

(see also sect. 2.3.2). If Equation (4.7) holds, then we say that V and V^* are *P-equivalent* or *identical P-almost surely*. If (4.6) holds, that is, if $V \underset{P}{=} V^*$ *for all pairs* $V, V^* \in \mathscr{E}(Y|X)$, then we say that $E(Y|X)$ *is P-unique*. ◁

Remark 4.8 [Two Versions of a Conditional Expectation] According to SN-Remark 10.14,

$$V, V^* \in \mathscr{E}(Y|X) \quad \Leftrightarrow \quad V \underset{P}{=} V^* \wedge V, V^* \text{ are } X\text{-measurable.} \tag{4.8}$$

Hence, V and V^* are two versions of the X-conditional expectation of Y if and only if V and V^* are identical P-almost surely and measurable with respect to X. ◁

Remark 4.9 [Implications of P-Almost Sure Identity] If two random variables are identical P-almost surely, then they have identical expectations, variances, and covariances with other random variables provided that these expectations, variances, and covariances exist [see Cor. 2.41, Box 3.1 (vi), Box 3.3 (v), and Box 3.4 (xi)]. If they also have identical value spaces, then they also have identical distributions (see Cor. 2.41). ◁

Remark 4.10 [Conditional Probability Given a Random Variable] Let $(\Omega, \mathscr{A}, P)$ be a probability space and let 1_A denote the indicator of the event $A \in \mathscr{A}$. We introduce the notation

$$P(A|X) := E(1_A|X) \tag{4.9}$$

and call it the *X-conditional probability of (the event) A* (with respect to P). Furthermore, considering the event $\{Y{=}y\} = \{\omega \in \Omega\colon Y(\omega) = y\}$, we also use the notation

$$P(Y{=}y|X) := P(\{Y{=}y\}|X) = E(1_{Y=y}|X), \tag{4.10}$$

and call it the *X-conditional probability of (the event) $\{Y{=}y\}$* (with respect to P). Note again that $P(A|X)$ and $P(Y{=}y|X)$ are X-measurable random variables on $(\Omega, \mathscr{A}, P)$. ◁

Remark 4.11 [$P(Y{=}y|X)$ and $P(Y{\neq}y|X)$] Let X, Y be random variables on the probability space $(\Omega, \mathscr{A}, P)$ and let $Y(\Omega) = \{Y(\omega)\colon \omega \in \Omega\}$ denote the image of Y. Then

$$1 \underset{P}{=} P\big(Y \in Y(\Omega)|X\big) \underset{P}{=} P(Y{=}y|X) + P(Y{\neq}y|X).$$

Therefore,

$$P(Y{\neq}y|X) \underset{P}{=} 1 - P(Y{=}y|X). \tag{4.11}$$

◁

Remark 4.12 [Multivariate X] If $X = (X_1, \dots, X_m)$ is an m-variate random variable on the probability space $(\Omega, \mathscr{A}, P)$ (see sect. 2.1.5), then a version $E(Y|X)$ of the conditional expectation is also denoted by $E(Y|X_1, \dots, X_m)$. In the same vein, if $Y = 1_A$ or $Y = 1_{X=x}$, then we use the notation $P(A|X_1, \dots, X_m)$ and $P(Y{=}y|X_1, \dots, X_m)$, respectively. ◁

Many important properties of a conditional expectation $E(Y|X)$ are gathered in Box 4.1, which has been adopted from SN-Box 10.2. A proof of these properties is found in the solution to SN-Exercise 10-4. Many of these properties will often be used in this and the remaining chapters of this book. Additional properties dealing with monotonicity of a conditional expectation $E(Y|X)$ are gathered in SN-Box 10.3.

A proposition on strict monotonicity is stated in the following lemma. Reading this lemma, note that $Y \underset{P}{>} \alpha$ is defined by

$$P\big(\{\omega \in \Omega \colon Y(\omega) > \alpha\}\big) = 1.$$

Lemma 4.13 [Strict Monotonicity of a Conditional Expectation]
Let the assumptions of Definition 4.4 hold and let $\alpha \in \mathbb{R}$. *Then*

$$Y \underset{P}{>} \alpha \quad \Rightarrow \quad E(Y|X) \underset{P}{>} \alpha. \tag{4.12}$$

(Proof p. 95)

Example 4.14 [Joe and Ann With Randomized Assignment] Consider again the example presented in Table 1.2. In this table we already displayed the values .45 and .6 of the conditional expectation $E(Y|X) = P(Y{=}1|X)$. This random variable satisfies conditions (a) and (b) of Definition 4.4 (see Exercise 4-4). In this example, there is only one single version of the X-conditional expectation of Y. That is, the set $\mathscr{E}(Y|X)$ has only one single element. Hence, in this example, $E(Y|X)$ is not only P-unique, but it is *uniquely* defined. This also holds for the conditional expectation $E(Y|X,U)$ specified in the same table. ◁

Example 4.15 [No Treatment for Joe] Table 2.1 displays another example illustrating the concept of a conditional expectation. In this example, there are uncountably many versions of the (X,U)-conditional expectation of Y. The column headed by $P(Y{=}1|X,U)$ displays *one* such version. Another one is obtained if, in this column, we replace the value 99 by any other number. Suppose $V = P(Y{=}1|X,U)$ is the version displayed in the table and $P(Y{=}1|X,U)^*$ is another version obtained by assigning to ω_3 and ω_4 the (arbitrarily chosen) number .8, leaving the assignments to the other $\omega_i \in \Omega$ untouched [see the column headed by $P(Y{=}1|X,U)^*$]. Obviously, $P(Y{=}1|X,U) \underset{P}{=} P(Y{=}1|X,U)^*$, because the values of these two versions only differ for the null set $\{\omega_3, \omega_4\}$. This illustrates that, in this example, the (X,U)-conditional expectation of Y is not uniquely defined. However, it is P-unique, because $V \underset{P}{=} V^*$, for all $V, V^* \in \mathscr{E}(Y|X,U)$. In contrast to $P(Y{=}1|X,U)$, the conditional expectations $P(Y{=}1|X)$ and $P(X{=}1|U)$ *are* uniquely defined in this example. In other words, the sets $\mathscr{E}(Y|X)$ and $\mathscr{E}(X|U)$ consist of only one single element, the random variables $P(Y{=}1|X)$ and $P(X{=}1|U)$, respectively, which are specified in Table 2.1. ◁

Box 4.1 Rules of Computation for X-Conditional Expectations

Let the assumptions of Definition 4.4 hold. Then:

$$E(\alpha \mid X) \underset{P}{=} \alpha, \quad \text{if } \alpha \in \mathbb{R}. \tag{i}$$

$$E(\alpha + Y \mid X) \underset{P}{=} \alpha + E(Y \mid X), \quad \text{if } \alpha \in \mathbb{R}. \tag{ii}$$

$$E(\alpha \cdot Y \mid X) \underset{P}{=} \alpha \cdot E(Y \mid X), \quad \text{if } \alpha \in \mathbb{R}. \tag{iii}$$

$$E\big(E(Y \mid X)\big) = E(Y). \tag{iv}$$

$$E(Y \mid X) \underset{P}{=} E(Y), \quad \text{if } Y \perp\!\!\!\perp X. \tag{v}$$

$$E(Y \mid X) \underset{P}{=} E(Y) \text{ and } E(Y \mid X, Z) \underset{P}{=} E(Y \mid Z), \quad \text{if } X \underset{P}{=} x. \tag{vi}$$

$$E(Y \mid X) \underset{P}{=} E\big(Y \,\big|\, E(Y \mid X)\big). \tag{vii}$$

$$Cov\big(Y, E(Y \mid X)\big) = Var\big(E(Y \mid X)\big), \quad \text{if } E(Y^2) < \infty. \tag{viii}$$

$$E(Y) \text{ is finite} \Rightarrow \exists V \in \mathscr{E}(Y \mid X)\colon V \text{ is real-valued.} \tag{ix}$$

$$E(Y^2) < \infty \Rightarrow E\big(E(Y \mid X)^2\big) < \infty. \tag{x}$$

$$E(Y \mid X) \underset{P}{=} Y, \quad \text{if } \sigma(Y) \subset \sigma(X). \tag{xi}$$

Additionally, let Z be a random variable on $(\Omega, \mathscr{A}, P)$. Then:

$$E\big(E(Y \mid X) \,\big|\, Z\big) \underset{P}{=} E(Y \mid Z), \quad \text{if } \sigma(Z) \subset \sigma(X). \tag{xii}$$

Additionally, let Z be numerical. Then:

$$E(Y \mid X) \underset{P}{=} E(Z \mid X), \quad \text{if } Y \underset{P}{=} Z. \tag{xiii}$$

$$Cov\big(Z, E(Y \mid X)\big) = Cov(Z, Y), \quad \text{if } E(Z^2), E(Y^2) < \infty. \tag{xiv}$$

Additionally, let Y be X-measurable, and assume $E(Y^2), E(Z^2) < \infty$ or $Y, Z \geq 0$. Then:

$$E(Y \cdot Z \mid X) \underset{P}{=} Y \cdot E(Z \mid X). \tag{xv}$$

If Y and Z are numerical and nonnegative or real-valued with finite expectations, then there is a nonnegative (if Y is nonnegative) or real-valued (if Y is real-valued) version $E(Y \mid X) \in \mathscr{E}(Y \mid X)$ and a nonnegative (if Z is nonnegative) or real-valued (if Z is real-valued) version $E(Z \mid X) \in \mathscr{E}(Z \mid X)$ such that

$$E(Y + Z \mid X) \underset{P}{=} E(Y \mid X) + E(Z \mid X). \tag{xvi}$$

If $Y_1, \dots, Y_n$ are real-valued random variables on $(\Omega, \mathscr{A}, P)$ with finite expectations and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, then

$$E\Big(\sum_{i=1}^{n} \alpha_i \cdot Y_i \,\Big|\, X\Big) \underset{P}{=} \sum_{i=1}^{n} \alpha_i \cdot E(Y_i \mid X). \tag{xvii}$$

4.3 Factorization of a Conditional Expectation and Regression

According to the following corollary, a version $E(Y|X) \in \mathscr{E}(Y|X)$ can always be written as a composition $g(X)$ (sometimes also denoted by $g \circ X$) of X and a numerical measurable function g (see SN-Cor. 10.23). (For the concept of a measurable function see Def. 2.2 and Rem. 2.5).

Corollary 4.16 [Existence of the Factorization]
Let the assumptions of Definition 4.4 hold, let $(\Omega'_X, \mathscr{A}'_X)$ denote the value space of X, and let $E(Y|X) \in \mathscr{E}(Y|X)$. Then there is a measurable function $g\colon (\Omega'_X, \mathscr{A}'_X) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ such that

$$E(Y|X) = g(X). \tag{4.13}$$

The function g occurring in Corollary 4.16 plays an important role. Among other things, it is used for a general definition of a *regression of Y on X*. The measurable space $(\overline{\mathbb{R}}^m, \overline{\mathscr{B}}_m)$ mentioned in Definition 4.17 (ii) consists of the m-fold Cartesian product of the set $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and the Borel σ-algebra on this set (see SN-sect. 1.2.2).

Definition 4.17 [Factorization and Regression]
Let the assumptions of Corollary 4.16 hold.

(i) A measurable function $g\colon (\Omega'_X, \mathscr{A}'_X) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ that satisfies Equation (4.13) is called a factorization of $E(Y|X)$.

(ii) If $(\Omega'_X, \mathscr{A}'_X) = (\overline{\mathbb{R}}^m, \overline{\mathscr{B}}_m)$, $m \in \mathbb{N}$, then g is also called an (m-variate) regression of Y on X. If $m = 1$, then it is called a simple regression, if $m > 1$, a multiple regression.

Remark 4.18 [Conditional Expectation vs. Regression] Hence, while a conditional expectation $E(Y|X) = g(X)$ is a random variable on $(\Omega, \mathscr{A}, P)$ with domain Ω, a factorization g is a function with domain Ω'_X. In fact, g is a numerical random variable on the probability space $(\Omega'_X, \mathscr{A}'_X, P_X)$. If $\Omega'_X = \overline{\mathbb{R}}^m$, then a factorization g is also called a *regression of Y on X*, or a *regression of Y on $X_1, \dots, X_m$*. As is true for a conditional expectation, there can be many versions of a factorization and a regression. Note that the concept of a regression is defined without any reference to a specific parameterization and without any restrictions of the function $g\colon (\Omega'_X, \mathscr{A}'_X) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ other than $(\Omega'_X, \mathscr{A}'_X) = (\overline{\mathbb{R}}^m, \overline{\mathscr{B}}_m)$. ◁

In the following definition we introduce the concept of *the linear regression of Y on* $X_1, \dots, X_m$. In this definition, we refer to the covariance matrix of $X = (X_1, \dots, X_m)$, which has been introduced in Equation (3.55).

Definition 4.19 [Linear Regression]
Let the assumptions of Corollary 4.16 hold, let $(\Omega'_X, \mathscr{A}'_X) = (\overline{\mathbb{R}}^m, \overline{\mathscr{B}}_m)$, and let g denote a regression [see Def. 4.17 (ii)].

(i) *A simple regression g is called the linear regression of Y on X, if* $0 < Var(X) < \infty$ *and there are* $\beta_0, \beta_1 \in \mathbb{R}$ *such that*

$$g(X) = \beta_0 + \beta_1 \cdot X. \tag{4.14}$$

(ii) *An m-variate regression g is called the linear regression of Y on* $X_1, \ldots, X_m$*, if the covariance matrix of* $X = (X_1, \ldots, X_m)$ *is invertible and there are* $\beta_0, \beta_1, \ldots, \beta_m \in \mathbb{R}$ *such that*

$$g(X) = \beta_0 + \sum_{i=1}^{m} \beta_i \cdot X_i. \tag{4.15}$$

Remark 4.20 [The Linear Regression is Uniquely Defined] As is true for all factorizations, a regression of Y on $X_1, \ldots, X_m$ is not uniquely defined. Therefore, we talk about *a* factorization and *a* regression, not about *the* factorization and *the* regression. In contrast, if it exists, then the *linear regression* of Y on $X_1, \ldots, X_m$ is a uniquely defined random variable on the probability space $(\overline{\mathbb{R}}^m, \overline{\mathscr{B}}_m, P_X)$. Furthermore, if g is the linear regression of Y on X, then $g(X)$, the composition of $X = (X_1, \ldots, X_m)$ and g, is a uniquely defined random variable on the probability space $(\Omega, \mathscr{A}, P)$. $\lhd$

Remark 4.21 [Linear Regression vs. Linear Quasi-Regression] If $P(X=x) > 0$, then the value $g(x)$ of a linear regression is identical to the conditional expectation value $E(Y|X=x)$, which has been defined in Equation (3.23). In contrast, even if $P(X=x) > 0$, the value $f(x)$ of a *linear quasi-regression* (see Def. 3.43) is not necessarily identical to $E(Y|X=x)$. It has no meaning other than being a value of a linear quasi-regression (see the example in Figure 3.2 and Exercise 4-5). However, if the linear regression of Y on X exists and $P(X=x) > 0$, then $f(x) = g(x) = E(Y|X=x)$. $\lhd$

4.4 Conditional Expectation Value $E(Y|X=x)$

A factorization of $E(Y|X)$ is also used for general definition of the concept of an $(X=x)$-conditional expectation value. In contrast to the elementary definition in Equation (3.23), in this definition we do *not* presume $P(X=x) > 0$.

Definition 4.22 [$(X=x)$-Conditional Expectation Value]
Let the assumptions of Definition 4.4 hold, let $(\Omega'_X, \mathscr{A}'_X)$ *denote the value space of* X*, let* $E(Y|X) \in \mathscr{E}(Y|X)$*, and let* $g\colon (\Omega'_X, \mathscr{A}'_X) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ *be a measurable function satisfying Equation (4.13). Then, for all* $x \in \Omega'_X$*, we define an* $(X=x)$*-conditional expectation value of* Y *by*

$$E(Y|X=x) := g(x). \tag{4.16}$$

Remark 4.23 [Values of a Conditional Expectation] Let $E(Y|X) = g(X) \in \mathscr{E}(Y|X)$, that is, let $E(Y|X) = g(X)$ be a version of the X-conditional expectation of Y, and let $\{X=x\} = \{\omega \in \Omega\colon X(\omega) = x\}$. Then

$$\forall\, x \in X(\Omega), \forall\, \omega \in \Omega\colon \quad E(Y|X)(\omega) \;=\; g(x) \;=\; E(Y|X{=}x), \quad \text{if } \omega \in \{X{=}x\} \tag{4.17}$$

(see SN-Rem. 10.37). ◁

Remark 4.24 [Uniqueness of a Conditional Expectation Value] If $P(X{=}x) > 0$, then the conditional expectation value $E(Y|X{=}x)$ is a uniquely defined real number, and it is identical to the term defined in Equation (3.23) (see SN-Rem. 10.35). This justifies using the same symbol. ◁

In the following theorem we present a proposition that is equivalent to two X-conditional expectations being P-equivalent. According to this theorem we may formulate propositions in terms of conditional expectation values or, equivalently in terms of conditional expectations. (For a proof see SN-Cor. 10.39.)

Theorem 4.25 [Equivalent Propositions]
Let X, Y_1, Y_2 be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Let Y_1 and Y_2 be numerical and nonnegative or with finite expectations, and let $(\Omega'_X, \mathscr{A}'_X)$ denote the value space of X. Then

$$E(Y_1|X) \underset{P}{=} E(Y_2|X) \quad \Leftrightarrow \quad E(Y_1|X{=}x) = E(Y_2|X{=}x), \;\; \textit{for } P_X\textit{-almost all } x \in \Omega'_X. \tag{4.18}$$

Remark 4.26 [Propositions About Conditional Expectation Values] Let the assumptions of Theorem 4.25 hold, let $x \in \Omega'_X$, $\{X{=}x\} \in \mathscr{A}$, and $P(X{=}x) > 0$. Then

$$E(Y_1|X) \underset{P}{=} E(Y_2|X) \quad \Rightarrow \quad E(Y_1|X{=}x) = E(Y_2|X{=}x). \tag{4.19}$$

This proposition immediately follows from Theorem 4.25. ◁

Remark 4.27 [Propositions About a Single Conditional Expectation Value] Assume that X and Y are real-valued random variables on $(\Omega, \mathscr{A}, P)$ and that (X, Y) has a bivariate normal distribution (see SN-sect. 8.2.3). In this case, $P(X{=}x) = 0$, for all $x \in \mathbb{R}$ (see Rem. 2.66), which implies that the conditional expectation value $E(Y|X{=}x)$ is not uniquely defined. If $P(X{=}x) = 0$, then Equation (3.23) does not apply. In contrast, Equation (4.16) still does apply, but different versions $V, V^* \in \mathscr{E}(Y|X)$ have different factorizations g and g^*, and they can have different values $g(x)$ and $g^*(x)$ for the same value x of X. Hence, propositions about such a single conditional expectation value $E(Y|X{=}x)$ are meaningless unless $P(X{=}x) > 0$. Nevertheless, we can formulate meaningful propositions about conditional expectation values $E(Y|X{=}x)$ *for P_X-almost all* $x \in \mathbb{R}$. This will be exemplified for the special case in which (X, Y) has a bivariate normal distribution. ◁

Theorem 4.28 [Linear Regression if (X, Y) Has a Bivariate Normal Distribution]
Assume that X, Y are real-valued random variables on the probability space $(\Omega, \mathscr{A}, P)$, that (X, Y) has a bivariate normal distribution with $-1 < Corr(X, Y) < 1$ and $Var(X)$, $Var(Y) > 0$. Then the conditional expectation $E(Y|X)$ has a linear parameterization, that is, then

$$E(Y|X) \;=\; \beta_0 + \beta_1 \cdot X \tag{4.20}$$

is a version of the X-conditional expectation of Y, where

$$\beta_0 \;=\; E(Y) - \beta_1 \cdot E(X) \qquad \textit{and} \qquad \beta_1 \;=\; \frac{Cov(X,Y)}{Var(X)}. \tag{4.21}$$

(Proof p. 96)

Remark 4.29 [A Proposition About $E(Y|X{=}x)$ for P_X-almost all $x \in \mathbb{R}$] Under the assumptions of Theorem 4.28, we can write

$$E(Y|X{=}x) \;=\; g(x) \;=\; \beta_0 + \beta_1 \cdot x, \quad \text{for } P_X\text{-almost all } x \in \mathbb{R}, \tag{4.22}$$

and this proposition is equivalent to stating that $(\beta_0 + \beta_1 \cdot X) \in \mathscr{E}(Y|X)$ (see Eq. (4.20), Th. 2.43, and Rem. 2.39). In contrast, for a single value x of X, a proposition such as $E(Y|X{=}x) = \beta_0 + \beta_1 \cdot x$ is meaningful only for the specific factorization $g\colon \mathbb{R} \to \mathbb{R}$ defined by

$$g(x) \;=\; \beta_0 + \beta_1 \cdot x, \quad \forall\, x \in \mathbb{R}, \tag{4.23}$$

which is the factorization of the version $(\beta_0 + \beta_1 \cdot X) \in \mathscr{E}(Y|X)$. This factorization g is the *linear regression of Y on X* [see Def. 4.19 (i)]. ◁

Example 4.30 [Another Version of the Conditional Expectation $E(Y|X)$] Under the assumptions of Theorem 4.28, there are infinitely many factorizations of $E(Y|X)$ and it is easy to find a version $V^* \in \mathscr{E}(Y|X)$ and a factorization g^* such that $V^* = g^*(X)$ and

$$E(Y|X{=}x)^* = g^*(x) \;\neq\; \beta_0 + \beta_1 \cdot x = E(Y|X{=}x),$$

for the same value x of X. For example, such a factorization $g^*\colon \mathbb{R} \to \mathbb{R}$ can be defined by

$$g^*(x) \;=\; \begin{cases} \beta_0 + \beta_1 \cdot x, & \forall\, x \in \mathbb{R}, \text{ if } x \in \mathbb{R} \setminus \{10\} \\ \alpha,\ \alpha \in \mathbb{R}, \text{ otherwise}. \end{cases}$$

For any real number $\alpha \neq \beta_0 + \beta_1 \cdot 10$, this defines a new version $V^* = g^*(X) \in \mathscr{E}(Y|X)$ of the X-conditional expectation of Y, and a factorization such that $g^* \neq g$ but still satisfying $g(X) \underset{P}{=} g^*(X)$. However,

$$E(Y|X{=}10) = g(10) \;\neq\; g^*(10) = E(Y|X{=}10)^*.$$

Therefore, a proposition such as $E(Y|X{=}10) = \gamma$, where γ is a real number, is meaningless unless a proposition about the linear regression, which is defined by Equation (4.23), is intended. This linear regression is only the (easy to grasp and communicate) factorization of *one version* of the X-conditional expectation of Y, out of infinitely many versions of the X-conditional expectation of Y and their factorizations. ◁

Remark 4.31 [Uniqueness of a Factorization] Hence, $E(Y|X{=}x)$ is uniquely defined only if $P(X{=}x) > 0$, and in this case Definitions 4.1 and 4.22 are consistent (see SN-Rem. 10.35). Similary, the concept of a regression of Y on X is not uniquely defined. For two elements $V, V^* \in \mathscr{E}(Y|X)$ there can be different factorizations g and g^* with $V = g(X)$ and $V^* = g^*(X)$. This is true even if $V = V^*$. Hence, there can be different factorizations of a single element $V \in \mathscr{E}(Y|X)$ (see SN-Example 10.32). If $g(X) = g^*(X)$ with $g \neq g^*$, then $g(x) = g^*(x)$ for all $x \in X(\Omega)$, whereas $g(x) = g^*(x)$ does *not* hold for all $x \in \Omega'_X$ unless $X(\Omega) = \Omega'_X$. However, SN-Theorem 10.9 (ii) and Theorem 2.43 imply the following corollary: ◁

Corollary 4.32 [P_X**-Equivalence of Factorizations**]
Let the assumptions of Definition 4.4 hold and let $g, g^*\colon (\Omega'_X, \mathscr{A}'_X) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ *be measurable functions. Furthermore, let* $g(X), g^*(X) \in \mathscr{E}(Y|X)$*. Then*

$$g \underset{P_X}{=} g^*. \tag{4.24}$$

Hence, if $g(X), g^*(X) \in \mathscr{E}(Y|X)$, then the factorizations g and g^* are identical, P_X-almost surely. Remember, P_X, the distribution of X, is a probability measure on the value space $(\Omega'_X, \mathscr{A}'_X)$ of X (see Rem. 2.28).

Many properties of conditional expectations are presented, proved, and illustrated in SN-chapters 10 and 11. The most important of these properties are gathered in Box 4.1.

4.5 Residual

Further properties are related to the residual of Y with respect to $E(Y|X)$. Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and nonnegative or with finite expectation $E(Y)$. Furthermore, assume that $E(Y|X) \in \mathscr{E}(Y|X)$ is a real-valued version of the X-conditional expectation of Y. Then we define

$$\varepsilon := Y - E(Y|X) \tag{4.25}$$

and call it the *residual of Y with respect to* $E(Y|X)$. Some important properties are found in Box 4.2.

Remark 4.33 [Some Special Cases] Because $E(Y|X)$ is $\sigma(X)$-measurable, the following equations are special cases of Rules (vi) and (vii) of Box 4.2, respectively.

$$E\big(\varepsilon \,\big|\, E(Y|X)\big) \underset{P}{=} 0, \tag{4.26}$$

$$Cov\big(\varepsilon, E(Y|X)\big) = E\big(\varepsilon \cdot E(Y|X)\big) = 0, \quad \text{if } E(Y^2) < \infty. \tag{4.27}$$

According to Equation (4.26), the conditional expectation of the residual ε given $E(Y|X)$ is 0, P-almost surely. According to the second equation, the residual $\varepsilon = Y - E(Y|X)$ is uncorrelated with $E(Y|X)$ if $E(Y^2) < \infty$. [Note that finiteness of $E\big(E(Y|X)^2\big)$ follows from $E(Y^2) < \infty$; see Box 4.1 (x)].

Another special case of Box 4.2 (vi) is

$$E(\varepsilon\,|\,X) \underset{P}{=} 0. \tag{4.28}$$

Furthermore, if X is real-valued and $E(Y^2), E(X^2) < \infty$, then

$$Cov(\varepsilon, X) = E(\varepsilon \cdot X) = 0, \tag{4.29}$$

is a special case of Rule (vii).

Now let also Z be a random variable on $(\Omega, \mathscr{A}, P)$, assume that $E(Y|X,Z) \in \mathscr{E}(Y|X,Z)$ is a real-valued version of the (X,Z)-conditional expectation of Y, and consider the residual

$$\varepsilon := Y - E(Y|X,Z). \tag{4.30}$$

Box 4.2 Rules of Computation for a Residual

Let X, Y be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and nonnegative or with finite expectation $E(Y)$. Then the following properties hold for all real-valued versions $E(Y|X) \in \mathscr{E}(Y|X)$ and all versions of the residual ε defined in Equation (4.25):

$$\varepsilon \underset{P}{=} Y - E(Y|X) \tag{i}$$

$$Y \underset{P}{=} E(Y|X) + \varepsilon \tag{ii}$$

$$E(\varepsilon) = 0 \tag{iii}$$

$$Var(Y) = Var\big(E(Y|X)\big) + Var(\varepsilon), \quad \text{if } E(Y^2) < \infty \tag{iv}$$

$$\varepsilon \underset{P}{=} 0, \quad \text{if } Y \underset{P}{=} E(Y|X). \tag{v}$$

Additionally, let W be a random variable on $(\Omega, \mathscr{A}, P)$. Then

$$E(\varepsilon|W) \underset{P}{=} 0, \quad \text{if } \sigma(W) \subset \sigma(X). \tag{vi}$$

If W is real-valued, $\sigma(W) \subset \sigma(X)$, and $E(W^2), E(Y^2) < \infty$, then

$$Cov(\varepsilon, W) = E(\varepsilon \cdot W) = 0 \tag{vii}$$

$$Cov\big(W, E(Y|X)\big) = Cov\big(W, E(Y|X) + \varepsilon\big) = Cov(W, Y). \tag{viii}$$

Then

$$E(\varepsilon|X,Z) \underset{P}{=} E(\varepsilon|X) \underset{P}{=} E(\varepsilon|Z) \underset{P}{=} 0 \tag{4.31}$$

are special cases of Box 4.2 (vi). If we additionally assume X and Z to be real-valued and $E(Y^2), E(X^2), E(Z^2) < \infty$, then

$$Cov(X,\varepsilon) = E(X \cdot \varepsilon) = Cov(Z,\varepsilon) = E(Z \cdot \varepsilon) = 0 \tag{4.32}$$

are special cases of Box 4.2 (vii). ◁

4.6 Mean-Independence

The concept of a conditional expectation can be used to define a new kind of independence of a numerical random variable Y from a random variable X.

Definition 4.34 [Mean-Independence of Y from X]
Let the assumptions of Definition 4.4 hold. Then we define mean-independence of Y from X, denoted $Y \vDash X$, by

$$E(Y|X) \underset{P}{=} E(Y). \tag{4.33}$$

Hence,

$$Y \vDash X \quad :\Leftrightarrow \quad E(Y|X) \underset{P}{=} E(Y). \tag{4.34}$$

In the Theorem 4.35 we present an alternative formulation of mean-independence in terms of conditional expectation values. We also consider the special case in which X is discrete or even *binary*, that is, in which the image of X, $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$, has just two different values, that is, in which $X(\Omega) = \{0, 1\}$. Remember, the term 'almost all' (with respect to a probability measure such as the distribution P_X of X), abreviated P_X-a.a., has been introduced in Remark 2.39.

Theorem 4.35 [Mean-Independence and Conditional Expectation Values]
Let the assumptions of Definition 4.4 hold.

(i) Then

$$Y \vDash X \quad \Leftrightarrow \quad E(Y|X{=}x) = E(Y), \quad \textit{for } P_X\textit{-a.a. } x \in \Omega'_X. \tag{4.35}$$

(ii) If, additionally, $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, and $P(X{=}x) > 0$, then

$$Y \vDash X \quad \Rightarrow \quad E(Y|X{=}x) = E(Y). \tag{4.36}$$

(Proof p. 97)

If $X(\Omega)$ is finite or countable and $P(X{=}x) > 0$ for all $x \in X(\Omega)$, then we can use further properties of mean independence [see Th. 4.36 (i)]. If even $X(\Omega) = \{0, 1\}$, then it suffices to consider only one single value x of X [see Th. 4.36 (ii)].

Theorem 4.36 [Mean-Independence and Conditional Expectation Values]
Let the assumptions of Definition 4.4 hold and let $X(\Omega)$ be finite or countable. Furthermore, for all $x \in X(\Omega)$, assume that $\{X{=}x\} \in \mathscr{A}$ and $P(X{=}x) > 0$.

(i) Then

$$Y \vDash X \quad \Leftrightarrow \quad E(Y|X{=}x) = E(Y), \quad \forall\, x \in X(\Omega) \tag{4.37}$$

$$\Leftrightarrow \quad Y \vDash 1_{X=x}, \quad \forall\, x \in X(\Omega) \tag{4.38}$$

$$\Leftrightarrow \quad E(Y|1_{X=x}) \underset{P}{=} E(Y), \quad \forall\, x \in X(\Omega). \tag{4.39}$$

(ii) If, additionally, $X(\Omega) = \{0, 1\}$ and $x \in X(\Omega)$, then

$$Y \vDash X \quad \Leftrightarrow \quad E(Y|X{=}x) = E(Y) \tag{4.40}$$

$$\Leftrightarrow \quad Y \vDash 1_{X=x} \tag{4.41}$$

$$\Leftrightarrow \quad E(Y|1_{X=x}) \underset{P}{=} E(Y). \tag{4.42}$$

(Proof p. 98)

In the following theorem we consider an implication of mean-independence of Y from X for an X-measurable random variable Z.

Theorem 4.37 [An Implication of Mean-Independence]
Let X, Y, Z be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and non-negative or with finite expectation $E(Y)$. Then:

$$\big(E(Y|X) \underset{P}{=} E(Y) \ \wedge \ \sigma(Z) \subset \sigma(X)\big) \quad \Rightarrow \quad E(Y|Z) \underset{P}{=} E(Y). \tag{4.43}$$

(Proof p. 99)

Rewriting Proposition (4.43) in symbols of mean-independence (see Def. 4.34) yields

$$\big(Y \vDash X \ \wedge \ \sigma(Z) \subset \sigma(X)\big) \quad \Rightarrow \quad Y \vDash Z. \tag{4.44}$$

In the following theorem we address the relationship between independence and mean-independence of two random variables.

Theorem 4.38 [Independence Implies Mean-Independence]
Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and non-negative or with finite expectation $E(Y)$. Then

$$Y \perp\!\!\!\perp X \quad \Rightarrow \quad Y \vDash X. \tag{4.45}$$

According to this theorem, independence of Y and X implies mean-independence of Y from X (for a proof, see SN-Rem. 16.36). Note that X does not have to be a numerical random variable. It can also be a multidimensional random variable, consisting on several random variables that do not have to be numerical (see sect. 2.1.5.) For more details on mean-independence see SN-section 10.6.

4.7 Conditional Mean-Independence

Now we generalize the definition of mean-independence of Y from X introducing the concept of conditional mean-independence of Y from X given a random variable, say Z.

Definition 4.39 [Z-Conditional Mean-Independence of Y from X]
Let the assumptions of Definition 4.4 hold and let Z be a random variable on the probability space $(\Omega, \mathscr{A}, P)$. Then we define Z-conditional mean-independence of Y from X, denoted $Y \vDash X \,|\, Z$, by

$$E(Y|X,Z) \underset{P}{=} E(Y|Z). \tag{4.46}$$

Hence,

$$Y \vDash X \,|\, Z \quad :\Leftrightarrow \quad E(Y|X,Z) \underset{P}{=} E(Y|Z). \tag{4.47}$$

In Theorem 4.40 we present an alternative formulation of Z-conditional mean-independence in terms of conditional expectation values. Remember, $P_{X,Z}$ denotes the distribution (see Def. 2.27) of the (multivariate) random variable (X, Z).

Theorem 4.40 [Z-Conditional Mean-Independence]
Let the assumptions of Definition 4.39 hold and let $(\Omega'_X, \mathscr{A}'_X)$, $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value spaces of X and Z.

(i) Then $Y \vDash X \,|\, Z$ is equivalent to

$$E(Y|X{=}x, Z{=}z) \;=\; E(Y|Z{=}z), \quad \text{for } P_{X,Z}\text{-a.a. } (x,z) \in \Omega'_X \times \Omega'_Z. \tag{4.48}$$

(ii) If, additionally, $\{X{=}x, Z{=}z\} \in \mathscr{A}$, and $P(X{=}x, Z{=}z) > 0$, then $Y \vDash X \,|\, Z$ implies

$$E(Y|X{=}x, Z{=}z) \;=\; E(Y|Z{=}z). \tag{4.49}$$

(Proof p. 99)

In the following theorem we consider an implication of Z-conditional mean-independence of Y from X for an X-measurable random variable W.

Theorem 4.41 [An Implication of Z-Conditional Mean-Independence]
Let W, X, Y, Z be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and nonnegative or with finite expectation $E(Y)$. Then:

$$\big(E(Y|X,Z) \underset{P}{=} E(Y|Z) \;\wedge\; \sigma(W) \subset \sigma(X)\big) \quad \Rightarrow \quad E(Y|W,Z) \underset{P}{=} E(Y|Z). \tag{4.50}$$

(Proof p. 99)

For more details on mean-independence and conditional mean-independence see SN-section 10.6

4.8 Summary and Conclusions

In this chapter we introduced the concept of an *X-conditional expectation* $E(Y|X)$ of a numerical random variable Y, that is, a random variable with value space $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$. This concept is a tool to describe how the conditional expectation values $E(Y|X{=}x)$ depend on the values x of X. Because X can be multidimensional, such a conditional expectation $E(Y|X)$ and its values $E(Y|X{=}x)$ are what we try to estimate in many statistical procedures, for example, in regression analysis, analysis of variance, structural equation modeling, multilevel analysis, etc. We also introduced related concepts such as a factorization of a conditional expectation, a regression, the linear regression, a residual with respect to a conditional expectation, mean-independence, and conditional mean-independence (see Box 4.3).

Box 4.3 Glossary of new concepts

$E(Y \mid X)$	A (version of the) *X-conditional expectation of Y* (with respect to the probability measure P). Let X and Y be random variables on a probability space $(\Omega, \mathscr{A}, P)$. If Y is numerical and nonnegative or has a finite expectation, then $E(Y \mid X)$ is defined as a random variable V on $(\Omega, \mathscr{A}, P)$ satisfying (a) $\sigma(V) \subset \sigma(X)$ and (b) $\forall\, C \in \sigma(X)\colon E(1_C \cdot V) = E(1_C \cdot Y)$.
$\mathscr{E}(Y \mid X)$	The set of all random variables V on $(\Omega, \mathscr{A}, P)$ satisfying (a) and (b).
P*-unique*	In general, $E(Y \mid X)$ is not uniquely defined. However, it is *P-unique* in the following sense: If $V, V^* \in \mathscr{E}(Y \mid X)$, then $V \underset{P}{=} V^*$.
g	*A factorization of* $E(Y \mid X)$. It is a function on Ω'_X, the co-domain of X, satisfying $E(Y \mid X) = g(X)$. If $\Omega'_X = \mathbb{R}^m$, then g is also called the *(m-variate) regression of Y on X*. An m-variate regression is called *the linear regression* of Y on $X_1, \dots, X_m$, if the inverse of the covariance matrix of $X = (X_1, \dots, X_m)$ exists and there are $\beta_0, \beta_1, \dots, \beta_m \in \mathbb{R}$ such that $g(X) = \beta_0 + \beta_1 \cdot X_1 + \ldots + \beta_m \cdot X_m$.
$E(Y \mid X{=}x)$	*$(X{=}x)$-conditional expectation value of Y*. It is defined by $E(Y \mid X{=}x) := g(x)$, where g is a factorization of $E(Y \mid X)$. It is also defined if $P(X{=}x) = 0$. However, it is *uniquely* defined only if $P(X{=}x) > 0$.
ε	*The residual of Y with respect to* $E(Y \mid X)$. Presuming that $E(Y \mid X)$ is a real-valued version of the X-conditional expectation of Y, it is defined by $\varepsilon := Y - E(Y \mid X)$.
$Y \vDash X$	*Mean-independence of Y from X*. It is defined by $E(Y \mid X) \underset{P}{=} E(X)$.
$Y \vDash X \mid Z$	*Z-conditional mean-independence of Y from X*. Assuming that Z is also a random variable on $(\Omega, \mathscr{A}, P)$, it is defined by $E(Y \mid X, Z) \underset{P}{=} E(Y \mid X)$.

The examples presented so far show that in some applications (such as an experiment with randomized assignment of an observational unit to a treatment), conditional expectations can be used to describe causal dependencies [see $E(Y \mid X)$ in Table 1.2], whereas in other applications (such as an observational study with systematic self-selection) they can totally lead us astray if such a causal interpretation of a conditional expectation is intended [see $E(Y \mid X)$ in Table 1.4].

4.9 Proofs

Proof of Lemma 4.13

SN-Box 10.3 (ii) implies $Y \underset{P}{\geq} \alpha \;\Rightarrow\; E(Y \mid X) \underset{P}{\geq} \alpha$. Therefore, if $Y \underset{P}{\geq} \alpha$, then $V \underset{P}{\geq} \alpha$ for all $V \in \mathscr{E}(Y \mid X)$. Now consider the event $C \in \sigma(X)$ defined by $C := \{\omega \in \Omega\colon V(\omega) = \alpha\}$. We prove $P(C) = 0$ by contradiction. Hence, assume $P(C) > 0$. Then

$$\begin{aligned}\alpha\cdot P(C) &= \int 1_C\,\alpha\,dP && \text{[SN-Th. 3.36 (i), SN-(3.9)]}\\ &= \int 1_C\,V\,dP && [V(\omega)=\alpha \text{ if } \omega\in C]\\ &= \int 1_C\,Y\,dP && \text{[(3.3), Def. 4.4 (b)]}\\ &> \alpha\cdot P(C). && \text{[SN-Th. 3.52 (ii)]}\end{aligned}$$

This contradiction proves $P(C) = 0$ and Proposition (4.12). Note that SN-Theorem 3.52 can be applied because $(C, \mathscr{A}|_C, P|_C)$ is a measure space, where $\mathscr{A}|_C$ denotes the trace σ-algebra of $\mathscr{A}$ in C (see SN-Example 1.10), and $P|_C$ denotes the restriction of P on C (see SN-Example 1.61).

Proof of Theorem 4.28

According to SN-Lemma 17.73

$$E(Y|X{=}x) = \int y\, f_{Y|X=x}(y)\,dy, \quad \forall x\in\mathbb{R}, \tag{4.51}$$

defines a version of the X-conditional expectation of Y, where

$$f_{Y|X=x}(y) := \begin{cases}\dfrac{f_{X,Y}(x,y)}{f_X(x)}, & \text{if } f_X(x)>0,\\ 0, & \text{otherwise,}\end{cases} \quad \forall\, y\in\mathbb{R}, \tag{4.52}$$

(see SN-Def. 17.75). First we compute the denominator of the fraction occurring in Equation (4.52), then the numerator, and then combine the results in order to obtain $E(Y|X{=}x)$ via Equation (4.51).

Define $\mu_1 := E(X)$, $\mu_2 := E(Y)$, $\sigma_1 := SD(X)$, $\sigma_2 := SD(Y)$, $z_1 := (x-\mu_1)/\sigma_1$, and $z_2 := (y-\mu_2)/\sigma_2$, as well as $\rho = Corr(X,Y)$. Furthermore, we define

$$z := -\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{2(1-\rho^2)}. \tag{4.53}$$

Using this notation,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy && \text{[SN-(5.49)]}\\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\exp(z)\,\sigma_2\,dz_2 && [\text{SN-(8.35)}, dy=\sigma_2\,dz_2]\\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}}\cdot\exp(z)\,dz_2 \end{aligned} \tag{4.54}$$

is a density of X that can be used as the denominator in Equation (4.52).

Now we turn to a term involving the numerator of the fraction occurring in Equation (4.52) using

$$z_2 \;=\; -(1-\rho^2)\left(\frac{-1}{1-\rho^2}(z_2-\rho\, z_1)\right)+\rho\, z_1 \tag{4.55}$$

and

$$\alpha \;:=\; 2\pi\,\sigma_1\sqrt{1-\rho^2}\,. \tag{4.56}$$

$$\begin{aligned}
&\int_{-\infty}^{\infty} y\, f_{X,Y}(x,y)\,dy \\
&= \int_{-\infty}^{\infty} (\sigma_2\, z_2+\mu_2)\cdot \frac{1}{2\pi\,\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot \exp(z)\,\sigma_2\, dz_2 && [(4.53),\ \text{SN-}(8.35)] \\
&= \frac{\sigma_2}{\alpha}\cdot\int_{-\infty}^{\infty} z_2 \exp(z)\,dz_2 \;+\; \frac{\mu_2}{\alpha}\cdot\int_{-\infty}^{\infty}\exp(z)\,dz_2 && [(4.56)] \\
&= \frac{\sigma_2}{\alpha}\cdot\int_{-\infty}^{\infty}\left(-(1-\rho^2)\left(\frac{-1}{1-\rho^2}(z_2-\rho\, z_1)\right)+\rho\, z_1\right)\exp(z)\,dz_2+\mu_2\cdot f_X(x) && [(4.54),\ (4.55)] \\
&= \frac{\sigma_2}{\alpha}\cdot\big(-(1-\rho^2)\big)\left[\exp\left(-\frac{z_1^2-2\rho\, z_1 z_2+z_2^2}{2(1-\rho^2)}\right)\right]_{z_2=-\infty}^{z_2=\infty} \\
&\quad+\;\frac{\sigma_2}{\alpha}\cdot\rho\, z_1\int_{-\infty}^{\infty}\exp(z)\,dz_2+\mu_2\cdot f_X(x) \\
&= 0 \;+\; \sigma_2\,\rho\, z_1\, f_X(x) \;+\; \mu_2\, f_X(x) && [(4.54)] \\
&= f_X(x)\left(\mu_2+\sigma_2\,\rho\cdot\frac{x-\mu_1}{\sigma_1}\right) \\
&= f_X(x)\left(E(Y)-\frac{Cov(X,Y)}{Var(X)}\cdot E(X) \;+\; \frac{Cov(X,Y)}{Var(X)}\cdot x\right) \\
&= f_X(x)\,\big(\beta_0+\beta_1\cdot x\big),
\end{aligned}$$

where $\beta_0 = E(Y)-\beta_1\cdot E(X)$ and $\beta_1 = Cov(X,Y)/Var(X)$. Substituting these results into Equation (4.51) yields

$$\begin{aligned}
\forall x\in\mathbb{R}:\quad E(Y|X{=}x) &= \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\,dy \;=\; \int_{-\infty}^{\infty}\frac{y\, f_{X,Y}(x,y)}{f_X(x)}\,dy \\
&= \frac{f_X(x)\,\big(\beta_0+\beta_1 x\big)}{f_X(x)} \;=\; \beta_0+\beta_1 x.
\end{aligned}$$

According to SN-Lemma 17.73, this defines a version of the X-conditional expectation of Y, namely $E(Y|X)=\beta_0+\beta_1 X$.

Proof of Theorem 4.35

Proposition (i). We may consider the expectation $E(Y)$ as a constant mapping with domain Ω (see first line below) and also with domain Ω'_X (see lines two and three below). Hence,

$$Y \vDash X \quad\Leftrightarrow\quad E(Y|X) \underset{P}{=} E(Y) \qquad [(4.34)]$$

$$\begin{aligned}
&\Leftrightarrow \quad g \underset{P_X}{=} E(Y) && [E(Y|X) \underset{P}{=} g(X),\ (2.34)]\\
&\Leftrightarrow \quad g(x) = E(Y)(x), \ \text{for } P_X\text{-a.a. } x \in \Omega'_X && [(2.31)]\\
&\Leftrightarrow \quad E(Y|X{=}x) = E(Y), \ \text{for } P_X\text{-a.a. } x \in \Omega'_X. && [(4.16)]
\end{aligned}$$

Proposition (ii). According to Equation (2.32), this proposition immediately follows from Proposition (i).

Proof of Theorem 4.36

Proposition (4.37). According to Equation (2.32), this proposition immediately follows from Proposition (i) of Theorem 4.35.

Proposition (4.38). We start proving $Y \vDash X \ \Rightarrow\ \big(Y \vDash 1_{X=x}, \forall\, x \in X(\Omega)\big)$.

$$\begin{aligned}
\forall\, x \in X(\Omega):\ \ E(Y|1_{X=x}) &\underset{P}{=} E\big(E(Y|X)\,|1_{X=x}\big) && [\textit{Box 4.1 (xii)}]\\
&\underset{P}{=} E\big(E(Y)\,|1_{X=x}\big) && [\text{Def. } 4.34,\ Y \vDash X]\\
&= E(Y). && [\textit{Box 4.1 (i)}]
\end{aligned}$$

According to Definition 4.34, this proposition is equivalent to $Y \vDash 1_{X=x}, \forall\, x \in X(\Omega)$.

Now we show $\big(Y \vDash 1_{X=x}, \forall\, x \in X(\Omega)\big) \ \Rightarrow\ Y \vDash X$. First of all, note that

$$\forall\, x \in X(\Omega):\ E(Y|1_{X=x}) \cdot 1_{X=x} \underset{P}{=} E(Y|X{=}x) \cdot 1_{X=x}. \tag{4.57}$$

Hence,

$$\begin{aligned}
E(Y|X) &\underset{P}{=} \sum_{x \in X(\Omega)} E(Y|X{=}x) \cdot 1_{X=x} && [(4.1)]\\
&\underset{P}{=} \sum_{x \in X(\Omega)} E(Y|1_{X=x}) \cdot 1_{X=x} && [(4.57)]\\
&= \sum_{x \in X(\Omega)} E(Y) \cdot 1_{X=x} && \big[Y \vDash 1_{X=x}, \forall\, x \in X(\Omega)\big]\\
&= E(Y) \sum_{x \in X(\Omega)} 1_{X=x} && \\
&= E(Y). && \big[\textstyle\sum_{x \in X(\Omega)} 1_{X=x} = 1\big]
\end{aligned}$$

Proposition (4.39). According to Definition 4.34, this proposition is equivalent to $Y \vDash 1_{X=x}, \forall\, x \in X(\Omega)$.

Proposition (4.40). $Y \vDash X \ \Rightarrow\ E(Y|X{=}x) = E(Y)$, immediately follows from Theorem 4.35 (i), because we assume $P(X{=}x) > 0$. Therefore, we only have to prove

$$E(Y|X{=}x) = E(Y) \quad \Rightarrow \quad Y \vDash X,$$

provided that $x \in X(\Omega) = \{0, 1\}$ and $P(X{=}0), P(X{=}1) > 0$. Without restricting generality, define $x = 0$. Then

$$\begin{aligned}
& E(Y) = E(Y|X{=}0) \cdot P(X{=}0) + E(Y|X{=}1) \cdot P(X{=}1) && [(3.38)]\\
\Rightarrow\quad & E(Y) = E(Y) \cdot P(X{=}0) + E(Y|X{=}1) \cdot P(X{=}1) && [E(Y|X{=}0) = E(Y)]\\
\Rightarrow\quad & E(Y) \cdot \big(1 - P(X{=}0)\big) = E(Y|X{=}1) \cdot P(X{=}1) && \\
\Rightarrow\quad & E(Y) \cdot P(X{=}1) = E(Y|X{=}1) \cdot P(X{=}1) && [P(X{=}1) = 1 - P(X{=}0)]
\end{aligned}$$

$$\Rightarrow \quad E(Y) = E(Y \mid X{=}1). \qquad [P(X{=}1) > 0]$$

Hence, we have shown that $E(Y) = E(Y \mid X{=}0)$ implies $E(Y) = E(Y \mid X{=}1)$. Because we assume and $P(X{=}0), P(X{=}1) > 0$, Proposition (4.37) then implies $Y \vDash X$.

Proposition (4.41). If $X(\Omega) = \{0, 1\}$, then $\sigma(X) = \sigma(1_{X=x})$ and

$$\begin{aligned} Y \vDash X \quad &\Leftrightarrow \quad E(Y \mid X) \underset{P}{=} E(Y) && [\text{Def. } 4.34] \\ &\Leftrightarrow \quad E(Y \mid 1_{X=x}) \underset{P}{=} E(Y) && [\sigma(X) = \sigma(1_{X=x}), \text{Def. } 4.4 \text{ (b)}] \\ &\Leftrightarrow \quad Y \vDash 1_{X=x} && [\text{Def. } 4.34] \end{aligned}$$

Proposition (4.42). $Y \vDash 1_{X=x}$ is equivalent to $E(Y \mid 1_{X=x}) \underset{P}{=} E(Y)$ (see Def. 4.34).

Proof of Theorem 4.37

$$\begin{aligned} E(Y \mid Z) &\underset{P}{=} E\big(E(Y \mid X) \,\big|\, Z\big) && [\text{Box } 4.1 \text{ (xii)}] \\ &\underset{P}{=} E\big(E(Y) \,\big|\, Z\big) && [E(Y \mid X) \underset{P}{=} E(Y)] \\ &\underset{P}{=} E(Y). && [\text{Box } 4.1 \text{ (i)}] \end{aligned}$$

Proof of Theorem 4.40

(i). Let $f \colon \Omega'_Z \to \overline{\mathbb{R}}$ denote a factorization of $E(Y \mid Z)$, so that $E(Y \mid Z) = f(Z)$ [see Eq. (4.13)]. Then

$$\begin{aligned} & Y \vDash X \mid Z \\ \Leftrightarrow \quad & E(Y \mid X, Z) \underset{P}{=} E(Y \mid Z) && [(4.47)] \\ \Leftrightarrow \quad & g(X, Z) \underset{P}{=} f(Z) && [(4.13)] \\ \Leftrightarrow \quad & g(x, z) = f(z), \quad \text{for } P_{X,Z}\text{-a.a. } (x, z) \in \Omega'_X \times \Omega'_Z && [(2.31)] \\ \Leftrightarrow \quad & E(Y \mid X{=}x, Z{=}z) = E(Y \mid Z{=}z), \quad \text{for } P_{X,Z}\text{-a.a. } (x, z) \in \Omega'_X \times \Omega'_Z. && [(4.16)] \end{aligned}$$

(ii). This proposition follows from (i) and $P(X{=}x, Z{=}z) > 0$.

Proof of Theorem 4.41

$$\begin{aligned} E(Y \mid W, Z) &\underset{P}{=} E\big(E(Y \mid X, Z) \,\big|\, W, Z\big) && [\sigma(W, Z) \subset \sigma(X, Z), \text{Box } 4.1 \text{ (xii)}] \\ &\underset{P}{=} E\big(E(Y \mid Z) \,\big|\, Z\big) && [E(Y \mid X, Z) \underset{P}{=} E(Y \mid Z)] \\ &\underset{P}{=} E(Y \mid Z). && [\text{Box } 4.1 \text{ (xi)}] \end{aligned}$$

4.10 Exercises

▷ **Exercise 4-1** Let X and Y be real-valued random variables on a probability space $(\Omega, \mathscr{A}, P)$, that is, assume that $(\mathbb{R}, \mathscr{B})$ is the value space of Y and X. Furthermore, assume that $X(\Omega) = \{x_1, \ldots, x_n\}$ and $P(X{=}x) > 0$ for all $x \in X(\Omega)$. Specify an element of the set $\mathscr{E}(Y|X)$ (see Rem. 4.5) that is not identical to $E(Y|X)$ defined in Equation (4.2).

▷ **Exercise 4-2** Prove Rule (vi) of Box 4.1.

▷ **Exercise 4-3** Why is the conditional expectation value $E(Y|X{=}1, U{=}\mathit{Joe})$ not uniquely defined in the example presented in Table 2.1?

▷ **Exercise 4-4** Consider the example presented in Table 1.2. Show that the random variable $E(Y|X) = P(Y{=}1|X)$ specified in this table satisfies conditions (a) and (b) of Definition 4.4.

▷ **Exercise 4-5** Compute the four values of the conditional expectation $E(Y|X)$ in Example 3.46 and compare them to the values of the linear quasi-regression in Figure 3.2.

▷ **Exercise 4-6** Compute the values of the conditional expectation $E(Y|X)$ in the example presented in Table 2.1.

▷ **Exercise 4-7** Compute the values of the conditional expectation $E(Y|X,U)$ in the example presented in Table 2.1.

Solutions

▷ **Solution 4-1** Let $\alpha \in \mathbb{R} \setminus X(\Omega)$, let α be a real number that is not an element of the image $X(\Omega)$. Then $P(X{=}\alpha) = 0$. Now define

$$V \underset{P}{=} \alpha \cdot 1_{X=\alpha} + \sum_{i=1}^{n} E(Y|X{=}x_i) \cdot 1_{X=x_i}.$$

Then $V \in \mathscr{E}(Y|X)$ and $V \underset{P}{=} E(Y|X) = \sum_{i=1}^{n} E(Y|X{=}x_i) \cdot 1_{X=x_i}$, but $V \neq E(Y|X)$.

▷ **Solution 4-2**

$$X \underset{P}{=} \alpha \quad \Rightarrow \quad X \perp\!\!\!\perp Y \qquad \text{[Box 2.1 (iii)]}$$

$$\Rightarrow \quad E(Y|X) \underset{P}{=} E(Y). \qquad \text{[Box 4.1 (v)]}$$

▷ **Solution 4-3** In this example, $P(X{=}1, U{=}\mathit{Joe}) = 0$. This implies that the conditional probabilities $P(Y{=}y|X{=}1, U{=}\mathit{Joe})$ that are used in the definition of $E(Y|X{=}1, U{=}\mathit{Joe})$ [see Eqs. (3.23) and (3.27)] are not defined. If we use the version $E(Y|X,U) = P(Y{=}1|X{=}1, U{=}\mathit{Joe})$ specified in Table 2.1, then its factorization g and Definition 4.22 yield

$$E(Y|X{=}1, U{=}\mathit{Joe}) = 99.$$

If, in contrast, we use the version $E(Y|X,U)^* = P(Y{=}1|X{=}1, U{=}\mathit{Joe})^*$ also specified in Table 2.1, then its factorization g^* and Definition 4.22 yield

$$E(Y|X{=}1, U{=}\mathit{Joe}) = .8.$$

This exemplifies that the conditional expectation value $E(Y|X{=}1, U{=}\mathit{Joe})$ is not uniquely defined in this example. Therefore, it does not make sense to formulate propositions about this specific value. However, propositions about $P_{X,U}$-almost all values $(x,u) \in \Omega'_X \times \Omega_U$ are meaningful. An example for such a proposition is: $0 < E(Y|X{=}x, U{=}u) < 1$, for $P_{X,U}$-almost all pairs $(x,u) \in \Omega'_X \times \Omega_U$.

▷ **Solution 4-4** According to Definition 4.4, $E(Y|X) = P(Y{=}1|X)$ is a numerical random variable, that is, the value space of $E(Y|X)$ is $(\overline{\mathbb{R}}, \overline{\mathscr{B}})$. Hence, we consider all inverse images of the sets $B \in \overline{\mathscr{B}}$ under $E(Y|X)$. In this example, the σ-algebra generated by $E(Y|X)$ consists of the following four inverse images:

$$\forall B \in \overline{\mathscr{B}}: \quad E(Y|X)^{-1}(B) = \begin{cases} \Omega_U \times \{no\} \times \Omega_Y = \{\omega_1, \omega_2, \omega_5, \omega_6\}, & \text{if } .6 \notin B \text{ and } .45 \in B \\ \Omega_U \times \{yes\} \times \Omega_Y = \{\omega_3, \omega_4, \omega_7, \omega_8\}, & \text{if } .6 \in B \text{ and } .45 \notin B \\ \Omega = \{\omega_1, \dots, \omega_8\}, & \text{if } .6 \in B \text{ and } .45 \in B \\ \emptyset, & \text{if } .6 \notin B \text{ and } .45 \notin B \end{cases}$$

(see Table 1.2). These four inverse images are identical to the elements of $\sigma(X)$:

$$\forall B \in \overline{\mathscr{B}}: \quad X^{-1}(B) = \begin{cases} C_1 = \Omega_U \times \{no\} \times \Omega_Y = \{\omega_1, \omega_2, \omega_5, \omega_6\}, & \text{if } 1 \notin B \text{ and } 0 \in B \\ C_2 = \Omega_U \times \{yes\} \times \Omega_Y = \{\omega_3, \omega_4, \omega_7, \omega_8\}, & \text{if } 1 \in B \text{ and } 0 \notin B \\ C_3 = \Omega = \{\omega_1, \dots, \omega_8\}, & \text{if } 1 \in B \text{ and } 0 \in B \\ C_4 = \emptyset, & \text{if } 1 \notin B \text{ and } 0 \notin B \end{cases}$$

(see again Table 1.2). Hence, $\sigma\big(E(Y|X)\big) = \sigma(X)$, and this implies that condition (a) of Definition 4.4 holds.

Now we show that condition (b) of Definition 4.4 holds as well. Using Equation (3.5) for the random variable $1_{C_1} \cdot E(Y|X)$ and the probabilities listed in Table 1.2 yields

$$E\big(1_{C_1} \cdot E(Y|X)\big) = 1 \cdot .45 \cdot (.09 + .21 + .24 + .06) = .45 \cdot .6 = .27.$$

Using the same formula and the same table for $1_{C_1} \cdot Y$ yields

$$E(1_{C_1} \cdot Y) = 1 \cdot 1 \cdot (.21 + .06) = .27.$$

Hence $E\big(1_{C_1} \cdot E(Y|X)\big) = E(1_{C_1} \cdot Y)$ holds for $C_1 \in \sigma(X)$. The analog computations for C_2, C_3, and C_4 show that this equation also holds for the sets C_2, C_3, and C_4 in $\sigma(X)$. Hence, condition (b) of Definition 4.4 is satisfied as well. This proves that the random variable $E(Y|X)$ specified in Table 1.2 is in fact a version of the X-conditional expectation of Y.

▷ **Solution 4-5**

$$\begin{aligned} E(Y|X{=}1) &= \sum_{i=1}^{4} y_i \cdot P(Y{=}y_i|X{=}1) \\ &= 1 \cdot P(Y{=}1|X{=}1) + 2 \cdot P(Y{=}2|X{=}1) + 0 + 0 \\ &= 1 \cdot \frac{P(Y{=}1, X{=}1)}{P(X{=}1)} + 2 \cdot \frac{P(Y{=}2, X{=}1)}{P(X{=}1)} \\ &= 1 \cdot \frac{1/8}{1/4} + 2 \cdot \frac{1/8}{1/4} \\ &= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = 1.5. \end{aligned}$$

The corresponding equations yield

$$E(Y|X{=}2) = 3.5, \qquad E(Y|X{=}3) = 3.5, \qquad \text{and} \qquad E(Y|X{=}4) = 1.5.$$

In contrast, the values of the linear quasi-regression in Figure 3.2 are identical to 2.5. In this example, the linear regression of Y on X does not exist. In contrast, the linear regression of Y on (X, X^2) *does* exist (see Exercise 3-7).

▷ **Solution 4-6** The values of the conditional expectation $E(Y|X)$ are the two conditional expectation values $E(Y|X=0)$ and $E(Y|X=1)$. Because $E(Y|X=x) = P(Y=1|X=x)$, they can be computed from Table 2.1 as follows:

$$P(Y=1|X=0) = \frac{P(Y=1, X=0)}{P(X=0)} = \frac{.35+.06}{.15+.35+.24+.06} = .5125,$$

$$P(Y=1|X=1) = \frac{P(Y=1, X=1)}{P(X=1)} = \frac{0+.08}{0+0+.12+.08} = .4.$$

▷ **Solution 4-7** The values of the conditional expectation $E(Y|X,U)$ in Table 2.1 are the four conditional expectation values $E(Y|X=x, U=u)$. Because $E(Y|X=x, U=u) = P(Y=1|X=x, U=u)$, they can be computed as follows:

$$P(Y=1|X=0, U=Joe) = \frac{P(Y=1, X=0, U=Joe)}{P(X=0, U=Joe)} = \frac{.35}{.15+.35} = .7,$$

$$P(Y=1|X=0, U=Ann) = \frac{P(Y=1, X=0, U=Ann)}{P(X=0, U=Ann)} = \frac{.06}{.24+.06} = .2,$$

$$P(Y=1|X=1, U=Ann) = \frac{P(Y=1, X=1, U=Ann)}{P(X=1, U=Ann)} = \frac{.08}{.12+.08} = .4.$$

The conditional expectation value $E(Y|X=1, U=Joe) = P(Y=1|X=1, U=Joe)$ is undefined, because $P(X=1, U=Joe) = 0$. Choosing any number (such as 99) as a value of $E(Y|X,U)$ for $\omega_3 = (Joe, yes, -)$ and $\omega_4 = (Joe, yes, +)$ yields a version of $E(Y|X,U)$. Different versions of $E(Y|X,U)$ are identical almost surely with respect to the measure P [see Eq. (4.7)].

Chapter 5
Conditional Expectation With Respect to a Conditional-Probability Measure

In this chapter, we introduce the concept of a *conditional expectation* $E^{X=x}(Y|Z)$ *with respect to a conditional-probability measure* $P^{X=x}$, assuming $P(X=x) > 0$. A conditional expectation $E^{X=x}(Y|Z)$ with respect to $P^{X=x}$ has the general properties of a conditional expectation, but also additional properties that are related to its uniqueness. We also introduced the more general concept of a *partial conditional expectation* $E(Y|X=x, Z)$, whose definition does not presume $P(X=x) > 0$. Both concepts formalize the same intuitive idea: describing how the conditional expectation of Y depends on a random variable Z if another random variable X is kept constant at one of its values x. Among other things, we study the conditions under which the expectation (with respect to P) of $E^{X=x}(Y|Z)$, that is, $E\big(E^{X=x}(Y|Z)\big)$, is a uniquely defined number.

5.1 Conditional Expectation $E^{X=x}(Y|Z)$ With Respect to $P^{X=x}$

In this section we introduce the concept of a conditional expectation with respect to a probability measure $P^{X=x}$. Assume that P is a probability measure on $(\Omega, \mathscr{A})$, let $x \in X(\Omega)$, $\{X=x\} \in \mathscr{A}$, and $P(X=x) > 0$. Then the measure $P^{X=x}: \mathscr{A} \to [0,1]$ is defined by

$$P^{X=x}(A) \ = \ P(A|X=x) \ = \ \frac{P(A \cap \{X=x\})}{P(X=x)}, \quad \forall\, A \in \mathscr{A}. \tag{5.1}$$

Equation (5.1) shows how $P^{X=x}$ and P are related to each other. According to this equation, not only P, but also $P^{X=x}$ is a probability measure on $(\Omega, \mathscr{A})$, and it satisfies

$$P^{X=x}(A) \ = \ 0, \quad \forall\ A \in \mathscr{A} \text{ with } A \cap \{X=x\} = \emptyset. \tag{5.2}$$

Note that $P^{X=x}$ is a special case of the probability measure P^B introduced in Equation (1.21) for the event $B = \{X=x\} = \{\omega \in \Omega \colon X(\omega) = x\}$.

Example 5.1 [No Treatment for Joe] The measures P, $P^{X=0}$, and $P^{X=1}$ specified in columns two, three, and four of Table 2.1 illustrate the relationship between the measures P, $P^{X=0}$, and $P^{X=1}$ for the two values $x = 0, 1$ of the treatment variable X. ◁

Remark 5.2 [Notation $E^{X=x}(Y)$] The expectation of a numerical random variable Y on $(\Omega, \mathscr{A}, P^{X=x})$ with respect to the measure $P^{X=x}: \mathscr{A} \to [0,1]$ is defined by

$$E^{X=x}(Y) \ = \int Y\, dP^{X=x} \tag{5.3}$$

This expectation is well-defined if we presume that $\int Y^+ dP^{X=x}$ or $\int Y^- dP^{X=x}$ is finite (see Def. 3.1). ◁

Remark 5.3 [Discrete Y] If Y is discrete, then Equation (5.3) simplifies to

$$E^{X=x}(Y) = \sum_{y \in Y(\Omega)} y \cdot P^{X=x}(Y=y) \tag{5.4}$$

[see Eq. (3.7)]. If $Y(\Omega) = \{y_1, \ldots, y_n\}$, then this equation can also be written

$$E^{X=x}(Y) = \sum_{i=1}^{n} y_i \cdot P^{X=x}(Y=y_i). \tag{5.5}$$

◁

Referring to the measure $P^{X=x}$ and the expectation $E^{X=x}(Y)$ with respect to this measure, a conditional expectation with respect to $P^{X=x}$ can now be introduced as follows.

Definition 5.4 [Z-**Conditional Expectation With Respect to** $P^{X=x}$]
Let V_x, X, Y, and Z be random variables on $(\Omega, \mathscr{A}, P^{X=x})$, where $P^{X=x}$ is a probability measure on $\mathscr{A}$ defined by Equation (5.1). Hence, we assume $x \in X(\Omega)$, $\{X=x\} \in \mathscr{A}$, and $P(X=x) > 0$. Furthermore, assume that V_x and Y are numerical and that Y is nonnegative or has a finite expectation $E^{X=x}(Y)$ with respect to $P^{X=x}$. Then V_x is called a (version of the) Z-conditional expectation of Y with respect to $P^{X=x}$, if:

(a) $\sigma(V_x) \subset \sigma(Z)$.

(b) $E^{X=x}(1_C \cdot V_x) = E^{X=x}(1_C \cdot Y), \quad \forall\, C \in \sigma(Z)$.

If V_x satisfies (a) and (b), then we also use the notation $E^{X=x}(Y|Z) := V_x$.

Remark 5.5 [$E^{X=x}(Y|Z)$ **is a Conditional Expectation**] Comparing Definitions 4.4 and 5.4 to each other shows that we only replaced the measure P (which is used to define an expectation and a conditional expectation with respect to P) by the measure $P^{X=x}$, which is used to define an expectation with respect to the measure $P^{X=x}$ [see Eq. (5.3)]. Hence, $E^{X=x}(Y|Z)$ has all properties of a Z-conditional expectation of Y, provided that we replace the measure P by $P^{X=x}$ and the expectation $E(\,)$ with respect to P by the expectation $E^{X=x}(\,)$ with respect to $P^{X=x}$.

For example, the property $E\big(Y\big|E(Y|X)\big) \underset{P}{=} E(Y|X)$ [see Box 4.1 (vii)] can be translated to

$$E^{X=x}\big(Y\big|E^{X=x}(Y|Z)\big) \underset{P^{X=x}}{=} E^{X=x}(Y|Z). \tag{5.6}$$

Similarly, $E\big(E(Y|X)\big) = E(Y)$ [see Box 4.1 (iv)] can be translated to

$$E^{X=x}\big(E^{X=x}(Y|Z)\big) = E^{X=x}(Y). \tag{5.7}$$

◁

Remark 5.6 [Conditional Probability With Respect to $P^{X=x}$ Given a Random Variable] For convenience, we also explicitly introduce the Z-conditional probability with respect to $P^{X=x}$. Consider the probability space $(\Omega, \mathscr{A}, P^{X=x})$ and let 1_A denote the indicator of the event $A \in \mathscr{A}$. We introduce the notation

$$P^{X=x}(A|Z) := E^{X=x}(1_A|Z) \tag{5.8}$$

and call it the *Z-conditional probability of (the event) A* (with respect to $P^{X=x}$). Furthermore, considering the event $\{Y=y\} = \{\omega \in \Omega\colon Y(\omega) = y\}$, we also use the notation

$$P^{X=x}(Y=y|Z) := P^{X=x}(\{Y=y\}|Z) = E^{X=x}(1_{Y=y}|Z), \tag{5.9}$$

and call it the *Z-conditional probability of (the event)* $\{Y=y\}$ (with respect to $P^{X=x}$). ◁

Remark 5.7 [Expectation of $E^{X=x}(Y|Z)$ With Respect to P] Aside from the properties of a conditional expectation (see sect. 4.2), $E^{X=x}(Y|Z)$ has a number of properties referring to the probability measure P on $(\Omega,\mathscr{A})$. Note that $E^{X=x}(Y|Z)$ is not only a random variable on the probability space $(\Omega,\mathscr{A},P^{X=x})$, but also a random variable on $(\Omega,\mathscr{A},P)$. This follows from the fact that $(\Omega,\mathscr{A},P)$ and $(\Omega,\mathscr{A},P^{X=x})$ share the same measurable space $(\Omega,\mathscr{A})$ (see Def. 2.2). Hence, we may not only consider the expectation $E^{X=x}\big(E^{X=x}(Y\,|\,Z)\big)$ with respect to the measure $P^{X=x}$, but also the expectation $E\big(E^{X=x}(Y|Z)\big)$ with respect to the measure P. However, this expectation is not always meaningful and even if it is, then it is not necessarily identical to $E^{X=x}(Y)$. The crucial condition under which the expectation $E\big(E^{X=x}(Y|Z)\big)$ is meaningful is related to various kinds of uniqueness of $E^{X=x}(Y|Z)$ (see sect. 5.2). ◁

Remark 5.8 [The Set $\mathscr{E}^{X=x}(Y|Z)$] As is true for all conditional expectations, there can be more than one single version of the Z-conditional expectation of Y with respect to the probability measure $P^{X=x}$. Therefore, we use $\mathscr{E}^{X=x}(Y|Z)$ to denote the *set of all versions of the Z-conditional expectation of Y with respect to* $P^{X=x}$. ◁

Example 5.9 [Joe and Ann With Randomized Assignment] The last column of Table 1.2 shows the values of the conditional expectation $E^{X=1}(Y|U)$ for all eight elements $\omega_i \in \Omega$. This random variable takes on only two different values, the value .8 if $\omega_i \in \{U=Joe\}$ and .4 if $\omega_i \in \{U=Ann\}$. In this example, $E^{X=1}(Y|U)$ is uniquely defined. This means that the set $\mathscr{E}^{X=1}(Y|U)$ has just one single element. Using Equation (3.13), the expectation of $E^{X=1}(Y|U)$ with respect to the measure P is

$$\begin{aligned} &E\big(E^{X=1}(Y|U)\big)\\ &= E^{X=1}(Y|U=Joe)\cdot P(U=Joe) + E^{X=1}(Y|U=Ann)\cdot P(U=Ann)\\ &= .8\cdot .5 + .4\cdot .5 \;=\; .6. \end{aligned}$$

Furthermore, the expectation of $E^{X=1}(Y|U)$ with respect to $P^{X=1}$ is

$$\begin{aligned} &E^{X=1}\big(E^{X=1}(Y|U)\big)\\ &= E^{X=1}(Y|U=Joe)\cdot P^{X=1}(U=Joe) + E^{X=1}(Y|U=Ann)\cdot P^{X=1}(U=Ann)\\ &= .8\cdot .5 + .4\cdot .5 \;=\; .6. \end{aligned}$$

Hence, the two expectations are identical, which is due to the fact that, in this example, X and U are independent. ◁

Example 5.10 [Joe and Ann With Self-Selection] In the example presented in Table 1.4, the conditional expectation $E^{X=1}(Y|U) = P^{X=1}(Y=1|U)$ is also uniquely defined. In this example, the expectation of $E^{X=1}(Y|U)$ with respect to P is again

$$\begin{aligned}
&E\big(E^{X=1}(Y|U)\big)\\
&= P^{X=1}(Y=1|U=Joe)\cdot P(U=Joe) \;+\; P^{X=1}(Y=1|U=Ann)\cdot P(U=Ann)\\
&= .8\cdot .5 \;+\; .4\cdot .5 \;=\; .6.
\end{aligned}$$

However, the expectation of $E^{X=1}(Y|U)$ with respect to $P^{X=1}$ is

$$\begin{aligned}
&E^{X=1}\big(E^{X=1}(Y|U)\big)\\
&= P^{X=1}(Y=1|U=Joe)\cdot P^{X=1}(U=Joe) \;+\; P^{X=1}(Y=1|U=Ann)\cdot P^{X=1}(U=Ann)\\
&= P^{X=1}(Y=1|U=Joe)\cdot \frac{P(U=Joe, X=1)}{P(X=1)} \;+\; P^{X=1}(Y=1|U=Ann)\cdot \frac{P(U=Ann, X=1)}{P(X=1)}\\
&= .8\cdot \frac{.02}{.4} \;+\; .4\cdot \frac{.38}{.4}\\
&= .8\cdot .05 \;+\; .4\cdot .95 \;=\; .42.
\end{aligned}$$

Hence, in this example, the two expectations $E\big(E^{X=1}(Y|U)\big)$ and $E^{X=1}\big(E^{X=1}(Y|U)\big)$ are not identical. In this example, X and U are *not* independent. ◁

Example 5.11 [No Treatment for Joe] The last but one column of Table 2.1 shows the values of a version of the conditional expectation $E^{X=1}(Y|U) = P^{X=1}(Y=1|U)$ for all eight elements $\omega_i \in \Omega$. This random variable takes on two different values, 99 if $\omega_i \in \{U=Joe\}$ and .4 if $\omega_i \in \{U=Ann\}$. In this example, the version $E^{X=1}(Y|U) \in \mathscr{E}^{X=1}(Y|U)$ is uniquely defined by the values in the last but one column of this table.

Using Equation (3.13), the expectation of $E^{X=1}(Y|U)$ with respect to P is

$$\begin{aligned}
&E\big(E^{X=1}(Y|U)\big)\\
&= E^{X=1}(Y|U=Joe)\cdot P(U=Joe) + E^{X=1}(Y|U=Ann)\cdot P(U=Ann)\\
&= 99\cdot .5 + .4\cdot .5 \;=\; 49.5 + .2 \;=\; 49.7.
\end{aligned}$$

However, there are infinitely many other versions $V_1 \in \mathscr{E}^{X=1}(Y|U)$, and each of these versions can have a different expectation with respect to the measure P. This is easily seen replacing, in the last displayed equation, the value 99 by any other number. Exchanging the number 99 by any other real number creates a new version $V_1 \in \mathscr{E}^{X=1}(Y|U)$ provided that the assignment of .4 to $\omega_5, \ldots, \omega_8$ remains unchanged. The last column of Table 2.1 is an example, in which, instead of 99, we assigned the number .8 to $\omega_1, \ldots, \omega_4$.

In contrast to $E\big(E^{X=1}(Y|U)\big)$, the expectation of $E^{X=1}(Y|U)$ with respect to $P^{X=1}$ is

$$\begin{aligned}
&E^{X=1}\big(E^{X=1}(Y|U)\big)\\
&= E^{X=1}(Y|U=Joe)\cdot P^{X=1}(U=Joe) + E^{X=1}(Y|U=Ann)\cdot P^{X=1}(U=Ann)\\
&= 99\cdot 0 + .4\cdot 1 \;=\; .4.
\end{aligned}$$

Hence, $E^{X=1}\big(E^{X=1}(Y|U)\big)$ remains unchanged if we replace the value 99 by any other number, and with it, the version $V_1 \in \mathscr{E}^{X=1}(Y|U)$ by another version $V_1^* \in \mathscr{E}^{X=1}(Y|U)$. ◁

5.2 Uniqueness, $P^{X=x}$-Uniqueness, and P-Uniqueness of $E^{X=x}(Y|Z)$

Remark 5.12 [$P^{X=x}$-Uniqueness of $E^{X=x}(Y|Z)$] Even if the set $\mathscr{E}^{X=x}(Y|Z)$ has more than one element, then

$$\forall\, V_x, V_x^* \in \mathscr{E}^{X=x}(Y|Z)\colon\; V_x \underset{P^{X=x}}{=} V_x^* \tag{5.10}$$

and, equivalently,

$$\forall\, V_x, V_x^* \in \mathscr{E}^{X=x}(Y|Z)\colon\; P^{X=x}\big(\{\omega \in \Omega\colon V_x(\omega) = V_x^*(\omega)\}\big) \;=\; 1 \tag{5.11}$$

(cf. Rem. 4.8). Equation (5.10) is what we mean saying that *$E^{X=x}(Y|Z)$ is $P^{X=x}$-unique*. ◁

Example 5.13 [No Treatment for Joe] Using the probabilities $P^{X=1}(\{\omega_i\})$ displayed in the fourth column of Table 2.1 shows that

$$P^{X=1}\big(\{\omega \in \Omega\colon V_1(\omega) = V_1^*(\omega)\}\big) = 1$$

holds for $V_1 = E^{X=1}(Y|U) = P^{X=1}(Y{=}1|U)$ and $V_1^* = E^{X=1}(Y|U)^* = P^{X=1}(Y{=}1|U)^*$ specified in Table 2.1. ◁

Remark 5.14 [$E^{X=x}(Y|Z)$ is Not Necessarily P-Unique] Although $E^{X=x}(Y|Z)$ is always $P^{X=x}$-unique, $E^{X=x}(Y|Z)$ is not necessarily P-unique, that is,

$$\forall\, V_x, V_x^* \in \mathscr{E}^{X=x}(Y|Z)\colon\; V_x \underset{P}{=} V_x^* \tag{5.12}$$

and, equivalently

$$\forall\, V_x, V_x^* \in \mathscr{E}^{X=x}(Y|Z)\colon\; P\big(\{\omega \in \Omega\colon V_x(\omega) = V_x^*(\omega)\}\big) \;=\; 1 \tag{5.13}$$

do *not necessarily hold*. In other words, although the definition of $E^{X=x}(Y|Z)$ implies that it is $P^{X=x}$-unique, it does not imply that is P-unique. ◁

Example 5.15 [No Treatment for Joe] Consider the versions $V_1, V_1^* \in \mathscr{E}^{X=1}(Y|U)$ considered in Example 5.13. Using the probabilities $P(\{\omega_i\})$ displayed in the second column of Table 2.1 shows: $P\big(\{\omega \in \Omega\colon V_1(\omega) = V_1^*(\omega)\}\big) = .5. \neq 1$. Hence, *$E^{X=1}(Y|U)$ is not P-unique*. In contrast, in the same example, the conditional expectation $E^{X=0}(Y|U) = P^{X=0}(Y{=}1|U)$ *is P-unique*. It is even uniquely defined, that is, if $V_0, V_0^* \in \mathscr{E}^{X=0}(Y|U)$, then $V_0 = V_0^*$. ◁

The concept of P-uniqueness is closely related to concept of absolute continuity of a measure with respect to another one. This term is defined as follows [see SN-Def. 3.70 (i)].

Definition 5.16 [Absolute Continuity of a Measure With Respect to Another One] *Let P and Q be measures on the measurable space $(\Omega, \mathscr{A})$. Then the measure Q is called absolutely continuous on $\mathscr{A}$ with respect to P, denoted $Q \underset{\mathscr{A}}{\ll} P$, if*

$$\forall\, A \in \mathscr{A}\colon\; P(A) = 0 \quad\Rightarrow\quad Q(A) = 0. \tag{5.14}$$

Hence, if Proposition (5.14) holds, then this means: If $A \in \mathscr{A}$ is a null set with respect to P, then it is also a null set with respect to the measure Q.

Lemma 5.17 [$P^{X=x}$ **is Absolutely Continuous With Respect to** P]
Let X be a random variable on the probability space $(\Omega, \mathscr{A}, P)$, let $x \in \Omega'_X$, $\{X=x\} \in \mathscr{A}$, and $P(X=x) > 0$. Furthermore, let $P^{X=x}$ be the probability measure on $(\Omega, \mathscr{A})$ defined by Equation (5.1). Then $P^{X=x}$ is absolutely continuous on $\mathscr{A}$ with respect to P, that is,

$$P^{X=x} \underset{\mathscr{A}}{\ll} P. \tag{5.15}$$

(Proof p. 121)

Hence, if $A \in \mathscr{A}$ is a null set with respect to P, then A is also a null set with respect to $P^{X=x}$.

Remark 5.18 [P **is Not Necessarily Absolute Continuous With Respect to** $P^{X=x}$] In contrast, under the assumptions of Lemma 5.17, $P \underset{\mathscr{A}}{\ll} P^{X=x}$ does *not* always hold (see Example 5.19). ◁

Example 5.19 [No Treatment for Joe] In the example displayed in Table 2.1, the event

$$\{X=1\} = \big\{(\mathit{Joe}, \mathit{yes}, -),\ (\mathit{Joe}, \mathit{yes}, +),\ (\mathit{Ann}, \mathit{yes}, -),\ (\mathit{Ann}, \mathit{yes}, +)\big\}$$

that the *drawn person is treated*, has the probability $P(X=1) = .12+.08 = .2$, and the $(X=1)$-conditional probability of the event

$$\{U=\mathit{Joe}\} = \big\{(\mathit{Joe}, \mathit{no}, -),\ (\mathit{Joe}, \mathit{no}, +),\ (\mathit{Joe}, \mathit{yes}, -),\ (\mathit{Joe}, \mathit{yes}, +)\big\}$$

that Joe is drawn is $P^{X=1}(U=\mathit{Joe}) = 0$, whereas $P(U=\mathit{Joe}) = .5$. Hence, P *is not absolutely continuous* on $\mathscr{A}$ with respect to $P^{X=1}$.

In contrast,

$$\begin{aligned} P(\emptyset) &= P\big(\{(\mathit{Joe}, \mathit{yes}, -)\}\big) = P\big(\{(\mathit{Joe}, \mathit{yes}, +)\}\big) \\ &= P\big(\{(\mathit{Joe}, \mathit{yes}, -), (\mathit{Joe}, \mathit{yes}, +)\}\big) = 0 \end{aligned}$$

and

$$\begin{aligned} P^{X=1}(\emptyset) &= P^{X=1}\big(\{(\mathit{Joe}, \mathit{yes}, -)\}\big) = P^{X=1}\big(\{(\mathit{Joe}, \mathit{yes}, +)\}\big) \\ &= P^{X=1}\big(\{(\mathit{Joe}, \mathit{yes}, -), (\mathit{Joe}, \mathit{yes}, +)\}\big) = 0 \end{aligned}$$

(see columns two and four of Table 2.1). Hence, $P^{X=1}$ *is absolutely continuous* on $\mathscr{A}$ with respect to P. ◁

In the following lemma we consider the relationship between $P^{X=x}$-equivalence and P-equivalence of two random variables, presuming $P \underset{\mathscr{A}}{\ll} P^{X=x}$, that is, presuming absolute continuity on $\mathscr{A}$ of P with respect to $P^{X=x}$. Reading this lemma, remember that the set $\{X \neq Y\} = \{\omega \in \Omega \colon X(\omega) \neq Y(\omega)\}$ denotes the event that X and Y do not take on the same value.

Lemma 5.20 [An Implication of Absolute Continuity of P With Respect to $P^{X=x}$] *Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$, let $x \in \Omega'_X$, $\{X{=}x\} \in \mathscr{A}$, $P(X{=}x) > 0$, and assume that $P^{X=x}$ is the probability measure on $(\Omega, \mathscr{A})$ defined by Equation (5.1). Furthermore, assume $\{X \neq Y\} \in \mathscr{A}$. Then*

$$\big(X \underset{P^{X=x}}{=} Y \;\wedge\; P \underset{\mathscr{A}}{\ll} P^{X=x}\big) \quad \Rightarrow \quad X \underset{P}{=} Y. \tag{5.16}$$

(Proof p. 121)

Example 5.21 [No Treatment for Joe] Consider the two random variables $P^{X=1}(Y{=}1|U)$ and $P^{X=1}(Y{=}1|U)^*$ displayed in the last two colums of Table 2.1. Inspection of this table shows that these two random variables *are not identical*, because Equation (2.27) does not hold. For example,

$$P^{X=1}(Y{=}1|U)(\omega_1) \;\neq\; P^{X=1}(Y{=}1|U)^*(\omega_1).$$

Furthermore, they are also *not P-equivalent*, because Proposition (2.28) does not hold: There is no $A \in \mathscr{A}$ such that $P(A) = 0$ and

$$\forall\, \omega \in \Omega \setminus A\colon\; P^{X=1}(Y{=}1|U)(\omega) = P^{X=1}(Y{=}1|U)^*(\omega),$$

because the values of the two random variables differ for all elements of the set $\{\omega_1,\ldots,\omega_4\}$, and $P(\{\omega_1,\ldots,\omega_4\}) = .5$. In contrast, there is an $A \in \mathscr{A}$, namely $A = \{\omega_1,\ldots,\omega_4\}$, such that $P^{X=1}(A) = 0$ and

$$\forall\, \omega \in \Omega \setminus A\colon\; P^{X=1}(Y{=}1|U)(\omega) = P^{X=1}(Y{=}1|U)^*(\omega).$$

In other words

$$P^{X=1}(Y{=}1|U) \underset{P^{X=1}}{=} P^{X=1}(Y{=}1|U)^*$$

holds, but neither

$$P^{X=1}(Y{=}1|U) \underset{P}{=} P^{X=1}(Y{=}1|U)^*$$

nor

$$P^{X=1}(Y{=}1|U) \;=\; P^{X=1}(Y{=}1|U)^*.$$

Hence, this example shows: *$P^{X=x}$-equivalence of two random variables does not imply that they are P-equivalent.* According to Lemma 5.20, we need the additional requirement $P \underset{\mathscr{A}}{\ll} P^{X=x}$. ◁

Remark 5.22 [A New Notation] Let Z be a real-valued random variable on a probability space $(\Omega, \mathscr{A}, Q)$ and $B \in \mathscr{B}$. Then we define

$$Z \underset{Q}{\in} B \quad :\Leftrightarrow \quad \big(\exists\, A \in \mathscr{A}\colon\, Q(A) = 0 \;\wedge\; \forall\, \omega \in \Omega \setminus A\colon\, Z(\omega) \in B\big). \tag{5.17}$$

◁

Example 5.23 [Some Special Cases] Special cases in which we can apply the notation introduced in Proposition (5.17) are:

(i) If $B = (0,\infty)$, then $Z \underset{Q}{\in} B \;\Leftrightarrow\; Z \underset{Q}{>} 0$.

(ii) If $\alpha \in \mathbb{R}$ and $B = \{\alpha\}$, then $Z \underset{Q}{\in} B \;\Leftrightarrow\; Z \underset{Q}{=} \alpha$.

(iii) If $B = (-\infty, 0]$, then $Z \underset{Q}{\in} B \Leftrightarrow Z \underset{Q}{\le} 0$.

◁

Lemma 5.24 [Almost Sure Properties With Respect to $P^{X=x}$ and P]
Let X and Z_x, $x \in X(\Omega)$, be random variables on a probability space $(\Omega, \mathscr{A}, P)$ and let $(\mathbb{R}, \mathscr{B})$ the value space of all Z_x, $x \in X(\Omega)$. Furthermore, let $B \in \mathscr{B}$, and for all $x \in X(\Omega)$, let $\{X=x\} \in \mathscr{A}$, $P(X=x) > 0$, and assume that $P^{X=x}$ is the probability measure on $(\Omega, \mathscr{A})$ defined by Equation (5.1). Then $X(\Omega)$ is finite or countable and

$$\big(\forall\, x \in X(\Omega)\colon\ Z_x \underset{P^{X=x}}{\in} B\big) \quad \Leftrightarrow \quad \left(\sum_{x \in X(\Omega)} Z_x \cdot 1_{X=x}\right) \underset{P}{\in} B. \tag{5.18}$$

(Proof p. 122)

Remark 5.25 [Absolute Continuity] In the Theorem 5.27, we present conditions that are equivalent to P-uniqueness of a conditional expectation $E^{X=x}(Y|Z)$ [see Eq. (5.12)]. In this theorem, we use the notation

$$P(X=x\,|\,Z) \underset{P}{>} 0 \quad :\Leftrightarrow \quad P\big(\{\omega \in \Omega\colon P(X=x\,|\,Z)(\omega) > 0\}\big) = 1, \tag{5.19}$$

$$P \underset{\sigma(Z)}{\ll} P^{X=x} \quad :\Leftrightarrow \quad \forall C \in \sigma(Z)\colon \big(P^{X=x}(C) = 0 \ \Rightarrow\ P(C) = 0\big), \tag{5.20}$$

and

$$P_Z \underset{\mathscr{A}'_Z}{\ll} P_Z^{X=x} \quad :\Leftrightarrow \quad \forall\, A' \in \mathscr{A}'_Z\colon \big(P_Z^{X=x}(A') = 0 \ \Rightarrow\ P_Z(A') = 0\big). \tag{5.21}$$

The first condition, $P(X=x\,|\,Z) \underset{P}{>} 0$, means that the Z-conditional probability of the event $\{X=x\}$ is greater than zero, P-almost surely. Furthermore, $P \underset{\sigma(Z)}{\ll} P^{X=x}$ means that the measure P is absolutely continuous on the σ-algebra $\sigma(Z)$ with respect to the measure $P^{X=x}$ (see Rem. 5.18). And finally, $P_Z \underset{\mathscr{A}'_Z}{\ll} P_Z^{X=x}$ means that the distribution P_Z is absolutely continuous on the σ-algebra $\mathscr{A}'_Z$ with respect to the measure $P_Z^{X=x}$. ◁

Remark 5.26 [Distribution of Z with respect to $P^{X=x}$] Assume that Z is a random variable on the probability space $(\Omega, \mathscr{A}, P^{X=x})$, where $P^{X=x}$ is defined by Equation (5.1), and let $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value space of Z. Then the function $P_Z^{X=x}\colon \mathscr{A}'_Z \to [0,1]$ defined by

$$P_Z^{X=x}(A') \;=\; P^{X=x}(\{Z \in A'\}), \quad \forall\, A' \in \mathscr{A}'_Z, \tag{5.22}$$

is called the *distribution of Z with respect to $P^{X=x}$* (cf. Def. 2.27). ◁

Reading the following theorem, which is an adaptation of SN-Corollary 14.48, also remember that P_Z denotes the distribution of Z.

Theorem 5.27 [**P-Uniqueness of $E^{X=x}(Y|Z)$**]
Let the assumptions of Definition 5.4 hold, let $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value space of Z, and assume that $P^{X=x}$ is the probability measure on $(\Omega, \mathscr{A})$ defined by Equation (5.1). Then the following propositions are equivalent to each other.

(a) $E^{X=x}(Y|Z)$ *is* P*-unique*

(b) $P(X=x|Z) \underset{P}{>} 0$

(c) $P \underset{\sigma(Z)}{\ll} P^{X=x}$

(d) $P_Z \underset{\mathscr{A}'_Z}{\ll} P_Z^{X=x}$.

Furthermore, if there is a version $V_x \in \mathscr{E}^{X=x}(Y|Z)$ *such that* $E(V_x)$ *is finite, then each of (a) to (d) is also equivalent to*

(e) $\forall\, V_x, V_x^* \in \mathscr{E}^{X=x}(Y|Z):\ E(V_x) = E(V_x^*)$.

According to Proposition (e) of Theorem 5.27, if there is a version $V_x \in \mathscr{E}^{X=x}(Y|Z)$ such that $E(V_x)$ is finite, then the expectation $E\big(E^{X=x}(Y|Z)\big)$ is a uniquely defined number if and only if $E^{X=x}(Y|Z)$ is P-unique. The following example shows why this property is important.

Example 5.28 [No Treatment for Joe] The last columns of Table 2.1 display versions of the conditional expectations $E^{X=0}(Y|U) = P^{X=0}(Y{=}1|U)$ and $E^{X=1}(Y|U) = P^{X=1}(Y{=}1|U)$, respectively. In this example, $P^{X=0}(Y{=}1|U)$ is $P^{X=0}$-unique and P-unique. It is even uniquely defined. In contrast, the conditional expectation $E^{X=1}(Y|U) = P^{X=1}(Y{=}1|U)$ is neither unique nor P-unique. However, it is $P^{X=1}$-unique (see Example 5.21). In this example, condition (b) of Theorem 5.27 does not hold for $P(X{=}1|U)$, because $P(X{=}1|U)(\omega_1) = P(X{=}1|U)(\omega_2) = 0$ but $P(\{\omega_1, \omega_2\}) = .5 > 0$. Because $E^{X=1}(Y|U) = P^{X=1}(Y{=}1|U)$ is not P-unique, the expectations $E\big(P^{X=1}(Y{=}1|U)\big) = 49.7$ and $E\big(P^{X=1}(Y{=}1|U)^*\big) = .6$ are not identical and have no substantive meaning (see Exercise 5-2). ◁

5.3 Factorization of $E^{X=x}(Y|Z)$

Remark 5.29 [Factorization of $E^{X=x}(Y|Z)$] Let the assumptions of Definition 5.4 hold, let $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value space of Z, and let $E^{X=x}(Y|Z) \in \mathscr{E}^{X=x}(Y|Z)$. Then there is a measurable function $g_x\colon (\Omega'_Z, \mathscr{A}'_Z) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ such that

$$E^{X=x}(Y|Z) \;=\; g_x(Z) \tag{5.23}$$

(see Cor. 4.16). The function g_x is a *factorization of* $E^{X=x}(Y|Z)$ (see Def. 4.17). Note that a factorization g_x is a random variable on $(\Omega'_Z, \mathscr{A}'_Z, P_Z)$, whereas $E^{X=x}(Y|Z)$ and the composition $g_x(Z)$ are random variables on $(\Omega, \mathscr{A}, P^{X=x})$. ◁

A factorization of $E^{X=x}(Y|Z)$ can also be used for the definition of a $(Z{=}z)$-conditional expectation value of Y with respect to the probability measure $P^{X=x}$. In this definition we assume $P(X{=}x) > 0$, but not $P^{X=x}(Z{=}z) > 0$.

Definition 5.30 [$(Z{=}z)$-Conditional Expectation Value With Respect to $P^{X=x}$]
Let the assumptions of Definition 5.4 hold, let $(\Omega'_Z, \mathscr{A}'_Z)$ *denote the value space of* Z*, and let* $g_x\colon (\Omega'_Z, \mathscr{A}'_Z) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ *be a measurable function satisfying Equation (5.23). Then, for all* $z \in \Omega'_Z$*, we define*

$$E^{X=x}(Y|Z=z) := g_x(z), \tag{5.24}$$

and call it a $(Z=z)$*-conditional expectation value of* Y *with respect to the probability measure* $P^{X=x}$.

Remark 5.31 [Values of $E^{X=x}(Y|Z)$] The conditional expectation values $E^{X=x}(Y|Z=z)$ are the values of $E^{X=x}(Y|Z)$. In more formal terms,

$$\forall\, \omega \in \Omega: \quad E^{X=x}(Y|Z)(\omega) = E^{X=x}(Y|Z=z), \quad \text{if } \omega \in \{X=x, Z=z\}. \tag{5.25}$$

Furthermore, if $P(X=x, Z=z) > 0$, then $P(X=x) > 0$, $P(Z=z) > 0$, and

$$E^{X=x}(Y|Z=z) = E^{Z=z}(Y|X=x) = E^{X=x,\,Z=z}(Y) = E(Y|X=x, Z=z) \tag{5.26}$$

(see SN-Rem. 14.37). ◁

Example 5.32 [No Treatment for Joe] Now we compute the values of the conditional expectation $E^{X=0}(Y|U)$ in the example presented in Table 2.1. According to Remark 5.31, the values of the conditional expectation $E^{X=0}(Y|U)$ are the two conditional expectation values $E(Y|X=0, U=Joe)$ and $E(Y|X=0, U=Ann)$. Because, in this example, Y is binary,

$$E(Y|X=0, U=u) = P(Y=1|X=0, U=u),$$

the conditional expectation values $E(Y|X=0, U=Joe)$ and $E(Y|X=0, U=Ann)$ can be computed from the probabilities of the elementary events presented in Table 2.1 as follows:

$$P(Y=1|X=0, U=Joe) = \frac{P(Y=1, X=0, U=Joe)}{P(X=0, U=Joe)} = \frac{.35}{.15+.35} = .7$$

and

$$P(Y=1|X=0, U=Ann) = \frac{P(Y=1, X=0, U=Ann)}{P(X=0, U=Ann)} = \frac{.06}{.06+.24} = .2.$$

If we try to apply the equation that is analog to the equation for $P(Y=1|X=0, U=Ann)$ to the conditional probability $P(Y=1|X=1, U=Joe)$, then we find $P(X=1, U=Joe) = 0$. This implies that the fraction $P(Y=1, X=1, U=Joe)/P(X=1, U=Joe)$ is undefined. However, using the (version of the) conditional expectation $E^{X=1}(Y|U) = P^{X=1}(Y=1|U)$ specified in the last but one column of Table 2.1 and Equation (5.24) yields

$$E^{X=1}(Y|U=Joe) = E(Y|X=1, U=Joe) = P(Y=1|X=1, U=Joe) = 99.$$

Obviously, from a substantive point of view, this number, is meaningless, because we can replace it by any other real number. However, this does not mean that Definition 5.30 is a bad idea, because it still allows us to formulate meaningful propositions about $P^{X=x}$-almost all $(Z=z)$-conditional expectation values $E^{X=x}(Y|Z=z)$. The arguments presented in Example 4.30 still apply if we exchange the measure P and the conditional expectation $E(Y|X)$ by the probability measure $P^{X=x}$ and the conditional expectation $E^{X=x}(Y|Z)$, respectively. ◁

5.4 Partial Conditional Expectation $E(Y|X{=}x, Z)$

Now we turn to a second concept that can be used to describe how the conditional expectation of Y depends on a random variable Z if another random variable X is kept constant at one of its values x. For this new concept we do not have to assume $P(X{=}x) > 0$.

According to Remark 4.12, the term $E(Y|X,Z)$ denotes the (X,Z)-conditional expectation of Y with respect to P. Let $(\Omega'_X, \mathscr{A}'_X)$, $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value spaces of X and Z, respectively. Then the concept of a *partial* $(X{=}x, Z)$*-conditional expectation* of Y builds on Corollary 4.16, according to which, for each version $E(Y|X,Z) \in \mathscr{E}(Y|X,Z)$, there is a measurable function $g: (\Omega'_X \times \Omega'_Z, \mathscr{A}'_X \otimes \mathscr{A}'_Z) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ such that

$$E(Y|X,Z) \;=\; g(X,Z), \tag{5.27}$$

where $(\Omega'_X \times \Omega'_Z, \mathscr{A}'_X \otimes \mathscr{A}'_Z)$ denotes the value space of (X,Z) (see sect. 2.1.5). In this equation, $g(X,Z)$ denotes the composition of the multivariate random variable (X,Z) and g. The function g is a *factorization of* $E(Y|X,Z)$ (see Cor. 4.16). According to Equation (4.16), for $(x,z) \in \Omega'_X \times \Omega'_Z$,

$$E(Y|X{=}x, Z{=}z) \;=\; g(x,z), \tag{5.28}$$

is an $(X{=}x, Z{=}z)$-conditional expectation value of Y.

In Definition 5.34, we will refer to the function $g_x: \Omega'_Z \to \overline{\mathbb{R}}$ that, for $x \in \Omega'_X$, is defined by

$$g_x(z) \;=\; g(x,z), \quad \forall\, z \in \Omega'_Z. \tag{5.29}$$

Hence, a value $g_x(z)$ is identical to $E(Y|X{=}x, Z{=}z)$, that is,

$$g_x(z) \;=\; g(x,z) \;=\; E(Y|X{=}x, Z{=}z), \quad \forall\, z \in \Omega'_Z. \tag{5.30}$$

Remark 5.33 [Uniqueness of $E(Y|X{=}x, Z{=}z)$] In Equations (5.29) and (5.30) we do not assume $P(X{=}x, Z{=}z) > 0$. However, there can be several versions $E(Y|X,Z) \in \mathscr{E}(Y|X,Z)$, and even for a given version $E(Y|X,Z) \in \mathscr{E}(Y|X,Z)$, there can be several factorizations satisfying Equation (5.27) (see Rem. 4.31). This implies that $E(Y|X{=}x, Z{=}z)$ is not uniquely defined if $P(X{=}x, Z{=}z) = 0$. (For more details see SN-section 10.4.4). ◁

Definition 5.34 [Partial Conditional Expectation]
Let X, Y, and Z be random variables on $(\Omega, \mathscr{A}, P)$ and let $(\Omega'_X, \mathscr{A}'_X)$, $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value spaces of X and Z, respectively. Furthermore, assume that Y is numerical and nonnegative or with finite expectation $E(Y)$, let $E(Y|X,Z) \in \mathscr{E}(Y|X,Z)$ and let g be a factorization of $E(Y|X,Z)$ such that $g(X,Z) = E(Y|X,Z)$. Finally, for $x \in \Omega'_X$, let g_x be defined by Equation (5.29). Then we call the function $E(Y|X{=}x, Z): \Omega \to \overline{\mathbb{R}}$ defined by

$$E(Y|X{=}x, Z) \;:=\; g_x(Z) \tag{5.31}$$

a version of the partial $(X{=}x, Z)$-conditional expectation of Y (with respect to P).

Remark 5.35 [A Partial Conditional Expectation is a Random Variable] For all $x \in \Omega'_X$, the function $E(Y|X{=}x, Z) = g_x(Z)$ denotes the composition of Z and g_x. Hence, for $x \in \Omega'_X$, $E(Y|X{=}x, Z)$ is a Z-measurable random variable on $(\Omega, \mathscr{A}, P)$ (see Rem. 2.13 and Lemma 2.24). ◁

Remark 5.36 [Partial Conditional Probability] If $A \in \mathscr{A}$, then we also use the notation $P(A \mid X{=}x, Z) := E(1_A \mid X{=}x, Z)$ and call it a *partial* $(X{=}x, Z)$*-conditional probability* of (the event) A (with respect to P). Furthermore, if Y is binary with values 0 and 1, then we use the notation $P(Y{=}1 \mid X{=}x, Z) := E(Y \mid X{=}x, Z)$ and call it a *partial* $(X{=}x, Z)$*-conditional probability* of (the event) $\{Y{=}1\}$ (with respect to P). ◁

Remark 5.37 [The Set $\mathscr{E}(Y \mid X{=}x, Z)$] Note that $E(Y \mid X{=}x, Z)$ is not uniquely defined for two reasons. The first is that $E(Y \mid X, Z)$ is not uniquely defined (see Rem. 4.5). The second reason is that even for a given version $E(Y \mid X, Z)$, the factorization g of $E(Y \mid X, Z) = g(X, Z)$ is not uniquely defined (see Rem. 4.31 and Cor. 4.32). Therefore, we use

$$\mathscr{E}(Y \mid X{=}x, Z) := \big\{ g_x(Z) \colon g_x \text{ satisfies (5.30), where } g(X, Z) \in \mathscr{E}(Y \mid X, Z) \big\}$$

to denote the *set of all versions of the partial* $(X{=}x, Z)$*-conditional expectation of* Y. ◁

Remark 5.38 [Discrete X] Under the assumptions of Definition 5.34, suppose that X is a random variable on $(\Omega, \mathscr{A}, P)$ and that the image $X(\Omega)$ of X is finite or countable. Then

$$E(Y \mid X, Z) = \sum_{x \in X(\Omega)} E(Y \mid X{=}x, Z) \cdot 1_{X=x} \tag{5.32}$$

holds *for the specific version* $E(Y \mid X, Z) \in \mathscr{E}(Y \mid X, Z)$ that is used in Definition 5.34 (for a proof see SN-Exercise 14-6). Furthermore, under these assumptions,

$$V \underset{P}{=} \sum_{x \in X(\Omega)} E(Y \mid X{=}x, Z) \cdot 1_{X=x}, \quad \text{if } V \in \mathscr{E}(Y \mid X, Z). \tag{5.33}$$

That is, under these assumptions, each version of the conditional expectation $E(Y \mid X, Z)$ is P-equivalent to the sum on the right-hand side of Equation (5.32). ◁

Remark 5.39 [A Partial Conditional Expectation is not a Conditional Expectation] Note again, a partial conditional expectation is defined even if $P(X{=}x) = 0$. However, this definition is not unique (see Rem. 5.37). Also note that, in general, a partial conditional expectation is not a conditional expectation (with all its well-known properties) unless $P(X{=}x) > 0$. In the latter case a partial conditional expectation is in fact a version of a conditional expectation. This is detailed in the following theorem. ◁

Theorem 5.40 [Relationship Between $E(Y \mid X{=}x, Z)$ and $E^{X=x}(Y \mid Z)$]
Let the assumptions of Definitions 5.34 hold and assume that $x \in \Omega'_X$, $\{X{=}x\} \in \mathscr{A}$, and $P(X{=}x) > 0$. Furthermore, let $P^{X=x}$ be defined by Equation (5.1). Then

$$E(Y \mid X{=}x, Z) \in \mathscr{E}^{X=x}(Y \mid Z). \tag{5.34}$$

This implies

$$E(Y \mid X{=}x, Z) \underset{P^{X=x}}{=} E^{X=x}(Y \mid Z), \quad \forall\, E^{X=x}(Y \mid Z) \in \mathscr{E}^{X=x}(Y \mid Z). \tag{5.35}$$

For a proof see SN-Theorem 14.33. According to this theorem, the partial conditional expectation $E(Y \mid X{=}x, Z)$ is also a version of the Z-conditional expectation of Y with respect to $P^{X=x}$, provided that $P(X{=}x) > 0$ and $P^{X=x}$ is the measure defined by Equation (5.1).

Box 5.1 Properties of P-uniqueness of $E^{X=x}(Y|Z)$

Let the assumptions of Definition 5.4 hold, let W be a random variable on $(\Omega, \mathscr{A}, P)$, and assume that $P^{X=x}$ is the measure defined by Equation (5.1). Then:

$$E^{X=x}(Y|Z) \text{ is } P\text{-unique} \;:\Leftrightarrow\; \forall\, V, V^* \in \mathscr{E}^{X=x}(Y|Z):\; V \underset{P}{=} V^* \tag{i}$$

$$E^{X=x}(Y|Z) \text{ is } P\text{-unique} \;\Leftrightarrow\; P \underset{\sigma(Z)}{\ll} P^{X=x} \tag{ii}$$

$$E^{X=x}(Y|Z) \text{ is } P\text{-unique} \;\Leftrightarrow\; P(X{=}x\,|Z) \underset{P}{>} 0 \tag{iii}$$

$$E^{X=x}(Y|Z) \text{ is } P\text{-unique} \;\Rightarrow\; E^{X=x}(Y|W) \text{ is } P\text{-unique}, \quad \text{if } \sigma(W) \subset \sigma(Z). \tag{iv}$$

Additionally, let V, V^* be random variables on $(\Omega, \mathscr{A}, P)$, let $B \in \mathscr{A}$, $P(B) > 0$, and let P^B be the measure introduced in Equation (1.21). Furthermore, let P_V^B, $E^B(V)$ denote the distribution and the expectation of V with respect to P^B, respectively. Then

$$E^{X=x}(Y|Z) \text{ is } P\text{-unique} \;\Rightarrow\; E^{X=x}(Y|Z) \text{ is } P^B\text{-unique} \tag{v}$$

$$E^{X=x}(Y|Z) \text{ is } P\text{-unique} \;\Rightarrow\; \forall\, V, V^* \in \mathscr{E}^{X=x}(Y|Z):\; P_V^B = P_{V^*}^B \tag{vi}$$

$$E^{X=x}(Y|Z) \text{ is } P\text{-unique} \;\Rightarrow\; \forall\, V, V^* \in \mathscr{E}^{X=x}(Y|Z):\; E^B(V) = E^B(V^*). \tag{vii}$$

Let $P(W{=}w) > 0$, $E^{X=x}(Y|Z)$ or $E^{W=w}(Y|Z)$ be real-valued, and $\alpha, \beta \in \mathbb{R}$. Then

$$E^{X=x}(Y|Z),\, E^{W=w}(Y|Z) \text{ are } P\text{-unique} \;\Rightarrow\; \alpha E^{X=x}(Y|Z) + \beta E^{W=w}(Y|Z) \text{ is } P\text{-unique}. \tag{viii}$$

If $Z(\Omega)$ is finite or countable, then:

$$\big(\forall\, z \in Z(\Omega):\; P^{X=x}(Z{=}z) > 0\big) \;\Rightarrow\; E^{X=x}(Y|Z) \text{ is } P\text{-unique}. \tag{ix}$$

5.5 Further Properties of $E^{X=x}(Y|Z)$

Remark 5.41 [Relationship Between $E^{X=x}(Y|Z)$ and $E(Y|X,Z)$] Let the assumptions of Definition 5.4 hold, assume that the image $X(\Omega)$ of X is finite or countable with $\{X{=}x\} \in \mathscr{A}$ and $P(X{=}x) > 0$ for all $x \in X(\Omega)$, and that $P^{X=x}$ is defined by Equation (5.1). Then

$$E(Y|X,Z) \underset{P}{=} \sum_{x \in X(\Omega)} E^{X=x}(Y|Z) \cdot 1_{X=x} \tag{5.36}$$

(see SN-Rem. 14.34). Because the right-hand side of Equation (5.36) is (X, Z)-measurable, this equation implies that it is a version of the conditional expectation $E(Y|X,Z)$ [see Prop. (4.8)]. For an indicator variable $1_{X=x}$, Equation (5.36) simplifies to

$$E(Y|1_{X=x}, Z) \underset{P}{=} E^{X=x}(Y|Z) \cdot 1_{X=x} + E^{X\neq x}(Y|Z) \cdot 1_{X\neq x}, \tag{5.37}$$

because $\{X{=}x\} = \{1_{X=x}{=}1\} = \{\omega \in \Omega: 1_{X=x}(\omega) = 1\}$.

Furthermore, under the same assumptions,

$$\big(\forall\, x \in X(\Omega):\; E^{X=x}(Y|Z) \underset{P^{X=x}}{>} 0\big) \;\Leftrightarrow\; \Bigg(\sum_{x \in X(\Omega)} E^{X=x}(Y|Z) \cdot 1_{X=x}\Bigg) \underset{P}{>} 0 \tag{5.38}$$

$$\Leftrightarrow \quad E(Y|X,Z) \underset{P}{>} 0, \tag{5.39}$$

[for the last equivalence proposition see Eq. (5.36)], where

$$E^{X=x}(Y|Z) \underset{P^{X=x}}{>} 0 \quad :\Leftrightarrow \quad \big(\exists\, A \in \mathscr{A}\colon P^{X=x}(A) = 0 \;\wedge\; \forall\, \omega \in \Omega \setminus A\colon E^{X=x}(Y|Z)(\omega) > 0\big).$$

Proposition (5.38) follows from Lemma 5.24 if $E^{X=x}(Y|Z)$ takes the role of Z_x, we use

$$E^{X=x}(Y|Z) \underset{P^{X=x}}{>} 0 \quad \Leftrightarrow \quad E^{X=x}(Y|Z) \underset{P^{X=x}}{\in} \mathbb{R}^+, \tag{5.40}$$

and

$$\Big(\sum_{x \in X(\Omega)} E^{X=x}(Y|Z) \cdot 1_{X=x}\Big) \underset{P}{>} 0 \quad \Leftrightarrow \quad \Big(\sum_{x \in X(\Omega)} E^{X=x}(Y|Z) \cdot 1_{X=x}\Big) \underset{P}{\in} \mathbb{R}^+, \tag{5.41}$$

where $\mathbb{R}^+ = (0, \infty)$ denotes the set of all positive real numbers [see Example 5.23 (i)]. ◁

Remark 5.42 [Properties Related to P-Uniqueness of $E^{X=x}(Y|Z)$] Box 5.1 summarizes some properties related to P-uniqueness of a Z-conditional expectation $E^{X=x}(Y|Z)$ of Y with respect to the measure $P^{X=x}$. This box is adapted from SN-Box 14.1. Proofs are found in the solution to SN-Exercise 14-13. ◁

5.6 (Conditional) Mean-Independence With Respect to $P^{X=x}$

Now we turn to mean-independence of a numerical random variable Y from a random variable Z with respect to the measure $P^{X=x}$. This concept specifies the intuitive idea of keeping constant a value x of a random variable X and then postulating that the Z-conditional expectation of Y actually does not depend on Z.

Definition 5.43 [Mean-Independence of Y From Z With Respect to $P^{X=x}$]
Let the assumptions of Definition 5.4 hold. Then we define mean-independence of Y from Z with respect to $P^{X=x}$, denoted $Y \underset{P^{X=x}}{\vDash} Z$, by

$$E^{X=x}(Y|Z) \underset{P^{X=x}}{=} E^{X=x}(Y). \tag{5.42}$$

Hence,

$$Y \underset{P^{X=x}}{\vDash} Z \quad :\Leftrightarrow \quad E^{X=x}(Y|Z) \underset{P^{X=x}}{=} E^{X=x}(Y). \tag{5.43}$$

In the following corollary we translate Theorem 4.35 to mean-independence with respect to $P^{X=x}$. In this corollary, we refer to $P_Z^{X=x}$, the *distribution of Z with respect to $P^{X=x}$* [see Eq. (5.22)].

Theorem 5.44 [Mean-Independence With Respect to $P^{X=x}$]
Let X, Y, and Z be random variables on $(\Omega, \mathscr{A}, P^{X=x})$, let $P^{X=x}$ be defined by Equation (5.1), and let $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value space of Z. Finally, assume that Y is numerical and nonnegative or with finite expectation $E^{X=x}(Y)$.

(i) Then

$$Y \underset{P^{X=x}}{\vDash} Z \quad \Leftrightarrow \quad E^{X=x}(Y|Z{=}z) = E^{X=x}(Y), \quad \text{for } P_Z^{X=x}\text{-a.a. } z \in \Omega'_Z. \tag{5.44}$$

(ii) If $z \in Z(\Omega)$, $\{Z{=}z\} \in \mathscr{A}$, and $P^{X=x}(Z{=}z) > 0$, then

$$Y \underset{P^{X=x}}{\vDash} Z \quad \Rightarrow \quad E^{X=x}(Y|Z{=}z) = E^{X=x}(Y). \tag{5.45}$$

According to Theorem 5.44 (i), mean-independence of Y from Z with respect to $P^{X=x}$ is equivalent to stating that the $(Z{=}z)$-conditional expectation values of Y with respect to $P^{X=x}$ actually do not depend on the values z of Z, for $P_Z^{X=x}$-a.a. $z \in \Omega'_Z$. If $P^{X=x}(Z{=}z) > 0$, then mean-independence of Y from Z with respect to $P^{X=x}$ implies that the two conditional expectation values $E^{X=x}(Y|Z{=}z)$ and $E^{X=x}(Y)$ are identical [see Th. 5.44 (i)].

Theorem 5.45 [Mean-Independence With Respect to $P^{X=x}$: Special Cases]
Let X, Y, and Z be random variables on $(\Omega, \mathscr{A}, P^{X=x})$ and let $P^{X=x}$ be defined by Equation (5.1). Finally, assume that Y is numerical and nonnegative or with finite expectation $E^{X=x}(Y)$.

(i) If, for all $z \in Z(\Omega)$: $\{Z{=}z\} \in \mathscr{A}$ and $P^{X=x}(Z{=}z) > 0$, then

$$Y \underset{P^{X=x}}{\vDash} Z \quad \Leftrightarrow \quad E^{X=x}(Y|Z{=}z) = E^{X=x}(Y), \quad \forall\, z \in Z(\Omega). \tag{5.46}$$

(ii) If $Z(\Omega) = \{0, 1\}$, $z \in Z(\Omega)$, $\{Z{=}z\} \in \mathscr{A}$, and $P^{X=x}(Z{=}z) > 0$, then

$$Y \underset{P^{X=x}}{\vDash} Z \quad \Leftrightarrow \quad E^{X=x}(Y|Z{=}z) = E^{X=x}(Y). \tag{5.47}$$

Hence, if $P^{X=x}(Z{=}z) > 0$ for all values z of Z, then mean-independence of Y from Z with respect to $P^{X=x}$ is equivalent to stating that the $(Z{=}z)$-conditional expectation values of Y with respect to $P^{X=x}$ actually do not depend on the values z of Z [see Th. 5.45 (i)]. And, if Z is binary and $P^{X=x}(Z{=}z) > 0$, then mean-independence of Y from Z with respect to $P^{X=x}$ is equivalent to $E^{X=x}(Y|Z{=}z) = E^{X=x}(Y)$ [see Th. 5.45 (ii)]. Note that, under the assumptions of Theorem 5.45 and $P^{X=x}(Z{=}z) > 0$,

$$Z(\Omega) = \{0, 1\} \wedge E^{X=x}(Y|Z{=}z) = E^{X=x}(Y) \quad \Rightarrow \quad E^{X=x}(Y|Z{\neq}z) = E^{X=x}(Y). \tag{5.48}$$

Remark 5.46 [Rewriting Some Propositions] Let the assumptions of Theorem 5.44 hold. If we additionally assume $\{X{=}x\} \in \mathscr{A}$ and that P is a probability measure on $(\Omega, \mathscr{A})$ such that $P(X{=}x) > 0$, then, according to Equation (3.24),

$$E^{X=x}(Y) \;=\; E(Y|X=x). \tag{5.49}$$

Similarly, if additionallly, $z \in Z(\Omega)$, $\{Z=z\} \in \mathscr{A}$, and $P^{X=x}(Z=z) > 0$, then

$$E^{X=x}(Y|Z=z) \;=\; E(Y|X=x, Z=z) \tag{5.50}$$

[see Eq. (5.26)]. Hence, in these cases we can rewrite propositions (5.44) to (5.47) in terms of the conditional expectation values $E(Y|X=x)$ and $E(Y|X=x, Z=z)$ replacing $E^{X=x}(Y)$ by $E(Y|X=x)$ and $E^{X=x}(Y|Z=z)$ by $E(Y|X=x, Z=z)$. ◁

Example 5.47 [Nonorthogonal Factors] In the example presented in Table 1.3, conditional mean-independence of Y from U given $(X=x, Z=low)$, holds for all three values x of X, that is, for all $x \in X(\Omega)$,

$$E^{X=x,\,Z=low}(Y|U=u) \;=\; E^{X=x,\,Z=low}(Y), \quad \forall u \in U(\Omega),$$

(see the first two rows of the last three columns of Table 1.3). Similarly,

$$E^{X=x,\,Z=hi}(Y|U=u) \;=\; E^{X=x,\,Z=hi}(Y), \quad \forall u \in U(\Omega),$$

for all values x of X (see the last two rows of the last three columns of this table). In contrast, the corresponding equation for $E^{X=x,\,Z=med}(Y|U=u)$ does not hold for any value x of X (see rows three to six of the last three columns of this table). ◁

Now we generalize the concept of mean-independence of Y from Z with respect to $P^{X=x}$ to W-conditional mean-independence of Y from Z with respect to $P^{X=x}$. This concept formalizes the idea that the (Z,W)-conditional expectation of Y does not depend on Z keeping constant X at its value x, presuming that $P(X=x) > 0$.

Definition 5.48 [Conditional Mean-Independence of Y From Z With Respect to $P^{X=x}$] *Let the assumptions of Definition 5.4 hold and let W be a random variable on $(\Omega, \mathscr{A}, P^{X=x})$. Then we define W-conditional mean-independence of Y from Z with respect to $P^{X=x}$, denoted $Y \underset{P^{X=x}}{\vDash} Z \,|\, W$, by*

$$E^{X=x}(Y|Z,W) \underset{P^{X=x}}{=} E^{X=x}(Y\,|\,W). \tag{5.51}$$

Hence,

$$Y \underset{P^{X=x}}{\vDash} Z\,|\,W \quad :\Leftrightarrow \quad E^{X=x}(Y|Z,W) \underset{P^{X=x}}{=} E^{X=x}(Y\,|\,W). \tag{5.52}$$

5.7 Conditional Mean-Independence Revisited

In Definition 4.39 we introduced the concept of Z-conditional mean-independence of Y from X, denoted $Y \vDash X\,|\,Z$, by

$$E(Y|X,Z) \underset{P}{=} E(Y|Z). \tag{5.53}$$

In this section we present some propositions about Z-conditional mean-independence of Y from X involving the conditional expectation $E^{X=x}(Y|Z)$.

Theorem 5.49 [Some Proposition About Z-Conditional Mean-Independence]
Let X, Y, Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and nonnegative or with finite expectation $E(Y)$. Furthermore, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, $P(X{=}x) > 0$, and let $P^{X=x}$ denote the measure defined by Equation (5.1).

(i) Then

$$Y \vDash X \,|\, Z \quad \Rightarrow \quad E^{X=x}(Y \,|\, Z) \underset{P^{X=x}}{=} E(Y \,|\, Z) \tag{5.54}$$

$$\Leftrightarrow \quad E^{X=x}(Y \,|\, Z{=}z) = E(Y \,|\, Z{=}z), \quad \text{for } P_Z^{X=x}\text{-a. a. } z \in \Omega'_Z. \tag{5.55}$$

(ii) If, additionally, for all $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, and $P(X{=}x) > 0$, then

$$Y \vDash X \,|\, Z \quad \Leftrightarrow \quad E^{X=x}(Y \,|\, Z) \underset{P^{X=x}}{=} E(Y \,|\, Z), \quad \forall\, x \in X(\Omega). \tag{5.56}$$

(iii) If, additionally, for all pairs $(x, z) \in X(\Omega) \times Z(\Omega)$, we assume $\{X{=}x, Z{=}z\} = \{\omega \in \Omega\colon X(\omega) = x, Z(\omega) = z\} \in \mathscr{A}$, and $P^{X=x}(Z{=}z) > 0$, then

$$Y \vDash X \,|\, Z \quad \Leftrightarrow \quad E^{X=x}(Y \,|\, Z{=}z) \;=\; E(Y \,|\, Z{=}z), \quad \forall\, (x, z) \in X(\Omega) \times Z(\Omega). \tag{5.57}$$

(Proof p. 123)

In the following theorem we consider Z-conditional mean-independence from an indicator $1_{X=x}$. This theorem is useful because the right-hand sides of Propositions (5.59) and (5.60) only refer to a single value x of X. In this theorem we assume that not only P but also $P^{X \neq x}$ is a probability measure on $(\Omega, \mathscr{A})$. This measure is defined as follows: let $x \in X(\Omega)$, $\{X{\neq}x\} \in \mathscr{A}$, and $P(X{\neq}x) > 0$. Then

$$P^{X \neq x}(A) \;=\; P(A \,|\, X{\neq}x) \;=\; \frac{P(A \cap \{X{\neq}x\})}{P(X{\neq}x)}, \quad \forall\, A \in \mathscr{A}. \tag{5.58}$$

Theorem 5.50 [Z-Conditional Mean-Independence From an Indicator]
Let the assumptions of Theorem 5.49 hold, let $0 < P(X{=}x) < 1$, and let $P^{X=x}$ and $P^{X \neq x}$ denote the measures defined by Equations (5.1) and (5.58), respectively. Furtthermore, assume that $E^{X=x}(Y \,|\, Z)$ and $E^{X \neq x}(Y \,|\, Z)$ are P-unique. Then:

$$Y \vDash 1_{X=x} | Z \quad \Leftrightarrow \quad E^{X=x}(Y \,|\, Z) \underset{P}{=} E(Y \,|\, Z) \tag{5.59}$$

$$\Leftrightarrow \quad E^{X \neq x}(Y \,|\, Z) \underset{P}{=} E(Y \,|\, Z). \tag{5.60}$$

(Proof p. 123)

If Z is discrete and $P^{X=x}(Z{=}z) > 0$ for all $z \in Z(\Omega)$, then Theorem 5.50 implies the following corollary.

Corollary 5.51 [Z-Conditional Mean-Independence From an Indicator: Discrete Z]
Let the assumptions of Theorem 5.50 hold and, for all $z \in Z(\Omega)$, let $\{Z{=}z\} \in \mathscr{A}$, and $P^{X=x}(Z{=}z) > 0$. Then

Box 5.2 Glossary of new concepts

Let X, Y, W, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Let Y be numerical and nonnegative or have a finite expectation $E^{X=x}(Y)$. Furthermore, let $(\Omega'_X, \mathscr{A}'_X)$ and $(\Omega'_Z, \mathscr{A}'_Z)$ denote the value spaces of X and Z, let $x \in \Omega'_X$, $\{X=x\} \in \mathscr{A}$, and $P(X=x) > 0$.

$P^{X=x}$	The $(X=x)$-*conditional-probability measure* on the measurable space $(\Omega, \mathscr{A})$. It is defined by $P^{X=x}(A) = P(A \mid X=x), \ \forall A \in \mathscr{A}$.
$E^{X=x}(Y)$	The *expectation of Y with respect to* $P^{X=x}$. If $\int Y^+ dP^{X=x}$ or $\int Y^- dP^{X=x}$ is finite, then it is defined by $E^{X=x}(Y) = \int Y \, dP^{X=x}$.
$P_Z^{X=x}$	The *distribution of Z with respect to measure* $P^{X=x}$. A probability measure on the measurable space $(\Omega'_Z, \mathscr{A}'_Z)$ defined by $P_Z^{X=x}(A') = P^{X=x}(\{Z \in A'\}), \ \forall A' \in \mathscr{A}'_Z$.
$E^{X=x}(Y \mid Z)$	A (version of the) *Z-conditional expectation of Y with respect to* $P^{X=x}$. If Y is nonnegative or has a finite expectation $E^{X=x}(Y)$, then $E^{X=x}(Y \mid Z)$ is defined as a random variable V_x on $(\Omega, \mathscr{A}, P^{X=x})$ satisfying (a) $\sigma(V_x) \subset \sigma(Z)$ and (b) $E^{X=x}(1_C \cdot V_x) = E^{X=x}(1_C \cdot Y), \ \forall C \in \sigma(Z)$.
$\mathscr{E}^{X=x}(Y \mid Z)$	The set of all random variables V_x on $(\Omega, \mathscr{A}, P^{X=x})$ satisfying (a) and (b).
g_x	*Factorization of* $E^{X=x}(Y \mid Z)$. It is a measurable function $g_x: (\Omega'_Z, \mathscr{A}'_Z) \to (\overline{\mathbb{R}}, \overline{\mathscr{B}})$ such that $E^{X=x}(Y \mid Z) = g_x(Z)$.
$E^{X=x}(Y \mid Z=z)$	$(Z=z)$-*conditonal expectation value of Y with respect to* $P^{X=x}$. It is defined by $E^{X=x}(Y \mid Z=z) = g_x(z)$. If $P^{X=x}(Z=z) > 0$, then $E^{X=x}(Y \mid Z=z) = E(Y \mid X=x, Z=z)$ is a uniquely defined number.
$E(Y \mid X=x, Z)$	A (version of the) *partial* $(X=x, Z)$-*conditional expectation of Y*.
$\mathscr{E}(Y \mid X=x, Z)$	The set of all versions of the partial $(X=x, Z)$-conditional expectation of Y.
$P^{X=x}$-*unique*	In general, $E^{X=x}(Y \mid Z)$ is not uniquely defined. However, it is $P^{X=x}$-*unique* in the following sense: If $V_x, V_x^* \in \mathscr{E}^{X=x}(Y \mid Z)$, then $V_x \underset{P^{X=x}}{=} V_x^*$.
$P \underset{\sigma(Z)}{\ll} P^{X=x}$	*Absolute continuity of P on* $\sigma(Z)$ *with respect to* $P^{X=x}$. It is defined by $\forall A \in \sigma(Z): \ P^{X=x}(A) = 0 \ \Rightarrow \ P(A) = 0.$ It is equivalent to P-uniqueness of $E^{X=x}(Y \mid Z)$.
$Y \underset{P^{X=x}}{\vDash} Z$	*Mean-independence of Y from Z with respect to* $P^{X=x}$. It is defined by $E^{X=x}(Y \mid Z) \underset{P^{X=x}}{=} E^{X=x}(Y)$.
$Y \underset{P^{X=x}}{\vDash} Z \mid W$	*W-conditional mean-independence of Y from Z with respect to* $P^{X=x}$. It is defined by $E^{X=x}(Y \mid Z, W) \underset{P^{X=x}}{=} E^{X=x}(Y \mid W)$.

$$Y \vDash 1_{X=x} \mid Z \quad \Leftrightarrow \quad E^{X=x}(Y \mid Z=z) = E(Y \mid Z=z), \quad \forall z \in Z(\Omega) \tag{5.61}$$

$$\Leftrightarrow \quad E^{X \neq x}(Y \mid Z=z) = E(Y \mid Z=z), \quad \forall z \in Z(\Omega). \tag{5.62}$$

5.8 Summary and Conclusions

In this chapter we introduced the concept of a *conditional expectation* $E^{X=x}(Y|Z)$ *with respect to a conditional-probability measure* $P^{X=x}$. This concept can be used to describe how the Z-conditional expectation of Y depends on a random variable Z if another random variable X is kept constant at one of its values x.

A conditional expectation $E^{X=x}(Y|Z)$ with respect to $P^{X=x}$ has the general properties of a conditional expectation, but also additional properties that are related to its uniqueness. It is *uniquely defined* if $P^{X=x}(Z=z) > 0$, for all values $z \in Z(\Omega)$. By definition, $E^{X=x}(Y|Z)$ *is always* $P^{X=x}$*-unique*, which means that different versions of $E^{X=x}(Y|Z)$ are identical $P^{X=x}$-almost surely. However, $E^{X=x}(Y|Z)$ *can also be* P*-unique* so that different versions of $E^{X=x}(Y|Z)$ are identical almost surely with respect to P. In this case, the expectation $E\big(E^{X=x}(Y|Z)\big)$ of $E^{X=x}(Y|Z)$ with respect to P is a uniquely defined number that is identical for different versions $E^{X=x}(Y|Z)$ and $E^{X=x}(Y|Z)^*$ of the Z-conditional expectation of Y with respect to $P^{X=x}$.

We also introduced the more general concept of a *partial conditional expectation* $E(Y|X=x, Z)$, whose definition does not presume $P(X=x) > 0$. However, if $P(X=x) > 0$, then $E(Y|X=x, Z)$ is also a version of the Z-conditional expectation of Y, and vice versa, $E^{X=x}(Y|Z)$ is also a version of the $(X=x, Z)$-partial conditional expectation of Y.

Box 5.2 summarizes these and related concepts such as the condition expectation value $E^{X=x}(Y|Z=z)$, mean-independence of Y from Z with respect to $P^{X=x}$, and absolute continuity of P on the σ-algebra $\sigma(Z)$ with respect to $P^{X=x}$. The latter is equivalent to P-uniqueness of $E^{X=x}(Y|Z)$ and to $P(X=x|Z) \underset{P}{>} 0$ (see Th. 5.27).

5.9 Proofs

Proof of Lemma 5.17

$$\begin{aligned} P(A) = 0 \quad &\Rightarrow \quad P(A \cap \{X=x\}) = 0 && \text{[Box 1.1 (v)]}\\ &\Rightarrow \quad P^{X=x}(A) = 0. && \text{[(5.1)]} \end{aligned}$$

Proof of Lemma 5.20

$$\begin{aligned} & X \underset{P^{X=x}}{=} Y \ \wedge\ P \underset{\mathscr{A}}{\ll} P^{X=x} \\ \Leftrightarrow\quad & P^{X=x}(\{X \neq Y\}) = 0 \ \wedge\ \big(\forall A \in \mathscr{A}: P^{X=x}(A) = 0 \Rightarrow P(A) = 0\big) && \text{[Defs. 2.35, 5.16]}\\ \Rightarrow\quad & P(\{X \neq Y\}) = 0 && [\{X \neq Y\} \in \mathscr{A}]\\ \Leftrightarrow\quad & X \underset{P}{=} Y. && \text{[Def. 2.35]} \end{aligned}$$

Proof of Lemma 5.24

We start proving that $X(\Omega)$ is finite or countable, if $P(X{=}x) > 0$ for all $x \in X(\Omega)$, using the notation $\#\{A\}$ for the number of elements in the set A.

$$\#\left\{x \in X(\Omega) : P(X{=}x) \geq \frac{1}{n}\right\} \leq n. \qquad [P(\Omega) = 1, \text{additivity of } P]$$

Furthermore,

$$X(\Omega) \;=\; \bigcup_{n \in \mathbb{N}} \left\{x \in X(\Omega) : P(X{=}x) \geq \frac{1}{n}\right\}. \qquad [\forall\, x \in X(\Omega) : P(X{=}x) > 0]$$

However, a countable union of finite sets is finite or countable.

Now we show that for $A \in \mathscr{A}$, $P(A) > 0$ is equivalent to $\exists\, x \in X(\Omega) : P^{X=x}(A) > 0$.

$$\Omega \;=\; \bigcup_{x \in X(\Omega)} \{X{=}x\} \qquad [\text{def. of } X(\Omega)]$$

and $\{X{=}x\} \cap \{X{=}x'\} = \emptyset$ if $x \neq x'$ (def. of a mapping). Hence, for $A \subset \Omega$,

$$A \;=\; A \cap \Omega \;=\; \bigcup_{x \in X(\Omega)} A \cap \{X{=}x\} \tag{5.63}$$

and

$$(A \cap \{X{=}x\}) \cap (A \cap \{X{=}x'\}) \;=\; \emptyset, \quad \text{if } x \neq x'. \tag{5.64}$$

Hence, for $A \in \mathscr{A}$,

$$P(A) \;=\; \sum_{x \in X(\Omega)} P(A \cap \{X{=}x\}) \qquad [(5.63), (5.64), \sigma\text{-additivity of } P]$$

and therefore

$$\begin{aligned} & P(A) > 0 \\ \Leftrightarrow \quad & \exists\, x \in X(\Omega) : P(A \cap \{X{=}x\}) > 0 \qquad [P \text{ is nonnegative}] \\ \Leftrightarrow \quad & \exists\, x \in X(\Omega) : P^{X=x}(A) > 0. \end{aligned} \tag{5.65}$$

Finally, we prove Proposition (5.18) by contradiction:

$$\begin{aligned} & \neg\left(\sum_{x \in X(\Omega)} Z_x \cdot 1_{X=x} \underset{P}{\in} B\right) \\ \Leftrightarrow \quad & \exists\, A \in \mathscr{A} : P(A) > 0 \;\wedge\; \forall \omega \in A : \left(\sum_{x \in X(\Omega)} Z_x(\omega) \cdot 1_{X=x}(\omega)\right) \notin B \\ \Leftrightarrow \quad & \exists\, A \in \mathscr{A} : \exists\, x \in X(\Omega) : P^{X=x}(A \cap \{X{=}x\}) > 0 \;\wedge\; \forall x \in A \cap \{X{=}x\} : Z_x(\omega) \notin B \\ \Leftrightarrow \quad & \neg(\forall\, x \in X(\Omega) : Z_x \underset{P^{X=x}}{\in} B). \end{aligned}$$

The next to last equivalence holds because of (5.65), $P^{X=x}(A) = P^{X=x}(A \cap \{X{=}x\})$, and

$$\forall \omega \in A \cap \{X{=}x\} : 1_{X=x}(\omega) = 1.$$

Proof of Theorem 5.49

(i).

$$\begin{aligned}
& E(Y|X,Z),\ E^{X=x}(Y|Z) \in \mathscr{E}^{X=x}(Y|X,Z) && \text{[SN-Th. 14.64]}\\
\Rightarrow\ & E(Y|X,Z) \underset{P^{X=x}}{=} E^{X=x}(Y|X,Z) \underset{P^{X=x}}{=} E^{X=x}(Y|Z) && \text{[SN-Eq. (14.72)]}\\
\Rightarrow\ & E(Y|Z) \underset{P^{X=x}}{=} E^{X=x}(Y|Z). && [Y \vDash X|Z \Rightarrow E(Y|Z) \in \mathscr{E}(Y|X,Z)]
\end{aligned}$$

Now Proposition (5.55) follows immediately from Propositions (5.44) and (5.45).
(ii). The first implication, namely

$$Y \vDash X|Z \quad \Rightarrow \quad E^{X=x}(Y|Z) \underset{P^{X=x}}{=} E(Y|Z), \ \ \forall\, x \in X(\Omega),$$

immediately follows from (i). Hence, we only have to prove

$$E^{X=x}(Y|Z) \underset{P^{X=x}}{=} E(Y|Z), \ \ \forall\, x \in X(\Omega) \quad \Rightarrow \quad Y \vDash X|Z.$$

Now

$$\begin{aligned}
E(Y|X,Z) &\underset{P}{=} \sum_{x \in X(\Omega)} E^{X=x}(Y|Z) \cdot 1_{X=x} && \text{[(5.36)]}\\
&\underset{P}{=} \sum_{x \in X(\Omega)} E(Y|Z) \cdot 1_{X=x} && [E(Y|Z) \in \mathscr{E}^{X=x}(Y|Z),\ \forall\, x \in X(\Omega)]\\
&\underset{P}{=} E(Y|Z) \cdot \sum_{x \in X(\Omega)} 1_{X=x} \underset{P}{=} E(Y|Z) \cdot 1_\Omega \underset{P}{=} E(Y|Z) \cdot 1\\
&\underset{P}{=} E(Y|Z).
\end{aligned}$$

(iii). This proposition follows from Theorem 4.40 (i), $P^{X=x}(Z{=}z) > 0$ for all $(x,z) \in X(\Omega) \times Z(\Omega)$, and Equation (5.26).

Proof of Theorem 5.50

(5.59). We start proving $Y \vDash 1_{X=x}|Z \Rightarrow E^{X=x}(Y|Z) \underset{P^{X=x}}{=} E(Y|Z)$.

$$\begin{aligned}
& Y \vDash 1_{X=x}|Z\\
\Rightarrow\ & E^{X=x}(Y|Z) \underset{P^{X=x}}{=} E(Y|Z) && \text{[Th. 5.49 (i)]}\\
\Leftrightarrow\ & E^{X=x}(Y|Z) \underset{P}{=} E(Y|Z). && [E^{X=x}(Y|Z) \text{ is } P\text{-unique, Th. 5.27 (c), (5.16)}]
\end{aligned}$$

Now we prove $E^{X=x}(Y|Z) \underset{P}{=} E(Y|Z) \Rightarrow Y \vDash 1_{X=x}|Z$.

$$\begin{aligned}
& E(Y|1_{X=x},Z) \underset{P}{=} E^{X=x}(Y|Z) \cdot 1_{X=x} + E^{X\neq x}(Y|Z) \cdot 1_{X\neq x} && \text{[(5.37)]}\\
\Rightarrow\ & E(Y|1_{X=x},Z) \underset{P}{=} E(Y|Z) \cdot 1_{X=x} + E^{X\neq x}(Y|Z) \cdot 1_{X\neq x} && [E^{X=x}(Y|Z) \underset{P}{=} E(Y|Z)]\\
\Rightarrow\ & E(Y|1_{X=x},Z) \cdot 1_{X=x} \underset{P}{=} E(Y|Z) \cdot 1_{X=x} && [\,|\cdot 1_{X=x},\ 1_{X=x}^2 = 1_{X=x},\ 1_{X=x} \cdot 1_{X\neq x} = 0\,]\\
\Rightarrow\ & E(Y|1_{X=x},Z) \underset{P^{X=x}}{=} E(Y|Z) && \text{[Rem. 2.46]}
\end{aligned}$$

$\Rightarrow \quad E(Y|1_{X=x}, Z) \underset{P}{=} E(Y|Z) \qquad [E^{X=x}(Y|Z)$ is P-unique, Th. 5.27 (c), (5.16)]

$\Leftrightarrow \quad Y \vDash 1_{X=x}|Z,$ [Def. 4.39]

where $|\cdot 1_{X=x}$ means multiplying both sides of the equation by $1_{X=x}$.

(5.60).

$$\begin{aligned} Y \vDash 1_{X=x}|Z \quad &\Leftrightarrow \quad E(Y|1_{X=x}, Z) \underset{P}{=} E(Y|Z) && \text{[Def. 4.39]} \\ &\Leftrightarrow \quad E(Y|1_{X\neq x}, Z) \underset{P}{=} E(Y|Z) && [\sigma(1_{X=x}, Z) = \sigma(1_{X\neq x}, Z)] \\ &\Leftrightarrow \quad Y \vDash 1_{X\neq x}|Z && \text{[Def. 4.39]} \\ &\Leftrightarrow \quad E^{X\neq x}(Y|Z) \underset{P}{=} E(Y|Z). && [(5.59)] \end{aligned}$$

5.10 Exercises

▷ **Exercise 5-1** Specify the probability measure $P^{X=1}$ in the random experiment presented in Table 2.1.

▷ **Exercise 5-2** Consider Table 2.1 and compute the expectations (with respect to P) of the two different versions $P^{X=1}(Y=1|U)$, $P^{X=1}(Y=1|U)^* \in \mathscr{E}^{X=1}(Y|U)$ mentioned in Example 5.28.

▷ **Exercise 5-3** Consider Example 5.28 and show that $P^{X=1}(\{\omega \in \Omega\colon V_1(\omega) = V_1^*(\omega)\}) = 1$ holds for $V_1 = P^{X=1}(Y=1|U)$ and $V_1^* = P^{X=1}(Y=1|U)^*$, using the measure $P^{X=1}$ specified in Exercise 5-1.

▷ **Exercise 5-4** What does it mean when we assume that the conditional expectation $E^{X=x}(Y|Z)$ is P-unique?

▷ **Exercise 5-5** Which are the values of $E^{X=0}(Y|U)$ and the values of the conditional expectation $E(Y|X,U)$ for $\omega_4 = (\mathit{Joe}, \mathit{yes}, +)$ in the example presented in Table 1.2?

▷ **Exercise 5-6** Which are the values of a Z-conditional expectation $E^{X=x}(Y|Z)$ of Y with respect to the probability measure $P^{X=x}$?

▷ **Exercise 5-7** Compute the values of the conditional expectation $E(Y|X) = P(Y=1|X)$ in the example presented in Table 2.1.

▷ **Exercise 5-8** Compute the values of the conditional expectation $E(Y|X,U) = P(Y=1|X,U)$ in the example presented in Table 2.1.

Solutions

▷ **Solution 5-1** The measure $P^{X=1}$ is specified if the probabilities $P^{X=1}(\{\omega_i\})$ are specified for all eight elementary events $\{\omega_1\}, \dots, \{\omega_8\}$. The probabilities of all other events can be computed from the probabilities of these eight elementary events [see Box 1.1 (x)]. Hence, we just have to compute the following probabilities:

$$\text{For } i = 1, \dots, 6: \quad P^{X=1}(\{\omega_i\}) = \frac{P(\{\omega_i\} \cap \{X=1\})}{P(X=1)} = \frac{0}{.12 + .08} = 0.$$

$$P^{X=1}(\{\omega_7\}) = \frac{P(\{\omega_7\} \cap \{X=1\})}{P(X=1)} = \frac{.12}{.12+.08} = .6.$$

$$P^{X=1}(\{\omega_8\}) = \frac{P(\{\omega_8\} \cap \{X=1\})}{P(X=1)} = \frac{.08}{.12+.08} = .4.$$

▷ **Solution 5-2** The first version $V_1 = P^{X=1}(Y=1|U) \in \mathscr{E}^{X=1}(Y|U)$ is displayed in the last but one column of Table 2.1. The second version, $V_1^* = P^{X=1}(Y=1|U)^*$, is obtained by replacing the number 99 by the number .8 in the first four lines of the table. Hence,

$$E(V_1) = 99 \cdot .5 + .4 \cdot .5 = 49.7 \qquad \text{and} \qquad E(V_1^*) = .8 \cdot .5 + .4 \cdot .5 = .6.$$

This shows that the two expectations are meaningless from a substantive point of view.

▷ **Solution 5-3** The values of V_1 and V_1^* only differ for the outcomes $\omega_1, \ldots, \omega_4$. According to Exercise 5-1, $P^{X=1}(\{\omega_1, \ldots, \omega_4\}) = 0$. Hence, $P^{X=1}(\{\omega \in \Omega\colon V_1(\omega) = V_1^*(\omega)\}) = 1$.

▷ **Solution 5-4** By definition, there may be different versions $E^{X=x}(Y|Z)$ of the Z-conditional expectation of Y with respect to the probability measure $P^{X=x}$. In general, all pairs of such versions are identical, $P^{X=1}$-almost surely [see Eq. (5.11)]. If we additionally assume that $E^{X=x}(Y|Z)$ is P-unique, then we assume that all pairs of versions $V_x, V_x^* \in \mathscr{E}^{X=x}(Y|Z)$ are identical P-almost surely [see Eq. (5.13)].

▷ **Solution 5-5** $E^{X=0}(Y|U)(\omega_4) = E(Y|X=0, U=\mathit{Joe}) = .7$. In contrast, the value of $E(Y|X,U)$ is $E(Y|X,U)(\omega_4) = E(Y|X=1, U=\mathit{Joe}) = .8$ (see the fourth row in Table 1.2).

▷ **Solution 5-6** The values of $E^{X=x}(Y|Z)$ are identical to the conditional expectation values $E^{X=x}(Y|Z=z)$. In more formal terms, $E^{X=x}(Y|Z)(\omega) = E^{X=x}(Y|Z=z)$, if $\omega \in Z^{-1}(\{z\})$. If P is a probability measure on $(\Omega, \mathscr{A})$ and $P(X=x, Z=z) > 0$, then $E^{X=x}(Y|Z=z) = E(Y|X=x, Z=z)$.

▷ **Solution 5-7** The values of the conditional expectation $E(Y|X)$ are the two conditional expectation values $E(Y|X=0)$ and $E(Y|X=1)$. Because $E(Y|X=x) = P(Y=1|X=x)$, they can be computed from Table 2.1 as follows:

$$P(Y=1|X=0) = \frac{P(Y=1, X=0)}{P(X=0)} = \frac{.35+.06}{.15+.35+.24+.06} = .5125,$$

$$P(Y=1|X=1) = \frac{P(Y=1, X=1)}{P(X=1)} = \frac{0+.08}{0+0+.12+.08} = .4.$$

▷ **Solution 5-8** The values of the conditional expectation $E(Y|X,U) = P(Y=1|X,U)$ in Table 2.1 are the conditional probabilities $P(Y=1|X=x, U=u)$, which can be computed as follows:

$$P(Y=1|X=0, U=\mathit{Joe}) = \frac{P(Y=1, X=0, U=\mathit{Joe})}{P(X=0, U=\mathit{Joe})} = \frac{.35}{.15+.35} = .7,$$

$$P(Y=1|X=0, U=\mathit{Ann}) = \frac{P(Y=1, X=0, U=\mathit{Ann})}{P(X=0, U=\mathit{Ann})} = \frac{.06}{.24+.06} = .2,$$

$$P(Y=1|X=1, U=\mathit{Ann}) = \frac{P(Y=1, X=1, U=\mathit{Ann})}{P(X=1, U=\mathit{Ann})} = \frac{.08}{.12+.08} = .4.$$

The conditional probability $P(Y=1|X=1, U=\mathit{Joe})$ is undefined, because $P(X=1, U=\mathit{Joe}) = 0$. Choosing any number (such as 99) as a value of $P(Y=1|X,U)$ for $\omega_3 = (\mathit{Joe}, \mathit{yes}, -)$ and $\omega_4 = (\mathit{Joe}, \mathit{yes}, +)$ yields a version of $P(Y=1|X,U)$. Different versions of $P(Y=1|X,U)$ are identical almost surely with respect to the measure P [see Eq. (4.7)].

Chapter 6
Conditional Independence

In chapter 4, we introduced the concept of a conditional expectation and dealt with its properties in some detail. In the first section of this chapter, this concept is used to define *conditional independence of events* and *of random variables given a random variable*. In the second section, we revisit (unconditional) *independence of events* and *of random variables* as a special case of conditional independence given a random variable and add some useful properties of independence. In the third and fourth section, we turn to conditional independence and (unconditional) independence with respect to a probability measure $P^{X=x}$, focussing on properties that involve both measures, P and $P^{X=x}$.

6.1 Conditional Independence Given a Random Variable

In chapter 4, we introduced the concept of a conditional expectation. Now this concept is used to define conditional independence of *events* given a random variable and conditional independence of *random variables* given a random variable.

6.1.1 Conditional Independence of Events

Independence of two events A and B has already been introduced in Definition 1.39. Now we define *conditional independence* of two events A and B *given a random variable*. Remember, in Remark 4.10 we introduced $P(A|X)$, an X-conditional probability of the event A by

$$P(A|X) \underset{P}{=} E(1_A|X), \tag{6.1}$$

where $E(1_A|X)$ denotes a (version of the) X-conditional expectation of the indicator of the event A.

Definition 6.1 [X-**Conditional Independence of Two Events**]
Let X be a random variable on the probability space $(\Omega, \mathscr{A}, P)$ and let $A, B \in \mathscr{A}$. Then the events A and B are called X-conditionally independent (with respect to P), denoted $A \perp\!\!\!\perp B | X$, if

$$P(A \cap B|X) \underset{P}{=} P(A|X) \cdot P(B|X). \tag{6.2}$$

If there is ambiguity with respect to the probability measure, then we also add the reference to the measure and use the notation $A \underset{P}{\perp\!\!\!\perp} B|X$ instead of $A \perp\!\!\!\perp B|X$.

6.1.2 Conditional Independence of Random Variables

Now we define *conditional independence of two random variables Y and Z given a random variable X*, generalizing the concept of independence of two random variables (see sect. 2.4). Reading the following definition, remember that the terms $\sigma(Y)$ and $\sigma(Z)$, the σ-algebras generated by Y and by Z, have been introduced in Definition 2.12.

Definition 6.2 [X-Conditional Independence of Two Random Variables]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Then Y and Z are called X-conditionally independent (with respect to P), denoted $Y \perp\!\!\!\perp Z \,|\, X$ if

$$P(A \cap B \,|\, X) \underset{P}{=} P(A \,|\, X) \cdot P(B \,|\, X), \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z). \tag{6.3}$$

Comparing Equations (6.2) and (6.3) to each other shows

$$Y \perp\!\!\!\perp Z \,|\, X \quad \Leftrightarrow \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z): \; A \perp\!\!\!\perp B \,|\, X. \tag{6.4}$$

Hence, two random variables Y and Z are called conditionally independent given a random variable X if all pairs of events $(A, B) \in \sigma(Y) \times \sigma(Z)$ are X-conditionally independent. Again, if there is ambiguity with respect to the probability measure, then we add an explicite reference to the measure. Hence, in such a case we use $Y \underset{P}{\perp\!\!\!\perp} Z \,|\, X$ instead of $Y \perp\!\!\!\perp Z \,|\, X$ if conditional independence of X and Y is meant with respect to the measure P.

Some properties of conditional independence are gathered in Box 6.1. This box has been adapted from SN-Box 16.3. Proofs are found in the solution to SN-Exercise 16-3. Note that these properties hold for any random variables X, Y, Z, and W on the same probability space. None of them has to be numerical and none has to be discrete.

Remark 6.3 [An Implication of Independence] Of course, the list of properties of conditional independence displayed in Box 6.1 is not complete. For example, if X, Y, and Z are random variables on the probability space $(\Omega, \mathscr{A}, P)$, then another useful proposition about independence and conditional independence of two random variables is

$$\big(Y \perp\!\!\!\perp Z \;\wedge\; \sigma(X) \subset \sigma(Y)\big) \;\Rightarrow\; \big(X \perp\!\!\!\perp Z \;\wedge\; Y \perp\!\!\!\perp Z \,|\, X\big). \tag{6.5}$$

This proposition follows from Box 6.1 (ix), because $Y \perp\!\!\!\perp Z \Leftrightarrow (X, Y) \perp\!\!\!\perp Z$, if $\sigma(X) \subset \sigma(Y)$. Hence, if Y and Z are independent and X is Y-measurable, then this implies independence of X and Z as well as X-conditional independence of Y and Z. ◁

In the following theorem we present an implication of conditional independence that has some appeal to our intuition about conditional independence of random variables. According to Proposition (6.6), $Y \perp\!\!\!\perp Z \,|\, X$ implies that the (X, Z)-conditional probability of the event $\{Y = y\}$ that Y takes on the value y does not depend on Z, P-almost surely, once we condition on X. (For a proof see SN-Cor. 16.24).

Box 6.1 Conditional Independence of Random Variables

Let X, Y, Z, and W be random variables on a probability space $(\Omega, \mathscr{A}, P)$. Then:

$$Y \perp\!\!\!\perp Z \,|\, X \;\Leftrightarrow\; Z \perp\!\!\!\perp Y \,|\, X \tag{i}$$

$$Y \perp\!\!\!\perp (W,Z) \,|\, X \;\Rightarrow\; Y \perp\!\!\!\perp W \,|\, X \;\wedge\; Y \perp\!\!\!\perp Z \,|\, X \tag{ii}$$

$$X \perp\!\!\!\perp Y \perp\!\!\!\perp Z \;\Rightarrow\; Y \perp\!\!\!\perp Z \,|\, X \tag{iii}$$

$$\sigma(Z) \subset \sigma(X) \;\Rightarrow\; Y \perp\!\!\!\perp Z \,|\, X \tag{iv}$$

$$Z \underset{P}{=} z,\; z \in \Omega'_Z \;\Rightarrow\; Y \perp\!\!\!\perp Z \,|\, X, \quad \text{where } \Omega'_Z \text{ is the co-domain of } Z \tag{v}$$

$$\big(Y \perp\!\!\!\perp Z \,|\, X \;\wedge\; \sigma(W) \subset \sigma(Z)\big) \;\Rightarrow\; Y \perp\!\!\!\perp W \,|\, X \tag{vi}$$

$$Y \perp\!\!\!\perp Z \,|\, X \;\Leftrightarrow\; Y \perp\!\!\!\perp (X,Z) \,|\, X \tag{vii}$$

$$Y \perp\!\!\!\perp (W,Z) \,|\, X \;\Leftrightarrow\; Y \perp\!\!\!\perp Z \,|\, X \;\wedge\; Y \perp\!\!\!\perp W \,|\, (X,Z) \tag{viii}$$

$$Y \perp\!\!\!\perp (X,Z) \;\Leftrightarrow\; Y \perp\!\!\!\perp X \;\wedge\; Y \perp\!\!\!\perp Z \,|\, X \tag{ix}$$

$$\big(Z \underset{P}{=} W \;\wedge\; Y \perp\!\!\!\perp Z \,|\, X\big) \;\Rightarrow\; Y \perp\!\!\!\perp W \,|\, X. \tag{x}$$

If Y_0 and Z_0 are random variables on $(\Omega, \mathscr{A}, P)$ that are measurable with respect to Y and Z, respectively, then

$$Y \perp\!\!\!\perp Z \,|\, X \;\Rightarrow\; Y \perp\!\!\!\perp Z \,|\, (X, Y_0, Z_0). \tag{xi}$$

Theorem 6.4 [An Implication of Conditional Independence]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let Ω'_Y denote the co-domain of Y, let $y \in \Omega'_Y$, and $\{Y=y\} \in \mathscr{A}$. Then

$$Y \perp\!\!\!\perp Z \,|\, X \quad\Rightarrow\quad P(Y{=}y \,|\, X, Z) \underset{P}{=} P(Y{=}y \,|\, X). \tag{6.6}$$

Note that Theorem 6.4 applies irrespective of whether or not Y is discrete. In contrast, in the following theorem, we consider the case in which the image $Y(\Omega)$ of Y is finite or countable. In this case $Y \perp\!\!\!\perp Z \,|\, X$ does not only imply the equation on the right-hand of Proposition (6.6). Instead, if this equation holds for all values y of Y, then the resulting proposition is equivalent to $Y \perp\!\!\!\perp Z \,|\, X$ [see Prop. (6.7)]. (For a proof see SN-Theorem 16.26.) Furthermore, if $Y(\Omega)$ is finite or countable, then, according to Proposition (6.8), $Y \perp\!\!\!\perp Z \,|\, X$ is also equivalent to the right-hand side of Proposition (6.8).

Theorem 6.5 [A Property Equivalent to Conditional Independence]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Furthermore, assume that $Y(\Omega)$ is finite or countable and let $\{Y=y\} \in \mathscr{A}$, for all $y \in Y(\Omega)$. Then

$$Y \perp\!\!\!\perp Z \,|\, X \quad\Leftrightarrow\quad \forall\, y \in Y(\Omega)\colon\; P(Y{=}y \,|\, X, Z) \underset{P}{=} P(Y{=}y \,|\, X) \tag{6.7}$$

$$\Leftrightarrow \quad \forall\, y \in Y(\Omega){:}\ \mathit{1}_{Y=y} \perp\!\!\!\perp Z \,|\, X. \tag{6.8}$$

Now we turn to an even more special case, considering a single value y of Y and the indicator $\mathit{1}_{Y=y}$. That is, we consider the special case $\mathit{1}_{Y=y} \perp\!\!\!\perp Z \,|\, X$. Although the indicator $\mathit{1}_{Y=y}$ has two different values, 0 and 1, we only have to show $P(\mathit{1}_{Y=y}{=}1 \,|\, X, Z) \underset{P}{=} P(\mathit{1}_{Y=y}{=}1 \,|\, X)$ in order to prove $\mathit{1}_{Y=y} \perp\!\!\!\perp Z \,|\, X$. The corresponding equation for $P(\mathit{1}_{Y=y}{=}0 \,|\, X, Z)$ follows. Finally, because $\{Y{=}y\} = \{\mathit{1}_{Y=y}{=}1\} = \{\omega \in \Omega{:}\ Y(\omega) = y\}$, we can also conclude

$$\mathit{1}_{Y=y} \perp\!\!\!\perp Z \,|\, X \quad \Leftrightarrow \quad P(Y{=}y \,|\, X, Z) \underset{P}{=} P(Y{=}y \,|\, X) \tag{6.9}$$

(see Exercise 6-1). Hence, in order to prove that Z and the indicator $\mathit{1}_{Y=y}$ are X-conditionally independent, we only have to show that the conditional probability of the event $\{Y{=}y\}$ does not depend on Z, once we condition on X. This will be exemplified in Example 6.7.

Theorem 6.6 [Conditional Independence Involving an Indicator]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let Ω'_Y denote the co-domain of Y, let $y \in \Omega'_Y$, and $\{Y{=}y\} \in \mathscr{A}$. Then the following propositions are equivalent to each other:

(a) $\mathit{1}_{Y=y} \perp\!\!\!\perp Z \,|\, X$

(b) $P(\mathit{1}_{Y=y}{=}1 \,|\, X, Z) \underset{P}{=} P(\mathit{1}_{Y=y}{=}1 \,|\, X)$

(c) $P(\mathit{1}_{Y=y}{=}0 \,|\, X, Z) \underset{P}{=} P(\mathit{1}_{Y=y}{=}0 \,|\, X)$

(d) $P(Y{=}y \,|\, X, Z) \underset{P}{=} P(Y{=}y \,|\, X)$

(e) $P(Y{\neq}y \,|\, X, Z) \underset{P}{=} P(Y{\neq}y \,|\, X)$.

(Proof p. 143)

Example 6.7 [Mortality] Table 6.1 displays the parameters of a random experiment that consists of sampling a person from the set $\Omega_U = \{\mathit{Joe}, \mathit{Jim}, \mathit{Ann}, \mathit{Sue}\}$ of four persons and observing whether $(Y{=}1)$ or not $(Y{=}0)$ this person passes away in a specified year. *Sex* and *age* of these persons are displayed as well, while, for simplicity, other attributes of these persons (such as their physical, mental health, social support) are not.

The set of possible outcomes of this random experiment is

$$\Omega \;=\; \Omega_U \times \Omega_Y,$$

where $\Omega_Y = \{\mathit{dies},\ \mathit{survives}\}$, and the elements of this set refer to the specified year. Hence, Ω has eight elements, the pairs $(\mathit{Joe}, \mathit{dies}), (\mathit{Joe}, \mathit{survives}), \ldots, (\mathit{Sue}, \mathit{survives})$. The σ-algebra on Ω is its power set, that is, $\mathscr{A} = \mathscr{P}(\Omega)$, which has $2^8 = 256$ elements. The probability measure P on $\mathscr{A}$ is completely specified by the parameters $P(\{\omega_i\})$ listed in the second column of Table 6.1. The probability of each of the 256 elements can be computed from these eight parameters, because, except for the empty set, each of these events is a finite union of the eight elementary events $\{\omega_i\}$.

Table 6.1. Mortality

Possible outcomes ω_i			Observables				Conditional probabilities			
Unit	Outcome	$P(\{\omega_i\})$	Person variable U	Sex W	Age Z	Outcome variable Y	$P(Y=1\|U)$	$P(Y=1\|W)$	$P(Y=1\|Z)$	$P(Y=1\|W,Z)$
ω_1 = (*Joe, survives*)		.24	*Joe*	*m*	86	0	.04	.05	.04	.04
ω_2 = (*Joe, dies*)		.01	*Joe*	*m*	86	1	.04	.05	.04	.04
ω_3 = (*Jim, survives*)		.235	*Jim*	*m*	88	0	.06	.05	.06	.06
ω_4 = (*Jim, dies*)		.015	*Jim*	*m*	88	1	.06	.05	.06	.06
ω_5 = (*Ann, survives*)		.24	*Ann*	*f*	86	0	.04	.06	.04	.04
ω_6 = (*Ann, dies*)		.01	*Ann*	*f*	86	1	.04	.06	.04	.04
ω_7 = (*Sue, survives*)		.23	*Sue*	*f*	92	0	.08	.06	.08	.08
ω_8 = (*Sue, dies*)		.02	*Sue*	*f*	92	1	.08	.06	.08	.08

Comparing the corresponding columns of Table 6.1 shows

$$P(Y=1|U) \underset{P}{=} P(Y=1|Z) \underset{P}{=} P(Y=1|W,Z).$$

Hence, $W \perp\!\!\!\perp Y \,|\, Z$ holds, because Y is an indicator variable so that we can apply Theorem 6.6. For details of computing the values of $P(Y=1|W,Z)$ and $P(Y=1|Z)$ from the probabilities $P(\{\omega_i\})$, see Exercise 6-2. ◁

6.1.3 *Conditional Independence and Conditional Mean-Independence*

Now we address the relationship between conditional independence and conditional mean-independence of two random variables. Remember, the notation $Y \vDash Z \,|\, X$ is a shortcut for

$$E(Y|X,Z) \underset{P}{=} E(Y|X) \tag{6.10}$$

(see Def. 4.34).

Theorem 6.8 [Conditional Independence Implies Conditional Mean-Independence] *Let X, Y, and Z be random variables on $(\Omega, \mathscr{A}, P)$ and assume that Y is numerical and nonnegative or with finite expectation $E(Y)$. Then*

$$Y \perp\!\!\!\perp Z \,|\, X \quad \Rightarrow \quad Y \vDash Z \,|\, X. \tag{6.11}$$

(For a proof, see SN-Rem. 16.36). According to this theorem, if $E(Y|X)$ is defined, then X-conditional independence of Y and Z implies X-conditional mean-independence of

Y from Z. Note that, in contrast to Y, neither X nor Z have to be numerical. They can be unidimensional or multidimensional random variables consisting of several random variables that do not have to be numerical.

Remember, according to Theorem 6.6,

$$1_{Y=y} \perp\!\!\!\perp Z \mid X \quad \Leftrightarrow \quad P(Y=y \mid X, Z) \underset{P}{=} P(Y=y \mid X). \tag{6.12}$$

Because $P(Y=y \mid X,Z) \underset{P}{=} E(1_{Y=y} \mid X,Z)$ and $P(Y=y \mid X) \underset{P}{=} E(1_{Y=y} \mid X)$ [see Eq. (4.10)], this implies

$$1_{Y=y} \perp\!\!\!\perp Z \mid X \quad \Leftrightarrow \quad 1_{Y=y} \vDash Z \mid X. \tag{6.13}$$

6.1.4 Conditional Independence of a Family of Events

Now we turn to conditional independence of a family of events given a random variable. Such a family of events $(A_i, i \in I)$ may consist, for example, of two $(I = \{1,2\})$, a finite number $(I = \{1,\dots,m\})$, or a countable number $(I = \{1,2,\dots\})$ of events. The index set I may even be uncountable.

Definition 6.9 [Conditional Independence of a Family of Events]
Let X be a random variable on the probability space $(\Omega, \mathscr{A}, P)$. Furthermore, let I be a nonempty set and let $A_i \in \mathscr{A}$, $i \in I$. Then $(A_i, i \in I)$ is called a family of X-conditionally independent events, denoted $\perp\!\!\!\perp (A_i, i \in I) \mid X$, if

$$P\Big(\bigcap_{i \in J} A_i \,\Big|\, X\Big) \underset{P}{=} \prod_{i \in J} P(A_i \mid X), \quad \forall \text{ finite nonempty } J \subset I. \tag{6.14}$$

Remark 6.10 [Sequence of Conditionally Independent Events] If the index set I is finite, that is, if $I = \{1,\dots,m\}$, then a family $(A_i, i \in I)$ of events is also called a *finite sequence of events*. In this case, we also use the notation $A_1,\dots,A_m$ instead of $(A_i, i \in I)$. Similarly, in this case, instead of $\perp\!\!\!\perp (A_i, i \in I) \mid X$, we use the notation $A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_m \mid X$ and say that the events $A_1,\dots,A_m$ are X-conditionally independent. For $m = 3$ events, for instance, $A_1 \perp\!\!\!\perp A_2 \perp\!\!\!\perp A_3 \mid X$ means that

$$P(A_i \cap A_j \mid X) \underset{P}{=} P(A_i \mid X) \cdot P(A_j \mid X), \quad i \neq j, \quad i,j = 1,2,3, \tag{6.15}$$

(*pairwise X-conditional independence*) and

$$P(A_1 \cap A_2 \cap A_3 \mid X) \underset{P}{=} P(A_1 \mid X) \cdot P(A_2 \mid X) \cdot P(A_3 \mid X) \tag{6.16}$$

(*triplewise X-conditional independence*). Equation (6.15) results from Equation (6.14) if we consider the subsets $J_1 = \{1,2\}$, $J_2 = \{1,3\}$, and $J_3 = \{2,3\}$ of $I = \{1,2,3\}$, and Equation (6.16) results from Equation (6.14) if we consider the subset $J = I$ of I. ◁

Remark 6.11 [Triplewise and Pairwise Conditional Independence] Note that pairwise conditional independence of three events [i.e., Eq. (6.15)] does not imply their triple-wise conditional independence [i.e., Eq. (6.16)]. And vice versa, triplewise conditional independence of events does not imply their pairwise conditional independence. ◁

6.1.5 Family of Conditionally Independent Random Variables

Now we turn to the concept of a family $(Y_i, i \in I)$ of conditionally independent random variables given a random variable. Again, the index set I may be finite, countable, or uncountable.

Definition 6.12 [Family of Conditionally Independent Random Variables]
Let $(Y_i, i \in I)$ be a family of random variables and X a random variable on the probability space $(\Omega, \mathscr{A}, P)$. Then $(Y_i, i \in I)$ is called a family of X-conditionally independent random variables, denoted $\perp\!\!\!\perp(Y_i, i \in I)\,|\,X$, if each family of events $(A_i, i \in I)$, $A_i \in \sigma(Y_i)$, $i \in I$, is X-conditionally independent.

Remark 6.13 [Sequence of Conditionally Independent Random Variables] Analogously to what we said for a sequence of events (see Rem. 6.10), if $I = \{1, \ldots, m\}$, then we also use the notation $Y_1, \ldots, Y_m$ instead of $(Y_i, i \in I)$ and call it a *finite sequence of random variables*. Similarly, in this case we also use the notation $Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_m | X$ instead of $\perp\!\!\!\perp(Y_i, i \in I)\,|\,X$ and say that the random variables $Y_1, \ldots, Y_m$ are *X-conditionally independent*. Note that $Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_m | X$ is equivalent to

$$\forall (A_1, \ldots, A_m) \in \sigma(Y_1) \times \ldots \times \sigma(Y_m):\ P(A_1 \cap \ldots \cap A_m | X) \underset{P}{=} P(A_1 | X) \cdot \ldots \cdot P(A_m | X). \tag{6.17}$$

In contrast, $P(A_1 \cap \ldots \cap A_m | X) \underset{P}{=} P(A_1 | X) \cdot \ldots \cdot P(A_m | X)$ is *not* equivalent to X-conditional independence of the events $A_1, \ldots, A_m$ (see Rem. 6.10). ◁

Remark 6.14 [An Implication of Conditional Independence of $Y_1, \ldots, Y_m$] If $(m - n) \geq 2$, then

$$Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_m | X \quad \Rightarrow \quad Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_{m-n} | X. \tag{6.18}$$

For example, *X-conditional independence* of Y_1, Y_2, Y_3 implies their *pairwise X-conditional independence*. More precisely,

$$Y_1 \perp\!\!\!\perp Y_2 \perp\!\!\!\perp Y_3 | X \quad \Rightarrow \quad Y_1 \perp\!\!\!\perp Y_2 | X \ \wedge\ Y_1 \perp\!\!\!\perp Y_3 | X \ \wedge\ Y_2 \perp\!\!\!\perp Y_3 | X, \tag{6.19}$$

but not vice versa (see Exercise 6-3). ◁

6.2 Independence of Events and of Random Variables Revisited

Independence of events has already been treated in section 1.3 and independence of random variables in section 2.4. Now we show that independence of events and of random variables are a special case of conditional independence given a random variable and add some useful properties.

6.2.1 Independence of Events

Now we consider

$$X \underset{P}{=} x, \quad \text{where } x \in \Omega'_X \tag{6.20}$$

and Ω'_X denotes the co-domain of X. Hence, Equation (6.20) means that *X is a constant, P-almost surely* (see Def. 2.35). In this special case,

$$P(A) = E(1_A) \in \mathscr{E}(1_A \mid X), \tag{6.21}$$

that is, if $X \underset{P}{=} x$, then the constant $P(A)$ is a version of the X-conditional probability $P(A \mid X)$ [see Eq. (4.9), Box 4.1 (vi), and Eq. (3.8)]. Of course, the corresponding argument holds for the events B and $A \cap B$.

Remark 6.15 [Independence of Two Events] Let the assumptions of Definition 6.1 hold. Then, for X being a constant, P-almost surely, Equation (6.2) is equivalent to

$$P(A \cap B) = P(A) \cdot P(B). \tag{6.22}$$

This equation defines $A \perp\!\!\!\perp B$, that is, *independence of the events A and B* (see Def. 1.39). Under the assumptions of Definition 6.1, independence of two events A and B, that is, $A \perp\!\!\!\perp B$, is a special case of $A \perp\!\!\!\perp B \mid X$ for X being a constant, P-almost surely, that is,

$$X \underset{P}{=} x \quad \Rightarrow \quad \big(A \perp\!\!\!\perp B \mid X \;\Leftrightarrow\; A \perp\!\!\!\perp B\big). \tag{6.23}$$

(see Exercise 6-4). ◁

6.2.2 Independence of Random Variables

Similarly, under the assumptions of Definition 6.2, if Equation (6.20) holds, then Equation (6.3) and Proposition (6.21) yield

$$P(A \cap B) = P(A) \cdot P(B), \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z), \tag{6.24}$$

which defines *independence of the random variables Y and Z* see (Def. 2.48). Hence, $Y \perp\!\!\!\perp Z$ is a special case of $Y \perp\!\!\!\perp Z \mid X$ for $X \underset{P}{=} x$, that is, for X being a constant, P-almost surely.

For $X \underset{P}{=} x$, Theorem 6.4 yields the following corollary.

Corollary 6.16 [An Implication of Independence]
Let Y and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let Ω'_Y denote the co-domain of Y, let $y \in \Omega'_Y$, and $\{Y = y\} \in \mathscr{A}$. Then

$$Y \perp\!\!\!\perp Z \quad \Rightarrow \quad P(Y=y \mid Z) \underset{P}{=} P(Y=y). \tag{6.25}$$

According to Proposition (6.25), $Y \perp\!\!\!\perp Z$ implies that the Z-conditional probability of the event $\{Y=y\}$ that Y takes on the value y does not depend on Z, P-almost surely. In other words, under $Y \perp\!\!\!\perp Z$, the probability $P(Y=y)$ is a version of $P(Y=y \mid Z)$.

In the following corollary we consider the case in which the image $Y(\Omega)$ of Y is finite or countable. This corollary is a special case of Theorem 6.5 for X being a constant.

Corollary 6.17 [A Property Equivalent to Independence]
Let Y and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$, assume that $Y(\Omega)$ is finite or countable, and, for all $y \in Y(\Omega)$, let $\{Y=y\} \in \mathscr{A}$. Then

$$Y \perp\!\!\!\perp Z \quad \Leftrightarrow \quad \forall\, y \in Y(\Omega)\colon\; P(Y=y\,|\,Z) \underset{P}{=} P(Y=y) \tag{6.26}$$

$$\Leftrightarrow \quad \forall\, y \in Y(\Omega)\colon\; 1_{Y=y} \perp\!\!\!\perp Z. \tag{6.27}$$

Hence, according to Proposition (6.26), if $Y(\Omega)$ is finite or countable, then independence of Y and Z (with respect to the measure P) is equivalent to $P(Y=y\,|\,Z) \underset{P}{=} P(Y=y)$ for all $y \in Y(\Omega)$. Furthermore, according to Proposition (6.27), $Y \perp\!\!\!\perp Z$ is equivalent to independence of Z and all indicators $1_{Y=y}$, for which $y \in Y(\Omega)$.

In the following corollary we consider a single value y of Y and the indicator $1_{Y=y}$, that is, we consider the special case $1_{Y=y} \perp\!\!\!\perp Z$. This corollary is a special case of Corollary 6.6 for X being a constant.

Corollary 6.18 [Independence Involving an Indicator]
Let Y and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let $y \in Y(\Omega)$ and $\{Y=y\} \in \mathscr{A}$. Then the following propositions are equivalent to each other:

(a) $1_{Y=y} \perp\!\!\!\perp Z$
(b) $P(1_{Y=y}=1\,|\,Z) \underset{P}{=} P(1_{Y=y}=1)$
(c) $P(1_{Y=y}=0\,|\,Z) \underset{P}{=} P(1_{Y=y}=0)$
(d) $P(Y=y\,|\,Z) \underset{P}{=} P(Y=y)$
(e) $P(Y\neq y\,|\,Z) \underset{P}{=} P(Y\neq y)$.

According to this corollary, in order to prove that Z and the indicator $1_{Y=y}$ are independent, we only have to show that the Z-conditional probability of the event $\{Y=y\}$ does not depend on Z, P-almost surely [see Prop. (d)].

6.3 Conditional Independence With Respect to $P^{X=x}$

Now we turn to conditional independence with respect to the measure $P^{X=x}$. This has three purposes. First, we provide equations and other propositions involving this measure. Second, and more important, we treat some propositions that involve the original measure P *and* the measure $P^{X=x}$. Third, we treat implications of $Y \perp\!\!\!\perp Z\,|\,X$ for $Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z$, that is, for independence of Y and Z with respect to the measure $P^{X=x}$.

Remark 6.19 [The Measure $P^{X=x}$] Remember, $P(X=x)$ and $P(A\,|\,X=x)$ are just a more convenient notation for $P(\{X=x\})$ and $P(A\,|\,\{X=x\})$, respectively, where we assume that $\{X=x\} = \{\omega \in \Omega\colon X(\omega) = x\} \in \mathscr{A}$. Hence, $\{X=x\}$ denotes the event that X takes on the value $x \in X(\Omega)$. Also remember, in Equation (1.10), we already defined the $(X=x)$-*conditional probability of the event* A by

$$P(A \mid X{=}x) \;=\; \frac{P(A \cap \{X{=}x\})}{P(X{=}x)}. \tag{6.28}$$

Now, if X is a random variable on the probability space $(\Omega, \mathscr{A}, P)$, $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, and $P(X{=}x) > 0$, then we define $P^{X=x}$ by

$$P^{X=x}(A) \;=\; P(A \mid X{=}x), \quad \forall\, A \in \mathscr{A}. \tag{6.29}$$

Remember, $P^{X=x}$ is a probability measure on the measurable space $(\Omega, \mathscr{A})$, and it satisfies

$$P^{X=x}(A) \;=\; 0, \quad \forall\, A \in \mathscr{A} \text{ with } A \cap \{X{=}x\} = \emptyset \tag{6.30}$$

(see Exercise 6-5). ◁

6.3.1 Conditional Independence of Events With Respect to $P^{X=x}$

We begin with the concept of *W-conditional independence of two events with respect to* $P^{X=x}$, or more briefly, *W-conditional* $P^{X=x}$*-independence of two events*. Except for using $P^{X=x}$ instead of P and exchanging Z by W, the definition is analog to Definition 6.2. Nevertheless, we repeat it for convenience.

Definition 6.20 [W**-Conditional** $P^{X=x}$**-Independence of Two Events**]
Let X and W be random variables on the probability space $(\Omega, \mathscr{A}, P)$ and let A, $B \in \mathscr{A}$. Furthermore, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, and $P^{X=x}$ be the probability measure on $(\Omega, \mathscr{A})$ defined by Equation (6.29). Then W-conditional independence of A and B with respect to $P^{X=x}$, denoted $A \underset{P^{X=x}}{\perp\!\!\!\perp} B \mid W$, is defined by

$$P^{X=x}(A \cap B \mid W) \underset{P^{X=x}}{=} P^{X=x}(A \mid W) \cdot P^{X=x}(B \mid W). \tag{6.31}$$

In the following theorem we show that (X, W)-conditional independence of two events A and B implies their W-conditional independence with respect to the probability measure $P^{X=x}$, provided that this measure is defined by Equation (6.29).

Theorem 6.21 [**An Implication of** $A \perp\!\!\!\perp B \mid (X, W)$]
Let X and W be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Furthermore, let $A, B \in \mathscr{A}$, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, assume $P(X{=}x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Then

$$A \perp\!\!\!\perp B \mid (X, W) \quad \Rightarrow \quad A \underset{P^{X=x}}{\perp\!\!\!\perp} B \mid W. \tag{6.32}$$

(Proof p. 143)

Remark 6.22 [**A Special Case**] Under the assumptions of Theorem 6.21, for W being a constant, Proposition (6.32) yields

$$A \perp\!\!\!\perp B \mid X \quad \Rightarrow \quad A \underset{P^{X=x}}{\perp\!\!\!\perp} B. \tag{6.33}$$

Hence, under these assumptions, X-conditional independence of the events A and B implies that they are independent with respect to the probability measure $P^{X=x}$ defined by Equation (6.29) (see Exercise 6-6). $\triangleleft$

6.3.2 Conditional Independence of Two Random Variables With Respect to $P^{X=x}$

Now we turn to *conditional independence with respect to* $P^{X=x}$ *of two random variables given a random variable*, or more briefly, *conditional* $P^{X=x}$*-independence of two random variables given a random variable.*

Definition 6.23 [W-Conditional $P^{X=x}$-Independence of Two Random Variables]
Let W, X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let $x \in X(\Omega)$, $\{X=x\} \in \mathscr{A}$, assume $P(X=x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Then W-conditional independence of Y and Z with respect to $P^{X=x}$, denoted $Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \,|\, W$, is defined by

$$P^{X=x}(A \cap B \,|\, W) \underset{P^{X=x}}{=} P^{X=x}(A \,|\, W) \cdot P^{X=x}(B \,|\, W), \quad \forall\, (A, B) \in \sigma(Y) \times \sigma(Z). \tag{6.34}$$

Hence,

$$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \,|\, W \quad \Leftrightarrow \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z):\; A \underset{P^{X=x}}{\perp\!\!\!\perp} B \,|\, W. \tag{6.35}$$

Analogously to what has been shown in Theorem 6.21 for events, now we show that (X, W)-conditional independence of two random variables Y and Z implies their W-conditional independence with respect to the probability measure $P^{X=x}$.

Theorem 6.24 [An Implication of $Y \perp\!\!\!\perp Z \,|\, (X, W)$]
Under the assumptions of Definition 6.23,

$$Y \perp\!\!\!\perp Z \,|\, (X, W) \quad \Rightarrow \quad Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \,|\, W. \tag{6.36}$$

(Proof p. 144)

Remark 6.25 [An Implication of Conditional Independence Involving $P^{X=x}$] Under the assumptions of Definition 6.23, for W being a constant, Proposition (6.36) yields

$$Y \perp\!\!\!\perp Z \,|\, X \quad \Rightarrow \quad Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z. \tag{6.37}$$

Hence, under these assumptions, X-conditional independence of the random variables Y and Z with respect to P implies that they are independent with respect to the probability measure $P^{X=x}$. $\triangleleft$

6.3.3 Implications on Conditional Mean-Independence With Respect to $P^{X=x}$

In the next theorem we present an implication of $X \perp\!\!\!\perp Y \,|\, Z$ for the conditional expectation $E^{X=x}(Y \,|\, Z)$, which been introduced in section 5.1. (For a proof see SN-Theorem 16.38).

Theorem 6.26 [An Implication of Z-Conditional Independence]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, $P(X{=}x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Furthermore, assume that Y is numerical and nonnegative or such that the expectation $E(Y)$ of Y is finite. Then:

$$X \perp\!\!\!\perp Y \mid Z \quad \Rightarrow \quad \mathscr{E}(Y \mid Z) \subset \mathscr{E}^{X=x}(Y \mid Z) \tag{6.38}$$

$$\Rightarrow \quad E(Y \mid Z) \underset{P^{X=x}}{=} E^{X=x}(Y \mid Z). \tag{6.39}$$

According to Proposition (6.38), $X \perp\!\!\!\perp Y \mid Z$ implies $E(Y \mid Z) \in \mathscr{E}^{X=x}(Y \mid Z)$. It does not imply $E^{X=x}(Y \mid Z) \in \mathscr{E}(Y \mid Z)$.

Remark 6.27 [An Additional Implication if $E^{X=x}(Y \mid Z)$ is P-Unique] If, additionally to the assumptions of Theorem 6.26, we assume that $E^{X=x}(Y \mid Z)$ is P-unique, then

$$X \perp\!\!\!\perp Y \mid Z \quad \Rightarrow \quad \mathscr{E}(Y \mid Z) = \mathscr{E}^{X=x}(Y \mid Z) \tag{6.40}$$

Hence, under the assumptions of Theorem 6.26,

$$X \perp\!\!\!\perp Y \mid Z \wedge E^{X=x}(Y \mid Z) \text{ is } P\text{-unique} \quad \Rightarrow \quad E(Y \mid Z) \underset{P}{=} E^{X=x}(Y \mid Z) \tag{6.41}$$

[see Prop. (4.6)]. ◁

6.4 Independence With Respect to $P^{X=x}$

Independence of events (with respect to P) has already been treated in section 1.3 and independence of random variables (with respect to P) in section 2.4. Now we turn to independence of events and independence of random variables with respect to the measure $P^{X=x}$, show that X-conditional independence of two events and of two random variables with respect to P implies their independence with respect to $P^{X=x}$, and add some useful properties.

6.4.1 Independence of Two Events With Respect to $P^{X=x}$

In the following definition we translate independence of two events A and B to their independence with respect to the measure $P^{X=x}$ defined by Equation (6.29). This will allow us to formulate some propositions involving both measures, P and $P^{X=x}$.

Definition 6.28 [Independence of Two Events With Respect to $P^{X=x}$]
Let X be a random variable on the probability space $(\Omega, \mathscr{A}, P)$ and let $A, B \in \mathscr{A}$. Furthermore, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, $P(X{=}x) > 0$, and let $P^{X=x}$ be the probability measure defined by Equation (6.29). Then independence of A and B with respect to $P^{X=x}$, denoted $A \underset{P^{X=x}}{\perp\!\!\!\perp} B$, is defined by

$$P^{X=x}(A \cap B) = P^{X=x}(A) \cdot P^{X=x}(B). \tag{6.42}$$

Hence,

$$A \underset{P^{X=x}}{\perp\!\!\!\perp} B \quad \Leftrightarrow \quad P^{X=x}(A \cap B) = P^{X=x}(A) \cdot P^{X=x}(B). \tag{6.43}$$

Comparing this definition to Definition 1.39 shows that we can apply all properties of independence of two events with respect to P (see, e. g., Box 1.2) to their independence with respect to $P^{X=x}$, provided that we replace the measure P by the measure $P^{X=x}$.

Remark 6.29 [A Special Case] Note that Equation (6.42) is a special case of Equation (6.31) for W being a constant (see again Exercise 6-6). ◁

6.4.2 Independence of Two Random Variables With Respect to $P^{X=x}$

Now we translate independence of two random variables with respect to P to their independence with respect to $P^{X=x}$. Again, this will allow us to formulate some propositions involving both measures, P and $P^{X=x}$.

Definition 6.30 [Independence of Two Random Variables With Respect to $P^{X=x}$]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Furthermore, let $x \in X(\Omega)$, $\{X=x\} \in \mathscr{A}$, $P(X=x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Then independence of Y and Z with respect to $P^{X=x}$, denoted $Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z$, is defined by

$$P^{X=x}(A \cap B) = P^{X=x}(A) \cdot P^{X=x}(B), \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z). \tag{6.44}$$

Hence,

$$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \quad \Leftrightarrow \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z)\colon\ A \underset{P^{X=x}}{\perp\!\!\!\perp} B. \tag{6.45}$$

Comparing this definition to Definition 2.48 shows that we can apply all properties of independence of two random variables (with respect to P) (see, e. g., Box 2.1) to their independence with respect to $P^{X=x}$, provided that we replace P by the measure $P^{X=x}$.

Remark 6.31 [An Implication of X-Conditional Independence of Y and Z] Under the assumptions of Definition 6.30,

$$Y \perp\!\!\!\perp Z \mid X \quad \Rightarrow \quad Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z. \tag{6.46}$$

Hence, X-conditional independence of two random variables implies their independence with respect to $P^{X=x}$ (see Exercise 6-7). ◁

The following corollary follows from Theorem 6.26 for Z being a constant.

Corollary 6.32 [An Implication of Independence]
Let X and Y be random variables on $(\Omega, \mathscr{A}, P)$, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, $P(X{=}x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Furthermore, assume that Y is numerical and nonnegative or such that $E(Y)$ is finite. Then

$$X \perp\!\!\!\perp Y \quad \Rightarrow \quad E(Y) \;=\; E^{X=x}(Y). \tag{6.47}$$

Hence, if Y is numerical and nonnegative or such that $E(Y)$ is finite, then $X \perp\!\!\!\perp Y$ does not only imply $E(Y|X) \underset{P}{=} E(Y)$ [see Box 4.1 (v)], but also $E^{X=x}(Y) = E(Y)$, provided that the assumptions of Corollary 6.32 hold.

Now we translate Corollary 6.16, replacing the measure P by $P^{X=x}$.

Corollary 6.33 [An Implication of Conditional Independence]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Furthermore, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, $P(X{=}x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Finally, let Ω'_Y denote the co-domain of Y, let $y \in \Omega'_Y$, and $\{Y{=}y\} \in \mathscr{A}$. Then

$$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \quad \Rightarrow \quad P^{X=x}(Y{=}y|Z) \underset{P^{X=x}}{=} P^{X=x}(Y{=}y). \tag{6.48}$$

According to Proposition (6.48), independence of Y and Z with respect to $P^{X=x}$ implies that the Z-conditional probability $P^{X=x}(Y{=}y|Z)$ does not depend on Z. Instead, it is a constant, $P^{X=x}$-almost surely.

In the following corollary we consider the case in which the image $Y(\Omega)$ of Y is finite or countable. Replacing the measure P by $P^{X=x}$ in Theorem 6.17 yields:

Corollary 6.34 [A Property Equivalent to Independence With Respect to $P^{X=x}$]
Let X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Furthermore, let $x \in X(\Omega)$, $\{X{=}x\} \in \mathscr{A}$, $P(X{=}x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Finally, let $Y(\Omega)$ be finite or countable and, for all $y \in Y(\Omega)$, let $\{Y{=}y\} \in \mathscr{A}$. Then

$$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \quad \Leftrightarrow \quad \forall\, y \in Y(\Omega)\colon\; P^{X=x}(Y{=}y|Z) \underset{P^{X=x}}{=} P^{X=x}(Y{=}y) \tag{6.49}$$

$$\Leftrightarrow \quad \forall\, y \in Y(\Omega)\colon\; 1_{Y=y} \underset{P^{X=x}}{\perp\!\!\!\perp} Z. \tag{6.50}$$

Hence, according to Proposition (6.49), if $Y(\Omega)$ is finite or countable, then independence of Y and Z with respect to the $P^{X=x}$ is equivalent to $P^{X=x}(Y{=}y|Z) \underset{P}{=} P^{X=x}(Y{=}y)$ for all $y \in Y(\Omega)$. Furthermore, according to Proposition (6.50), $Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z$ is equivalent to $P^{X=x}$-independence of Z and all indicators $1_{Y=y}$, $y \in Y(\Omega)$.

In the following corollary we consider a single value y of Y and the indicator $1_{Y=y}$, that is, we consider the special case $1_{Y=y} \underset{P^{X=x}}{\perp\!\!\!\perp} Z$.

Corollary 6.35 [Independence With Respect to $P^{X=x}$ Involving an Indicator]
Let X, Y and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$. Furthermore, let $x \in X(\Omega)$, $\{X=x\} \in \mathscr{A}$, $P(X=x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29). Finally, let $y \in Y(\Omega)$ and $\{Y=y\} \in \mathscr{A}$. Then the following propositions are equivalent to each other:

(a) $1_{Y=y} \underset{P^{X=x}}{\perp\!\!\!\perp} Z$

(b) $P^{X=x}(1_{Y=y}=1 \,|\, Z) \underset{P^{X=x}}{=} P^{X=x}(1_{Y=y}=1)$

(c) $P^{X=x}(1_{Y=y}=0 \,|\, Z) \underset{P^{X=x}}{=} P^{X=x}(1_{Y=y}=0)$

(d) $P^{X=x}(Y=y \,|\, Z) \underset{P^{X=x}}{=} P^{X=x}(Y=y)$

(e) $P^{X=x}(Y \neq y \,|\, Z) \underset{P^{X=x}}{=} P^{X=x}(Y \neq y)$.

According to this corollary, in order to prove that Z and the indicator $1_{Y=y}$ are independent, we only have to show that the Z-conditional probability with respect to $P^{X=x}$ of the event $\{Y=y\}$ does not depend on Z [see Prop. (d)].

Example 6.36 [Nonorthogonal Factors] Consider the random experiment presented in Table 1.3. In this random experiment, $X \underset{P^{Z=med}}{\perp\!\!\!\perp} U$ holds , because

$$\forall\ x \in X(\Omega) = \{0,1,2\}\colon\ \ P^{Z=med}(X=x \,|\, U) \underset{P^{Z=med}}{=} P^{Z=med}(X=x). \tag{6.51}$$

Inspecting the conditional probabilities $P(X=1 \,|\, U=u)$ and $P(X=2 \,|\, U=u)$ in rows three to six of Table 1.3 [and with them $P(X=0|U=u) = 1-P(X=1|U=u)-P(X=2 \,|\, U=u)$] shows

$\forall\ x \in X(\Omega) = \{0,1,2\}$:

$$
\begin{aligned}
P^{Z=med}(X=x \,|\, U)(\omega) \ &=\ P^{Z=med}(X=x \,|\, U=u), \quad \text{if } \omega \in \{Z=med, U=u\} && [(5.25)]\\
&= \frac{P(X=x, U=u, Z=med)}{P(U=u, Z=med)}, \quad \text{if } \omega \in \{Z=med, U=u\} && [(1.10), (1.26)]\\
&= \begin{cases} \dfrac{P(X=x, Z=med)}{P(Z=med)}, & \text{if } \omega \in \{Z=med\}\\ 0, \text{ otherwise} \end{cases}\\
&= \begin{cases} P^{Z=med}(X=x), & \text{if } \omega \in \{Z=med\}\\ 0, \text{ otherwise.} \end{cases} && [(1.10), (1.26)]
\end{aligned}
$$

Because $P^{Z=med}(A) = 0$, for $A = \big\{\omega \in \Omega\colon \omega \notin \{Z=med\}\big\}$ (see Def. 2.35), this implies Proposition (6.51), which, in this example, is equivalent to $X \underset{P^{Z=med}}{\perp\!\!\!\perp} U$. Note that, in this example, neither $X \underset{P^{Z=low}}{\perp\!\!\!\perp} U$ nor $X \underset{P^{Z=hi}}{\perp\!\!\!\perp} U$ hold. ◁

Box 6.2 Glossary of new concepts

Let W, X, Y, and Z be random variables on the probability space $(\Omega, \mathscr{A}, P)$ and let $A, B \in \mathscr{A}$. Furthermore, let $x \in X(\Omega)$, $\{X = x\} \in \mathscr{A}$, $P(X = x) > 0$, and let $P^{X=x}$ be defined by Equation (6.29).

$A \perp\!\!\!\perp B \,|\, X$ — *X-conditional independence of the events A and B* with respect to P. It is defined by

$$P(A \cap B \,|\, X) \underset{P}{=} P(A \,|\, X) \cdot P(B \,|\, X).$$

$Y \perp\!\!\!\perp Z \,|\, X$ — *X-conditional independence of the random variables Y and Z* on $(\Omega, \mathscr{A}, P)$ with respect to P. It is defined by

$$P(A \cap B \,|\, X) \underset{P}{=} P(A \,|\, X) \cdot P(B \,|\, X), \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z).$$

$A \underset{P^{X=x}}{\perp\!\!\!\perp} B \,|\, W$ — *W-conditional independence of the events A and B* with respect to $P^{X=x}$. It is defined by

$$P^{X=x}(A \cap B \,|\, W) \underset{P^{X=x}}{=} P^{X=x}(A \,|\, W) \cdot P^{X=x}(B \,|\, W).$$

$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \,|\, W$ — *W-conditional independence of the random variables Y and Z* on $(\Omega, \mathscr{A}, P)$ with respect to $P^{X=x}$. It is defined by

$$P^{X=x}(A \cap B \,|\, W) \underset{P^{X=x}}{=} P^{X=x}(A \,|\, W) \cdot P^{X=x}(B \,|\, W), \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z).$$

$A \underset{P^{X=x}}{\perp\!\!\!\perp} B$ — *Independence of the events A and B with respect to* $P^{X=x}$. It is defined by

$$P^{X=x}(A \cap B) = P^{X=x}(A) \cdot P^{X=x}(B).$$

$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z$ — *Independence of the random variables Y and Z with respect to* $P^{X=x}$. It is defined by

$$P^{X=x}(A \cap B) = P^{X=x}(A) \cdot P^{X=x}(B), \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z).$$

6.5 Summary and Conclusions

In this chapter we reviewed the concepts of *conditional independence of events given a random variable* and *conditional independence of random variables given a random variable*. We also revisited (unconditional) independence between events and independence between random variables, and showed that these concepts are special cases of conditional independence given a random variable. Furthermore, we added some propositions about independence that refer to the concepts of a conditional probability of an event or of a conditional expectation of a numerical random variable. Finally, we treated conditional independence with respect to a probability measure $P^{X=x}$, focussing on properties that involve both measures, P and $P^{X=x}$. Box 6.2 summarizes the most important concepts treated in this chapter. More details about conditional independence are found in SN-chapter 16.

6.6 Proofs

Proof of Theorem 6.6

(a) ⇒ (b), (c). These implications immediately follow from Theorem 6.5.

(b) ⇒ (a).

$$
\begin{aligned}
& P(1_{Y=y}{=}1 \mid X, Z) \underset{P}{=} P(1_{Y=y}{=}1 \mid X) \\
\Leftrightarrow\; & 1 - P(1_{Y=y}{=}1 \mid X, Z) \underset{P}{=} 1 - P(1_{Y=y}{=}1 \mid X) && [\,| \cdot (-1),\ |+1\,] \\
\Leftrightarrow\; & P(1_{Y=y}{=}0 \mid X, Z) \underset{P}{=} P(1_{Y=y}{=}0 \mid X) && [(4.11),\ P(1_{Y=y}{=}0 \mid X) \underset{P}{=} P(1_{Y=y}{\neq}1 \mid X)] \\
\Rightarrow\; & P(1_{Y=y}{=}0 \mid X, Z) \underset{P}{=} P(1_{Y=y}{=}0 \mid X), && [\text{Th. } 6.5]
\end{aligned}
$$

where $|\cdot(-1)$ means multiplying both sides of the equation by (-1) and $|+1$ means adding 1 on both sides.

(b) ⇔ (c). This equivalence has just been shown using $P(1_{Y=y}{=}0 \mid X) \underset{P}{=} P(1_{Y=y}{\neq}1 \mid X)$ and Equation (4.11).

(d) ⇔ (a), (e) ⇔ (c). These equivalences hold because

$$\{Y{=}y\} \;=\; \{1_{Y=y}{=}1\} \;=\; \{\omega \in \Omega\colon Y(\omega) = y\}$$

and

$$\{Y{=}y\}^c \;=\; \{Y{\neq}y\} \;=\; \{1_{Y=y}{=}0\} \;=\; \{\omega \in \Omega\colon Y(\omega) \neq y\}$$

(see Rem. 4.10).

Proof of Theorem 6.21

$$
\begin{aligned}
& A \perp\!\!\!\perp B \mid (X, W) \\
\Leftrightarrow\; & P(A \cap B \mid X, W) \underset{P}{=} P(A \mid X, W) \cdot P(B \mid X, W) && [(6.2)] \\
\Leftrightarrow\; & E(1_{A\cap B} \mid X, W) \underset{P}{=} E(1_A \mid X, W) \cdot E(1_B \mid X, W) && [(4.9)] \\
\Rightarrow\; & E^{X=x}\big(E(1_{A\cap B} \mid X, W) \,\big|\, W\big) \underset{P^{X=x}}{=} E^{X=x}\big(E(1_A \mid X, W) \cdot E(1_B \mid X, W) \,\big|\, W\big) \\
& && [\text{Box } 5.1 \text{ (v), Box } 4.1 \text{ (xiii)}] \\
\Rightarrow\; & E^{X=x}(1_{A\cap B} \mid W) \underset{P^{X=x}}{=} E^{X=x}\big(E(1_A \mid X, W) \cdot E(1_B \mid X, W) \,\big|\, W\big) \\
& && [\text{Box } 5.1 \text{ (v), Box } 4.1 \text{ (xii)}] \\
\Rightarrow\; & E^{X=x}(1_{A\cap B} \mid W) \underset{P^{X=x}}{=} E^{X=x}\big(E^{X=x}(1_A \mid W) \cdot E^{X=x}(1_B \mid W) \,\big|\, W\big) && [(5.35)] \\
\Rightarrow\; & E^{X=x}(1_{A\cap B} \mid W) \underset{P^{X=x}}{=} E^{X=x}(1_A \mid W) \cdot E^{X=x}(1_B \mid W) && [\text{SN-Th. } 2.57, \text{ Box } 4.1 \text{ (xi)}] \\
\Rightarrow\; & P^{X=x}(A \cap B \mid W) \underset{P^{X=x}}{=} P^{X=x}(A \mid W) \cdot P^{X=x}(B \mid W) && [(4.9)] \\
\Leftrightarrow\; & A \underset{P^{X=x}}{\perp\!\!\!\perp} B \mid W. && [\text{Def. } 6.1]
\end{aligned}
$$

Proof of Theorem 6.24

$$
\begin{aligned}
& Y \perp\!\!\!\perp Z \,|\, (X, W) \\
\Leftrightarrow\ & \forall (A,B) \in \sigma(Y) \times \sigma(Z):\ P(A \cap B \,|\, X, W) \underset{P}{=} P(A \,|\, X, W) \cdot P(B \,|\, X, W) && \text{[Def. 6.2]} \\
\Leftrightarrow\ & \forall (A,B) \in \sigma(Y) \times \sigma(Z):\ A \perp\!\!\!\perp B \,|\, (X, W) && \text{[Def. 6.1]} \\
\Rightarrow\ & \forall (A,B) \in \sigma(Y) \times \sigma(Z):\ A \underset{P^{X=x}}{\perp\!\!\!\perp} B \,|\, W && \text{[Prop. (6.32)]} \\
\Leftrightarrow\ & Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \,|\, W. && \text{[Def. 6.1]}
\end{aligned}
$$

6.7 Exercises

▷ **Exercise 6-1** Prove Proposition (6.9), using Proposition (6.7).

▷ **Exercise 6-2** Compute the conditional probabilities $P(Y=1|Z=86)$ and $P(Y=1|W=m, Z=86)$ shown in the first row of Table 6.1 from the probabilities $P(\{\omega_i\})$ displayed in the second column of this table.

▷ **Exercise 6-3** Show that $Y_1 \perp\!\!\!\perp Y_2 \perp\!\!\!\perp Y_3 | X$ implies $Y_1 \perp\!\!\!\perp Y_2 | X$.

▷ **Exercise 6-4** Prove Proposition (6.23).

▷ **Exercise 6-5** Show that the probability measure $P^{X=x}$ on the measurable space $(\Omega, \mathscr{A})$ defined by Equation (6.29) satisfies Equation (6.30).

▷ **Exercise 6-6** Show: Under the assumptions of Theorem 6.21, for W being a constant, P-almost surely, Proposition (6.33) is a special case of Proposition (6.32).

▷ **Exercise 6-7** Prove Proposition (6.46).

Solutions

▷ **Solution 6-1** According to Proposition (6.7), for $y \in Y(\Omega)$,

$$
X \perp\!\!\!\perp 1_{Y=y} | Z \quad \Leftrightarrow \quad P(Y=y|X,Z) \underset{P}{=} P(Y=y|Z) \ \wedge\ P(Y\neq y|X,Z) \underset{P}{=} P(Y\neq y|Z).
$$

Hence, $X \perp\!\!\!\perp 1_{Y=y} | Z$ implies $P(Y=y|X,Z) \underset{P}{=} P(Y=y|Z)$. Vice versa, $P(Y=y|X,Z) \underset{P}{=} P(Y=y|Z)$ implies $X \perp\!\!\!\perp 1_{Y=y} | Z$, because

$$
\begin{aligned}
& P(Y=y|X,Z) \underset{P}{=} P(Y=y|Z) \\
\Leftrightarrow\ & 1 - P(Y=y|X,Z) \underset{P}{=} 1 - P(Y=y|Z) \\
\Leftrightarrow\ & P(Y\neq y|X,Z) \underset{P}{=} P(Y\neq y|Z). && \text{[(4.11)]}
\end{aligned}
$$

Hence, if $P(Y=y|X,Z) \underset{P}{=} P(Y=y|Z)$, then $P(Y\neq y|X,Z) \underset{P}{=} P(Y\neq y|Z)$. However, according to Proposition (6.7) this implies $X \perp\!\!\!\perp 1_{Y=y} | Z$.

▷ **Solution 6-2**

$$P(Y{=}1|Z{=}86) = \frac{P(Y{=}1, Z{=}86)}{P(Z{=}86)} = \frac{.01+.01}{.24+.01+.24+.01} = \frac{.02}{.5} = .04.$$

$$P(Y{=}1|W{=}m, Z{=}86) = \frac{P(Y{=}1, W{=}m, Z{=}86)}{P(W{=}m, Z{=}86)} = \frac{.01}{.24+.01} = \frac{.01}{.25} = .04.$$

▷ **Solution 6-3** According to Proposition (6.17), $Y_1 \perp\!\!\!\perp Y_2 \perp\!\!\!\perp Y_3 | X$ is defined by

$$\forall (A_1, A_2, A_3) \in \sigma(Y_1) \times \sigma(Y_2) \times \sigma(Y_3): \quad P(A_1 \cap A_2 \cap A_3 | X) \underset{P}{=} P(A_1 | X) \cdot P(A_2 | X) \cdot P(A_3 | X).$$

Because $\Omega \in \sigma(Y_3)$, this implies

$$\forall (A_1, A_2) \in \sigma(Y_1) \times \sigma(Y_2): \quad P(A_1 \cap A_2 \cap \Omega | X) \underset{P}{=} P(A_1 | X) \cdot P(A_2 | X) \cdot P(\Omega | X).$$

Finally, because $A_1 \cap A_2 \cap \Omega = A_1 \cap A_2$ and $P(\Omega | X) \underset{P}{=} 1$, this implies

$$\forall (A_1, A_2) \in \sigma(Y_1) \times \sigma(Y_2): \quad P(A_1 \cap A_2 | X) \underset{P}{=} P(A_1 | X) \cdot P(A_2 | X),$$

and this is the condition that defines $Y_1 \perp\!\!\!\perp Y_2 | X$.

▷ **Solution 6-4** If $X \underset{P}{=} x$, that is, if X is identical to the constant x, P-almost surely, then

$$\begin{aligned}
& A \perp\!\!\!\perp B | X \\
\Leftrightarrow\ & P(A \cap B | X) \underset{P}{=} P(A | X) \cdot P(B | X) && \text{[Def. 6.2]} \\
\Leftrightarrow\ & E(1_{A \cap B} | X) \underset{P}{=} E(1_A | X) \cdot E(1_B | X) && \text{[(4.9)]} \\
\Leftrightarrow\ & E(1_{A \cap B}) = E(1_A) \cdot E(1_B) && [X \underset{P}{=} x, \text{ Box 4.1 (vi)}] \\
\Leftrightarrow\ & P(A \cap B) = P(A) \cdot P(B) && \text{[(3.8)]} \\
\Leftrightarrow\ & A \perp\!\!\!\perp B. && \text{[Def. 1.39]}
\end{aligned}$$

▷ **Solution 6-5** If $A \cap \{X{=}x\} = \emptyset$, then

$$\begin{aligned}
P^{X=x}(A) &= \frac{P(A \cap \{X{=}x\})}{P(X{=}x)} && \text{[(5.1)]} \\
&= \frac{P(\emptyset)}{P(X{=}x)} && [A \cap \{X{=}x\} = \emptyset] \\
&= 0. && \text{[Def. 1.4 (a), Box 1.1 (iv)]}
\end{aligned}$$

▷ **Solution 6-6** We show that $W \underset{P}{=} w$ implies

$$A \perp\!\!\!\perp B | (X, W) \Leftrightarrow A \perp\!\!\!\perp B | X \quad \text{and} \quad A \underset{P^{X=x}}{\perp\!\!\!\perp} B | W \Leftrightarrow A \underset{P^{X=x}}{\perp\!\!\!\perp} B.$$

If $W \underset{P}{=} w$, then

$$\begin{aligned}
& A \perp\!\!\!\perp B | (X, W) \\
\Leftrightarrow\ & P(A \cap B | X, W) \underset{P}{=} P(A | X, W) \cdot P(B | X, W) && \text{[Def. 6.1]} \\
\Leftrightarrow\ & P(A \cap B | X) \underset{P}{=} P(A | X) \cdot P(B | X) && [W \underset{P}{=} w, \text{ (4.9), Box 4.1 (vi), (3.8)}] \\
\Leftrightarrow\ & A \perp\!\!\!\perp B | X. && \text{[Def. 6.1]}
\end{aligned}$$

Because a nullset with respect to P (see Def. 2.35) is also a nullset with respect to $P^{X=x}$,

$$W \underset{P}{=} w \quad \Rightarrow \quad W \underset{P^{X=x}}{=} w,$$

[see Eqs. (6.28) and (6.29)]. Hence,

$$
\begin{aligned}
& A \underset{P^{X=x}}{\perp\!\!\!\perp} B \mid W \\
\Leftrightarrow\quad & P^{X=x}(A \cap B \mid W) \underset{P^{X=x}}{=} P^{X=x}(A \mid W) \cdot P^{X=x}(B \mid W) && \text{[Def. 6.20]} \\
\Leftrightarrow\quad & P^{X=x}(A \cap B) \underset{P^{X=x}}{=} P^{X=x}(A) \cdot P^{X=x}(B) && [W \underset{P^{X=x}}{=} w,\ (4.9), \text{Box 4.1 (vi)}, (3.8)] \\
\Leftrightarrow\quad & A \underset{P^{X=x}}{\perp\!\!\!\perp} B. && \text{[Def. 1.39 (i)]}
\end{aligned}
$$

▷ **Solution 6-7**

$$
\begin{aligned}
& Y \perp\!\!\!\perp Z \mid X \\
\Leftrightarrow\quad & A \perp\!\!\!\perp B \mid X, \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z) && [(6.4)] \\
\Rightarrow\quad & A \underset{P^{X=x}}{\perp\!\!\!\perp} B, \quad \forall (A, B) \in \sigma(Y) \times \sigma(Z) && [(6.33)] \\
\Leftrightarrow\quad & Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z. && [(6.45)]
\end{aligned}
$$

References

Klenke, A. (2020). *Probability theory – A comprehensive course* (3rd ed.). Cham, Switzerland: Springer. doi: 10.1007/978-3-030-56402-5 pages 33, 36

Kolmogorov, A. N. (1933/1977). *Grundbegriffe der Wahrscheinlichkeitsrechnung [Foundations of the Theory of Probability]* (Reprinted ed.). Berlin, Germany: Springer. pages 7, 81

Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd ed.; N. Morrison, Trans.). New York, NY: Chelsea. pages 7, 81

Steyer, R., & Nagel, W. (2017). *Probability and conditional expectation. Fundamentals for the Empirical Sciences.* Chichester, England: Wiley. pages V, VI

Author Index

Subject Index

List of Symbols

$\forall$	for all, 7
$\exists$	there exists, 27
$\neg A$	proposition A does not hold, 41
$A \wedge B$	conjunction of the propositions A and B, 45
$A \Rightarrow B$	proposition A implies the proposition B, 5
$A \Leftrightarrow B$	propositions A and B are equivalent to each other, 20
$\mathbb{N}$	set of natural numbers $1, 2, \ldots$, 8
$\mathbb{R}$	the set of real numbers, 30
$\overline{\mathbb{R}}$	the set $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$, 30
$\mathbb{R}^+$	set of all positive real numbers, 116
$A \in \mathscr{A}$	A is an element of the set $\mathscr{A}$, 4
$A \subset B$	A is a subset of B, i.e., if $a \in A$, then $a \in B$, 5
$A \cup B$	union of the sets A and B, 5
$\bigcup_{i=1}^{m} A_i$, $\bigcup_{i=1}^{m} A_i$	union of the sets $A_1, \ldots, A_m$, 5
$A \cap B$	intersection of the sets A and B, 5
$\bigcap_{i=1}^{m} A_i$, $\bigcap_{i=1}^{m} A_i$	intersection of the sets $A_1, \ldots, A_m$, 5
$\bigcap_{i \in I} A_i$, $\bigcap_{i \in I} A_i$	intersection of the sets A_i, $i \in I$, 5
$A \times B$	Cartesian product of the sets A and B, the set of all pairs (a, b) such that $a \in A$ and $b \in B$, 3
$\times_{i=1}^{m} A_i$, $\mathop{\times}_{i=1}^{m} A_i$	Cartesian product of the sets $A_1, \ldots, A_m$, 6
A^c	complement of the set A, 4
$\Omega \setminus A$	set difference, the set of all elements of Ω that are not elements of the set A, 4
$\sum_{i=1}^{n} \alpha_i$, $\sum_{i=1}^{n} \alpha_i$	sum of the real numbers $\alpha_1, \ldots, \alpha_n$, 8
$\sum_{i=1}^{\infty} \alpha_i$, $\sum_{i=1}^{\infty} \alpha_i$	sum of the sequence $\alpha_1, \alpha_2, \ldots$ of real numbers, 8
$\prod_{i=1}^{m} Y_i$, $\prod_{i=1}^{m} Y_i$	the product of numerical random variables $Y_1, \ldots, Y_m$, 59
$E(Y)$	expectation of the random variable Y, 55
$E^B(Y)$	expectation of the random variable Y w.r.t. the probability measure P^B, 60
$E(Y \mid B)$	conditional expectation value of the random variable Y given the event B, 60

$E^{X=x}(Y)$	expectation of Y w.r.t. $P^{X=x}$, 103
$E(Y\|X=x)$	conditional expectation value of Y given the event $\{X=x\}$, 61
$E(Y\|X=x)$	conditional expectation value of Y given the value x of the random variable X, 87
$E(Y\|X)$	a version of the X-conditional expectation of Y w.r.t. the measure P, 82
$\mathscr{E}(Y\|X)$	set of all versions of the X-conditional expectation of Y, 82
$E_Y(g)$	expectation of g w.r.t. the distribution of Y, 58
$E_Y^{X=x}(g)$	expectation of g w.r.t. the $(X=x)$-conditional distribution of Y, 63
$E(Y\|X=x,Z)$	version of the partial $(X=x,Z)$-conditional expectation of Y w.r.t. the measure P, 113
$\mathscr{E}(Y\|X=x,Z)$	set of all versions of the partial $(X=x,Z)$-conditional expectation of Y, 114
$E(Y\|X_1,\ldots,X_m)$	version of the conditional expectation of Y given $X = (X_1,\ldots,X_m)$, 84
$E^{X=x}(Y\|Z)$	version of the Z-conditional expectation of Y w.r.t. $P^{X=x}$, 104
$\mathscr{E}^{X=x}(Y\|Z)$	set of all versions of the Z-conditional expectation of Y w.r.t. the measure $P^{X=x}$, 105
$E^{X=x}(Y\|Z=z)$	$(Z=z)$-conditional expectation value of Y w.r.t. $P^{X=x}$, 112
ϵ	residual w.r.t. the linear quasi-regression of Y on X, 71
ε	residual of Y with respect to $E(Y\|X)$, 90
$\mathscr{A}$	set of possible events, a σ-algebra on Ω, 4
$\bigotimes_{i=1}^{m}\mathscr{A}_i$, $\otimes_{i=1}^{m}\mathscr{A}_i$	product of the σ-algebras $\mathscr{A}_1,\ldots,\mathscr{A}_m$, 6
$\mathscr{B}$	Borel σ-algebra, 6
$\mathscr{B}_m$	Borel σ-algebra on the set $\mathbb{R}^m$, 86
$\mathscr{E}$	set system on Ω, i.e., a set of subsets of Ω, 5
$\mathscr{P}(\Omega)$	power set of Ω, i.e., the set of all subsets of Ω, 6
$Cov(X,Y)$, σ_{XY}	covariance of the random variables X and Y, 67
$Corr(X,Y)$, ρ_{XY}	correlation of the random variables X and Y, 69
F_Y	distribution function of a real-valued random variable Y, 47
f_Y	probability density of a continuous real-valued random variable Y, 49
$g(X)$, $g \circ X$	the composition of the mapping $X: \Omega \to \Omega'$ and the mapping $g: \Omega' \to \Omega''$, 36
1_A	the indicator of the set A. In the context of a probability space $(\Omega,\mathscr{A},P)$, it is a binary random variable on $(\Omega,\mathscr{A},P)$ if $A \in \mathscr{A}$, 31
$1_{X=x}$	the indicator of the set $\{X=x\}$. In the context of a probability space $(\Omega,\mathscr{A},P)$, it is a binary random variable on $(\Omega,\mathscr{A},P)$ if $\{X=x\} \in \mathscr{A}$, 31

$A \perp\!\!\!\perp B,\ A \underset{P}{\perp\!\!\!\perp} B$	events A and B are independent w.r.t. the probability measure P, 20
$A \underset{P^B}{\perp\!\!\!\perp} C$	events A and C are independent w.r.t. the probability measure P^B, 21
$A \underset{P^{X=x}}{\perp\!\!\!\perp} B$	independence of the events A and B w.r.t. $P^{X=x}$, 138
$A_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp A_n,\ A_1 \underset{P}{\perp\!\!\!\perp} \ldots \underset{P}{\perp\!\!\!\perp} A_n$	independence of the events $A_1, \ldots, A_n$ w.r.t. P, 20
$\perp\!\!\!\perp (A_i, i \in I),\ \underset{P}{\perp\!\!\!\perp} (A_i, i \in I)$	family of events A_i that are independent w.r.t. the probability measure P, 20
$\underset{P^B}{\perp\!\!\!\perp} (A_i, i \in I)$	family of events A_i that are independent w.r.t. the probability measure P^B, 23
$A \perp\!\!\!\perp C \,\vert\, B,\ A \underset{P}{\perp\!\!\!\perp} C \,\vert\, B$	the events A and C are B-conditionally independent w.r.t. the probability measure P, 22
$A \perp\!\!\!\perp B \vert X,\ A \underset{P}{\perp\!\!\!\perp} B \vert X$	X-conditional independence of the events A and B w.r.t. P, 127
$A \underset{P^{X=x}}{\perp\!\!\!\perp} B \,\vert\, W$	W-conditional independence of A and B w.r.t. $P^{X=x}$, 136
$\perp\!\!\!\perp (A_i, i \in I) \,\vert\, X,\ \underset{P}{\perp\!\!\!\perp} (A_i, i \in I) \,\vert\, X$	X-conditional independence of a family $(A_i, i \in I)$ of events, 132
$\mathscr{E}_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp \mathscr{E}_n$	the set systems $\mathscr{E}_1, \ldots, \mathscr{E}_n$ are independent w.r.t. the probability measure P, 21
$\perp\!\!\!\perp (\mathscr{E}_i, i \in I)$	$(\mathscr{E}_i, i \in I)$ is a family of independent set systems $\mathscr{E}_i$ w.r.t. the probability measure P, 21
$\mathscr{E}_1 \perp\!\!\!\perp_{P^B} \ldots \perp\!\!\!\perp_{P^B} \mathscr{E}_m$	independence of the set systems $\mathscr{E}_1, \ldots, \mathscr{E}_m$ w.r.t. P^B, 24
$\perp\!\!\!\perp_{P^B} (\mathscr{E}_i, i \in I)$	independence of the family of set systems $\mathscr{E}_i$, $i \in I$, w.r.t. P^B, 24
$Y \perp\!\!\!\perp Z \vert X,\ Y \underset{P}{\perp\!\!\!\perp} Z \vert X$	X-conditional independence of the random variables Y and Z w.r.t. P, 128
$\perp\!\!\!\perp (Y_i, i \in I) \,\vert\, X,\ \underset{P}{\perp\!\!\!\perp} (Y_i, i \in I) \,\vert\, X$	X-conditional independence of a family $(Y_i, i \in I)$ of random variables, 133
$Y_1 \perp\!\!\!\perp \ldots \perp\!\!\!\perp Y_m \vert X,\ Y_1 \underset{P}{\perp\!\!\!\perp} \ldots \underset{P}{\perp\!\!\!\perp} Y_m \vert X$	X-conditional independence of the random variables $Y_1, \ldots, Y_m$ w.r.t. P, 133
$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z$	independence of the random variables Y and Z w.r.t. $P^{X=x}$, 139
$Y \underset{P^{X=x}}{\perp\!\!\!\perp} Z \,\vert\, W$	W-conditional independence of the random variables Y and Z w.r.t. $P^{X=x}$, 137
$Y \vDash X$	mean-independence of Y from X, 91
$Y \underset{P^{X=x}}{\vDash} Z$	mean-independence of Y from Z w.r.t. $P^{X=x}$, 116
$Y \vDash X \,\vert\, Z$	conditional mean-independence of Y from X given Z, 93
$Y \underset{P^{X=x}}{\vDash} Z \,\vert\, W$	W-conditional mean-independence of Y from Z w.r.t. $P^{X=x}$, 118
$\lambda\big((a,b]\big)$	value of the Lebesgue measure for the interval $(a,b]$, i.e., the length of the interval $(a,b]$, 48

Ω	set of possible outcomes, sample space, 2
$(\Omega, \mathscr{A})$	measurable space consisting of the set Ω and the σ-algebra $\mathscr{A}$ on Ω, 4
$(\Omega, \mathscr{A}, P)$	probability space, 7
P	probability measure, 7
$P(A)$	probability of the event A, 7
$P(X{=}x)$	probability of the event $\{X{=}x\}$ that X takes on the value x, 44
$P(Y \leq y)$	probability of the event $\{Y \leq y\}$ that Y takes on a value smaller than or equal to y, 47
P^B	B-conditional probability measure, 17
$P^{X=x}$	$(X{=}x)$-conditional-probability measure, 61
P-a.a. $\omega \in \Omega$	P-almost all $\omega \in \Omega$, 42
P_Y	distribution of the random variable Y (w.r.t. P), 38
$P_Y\big((-\infty, y]\big)$	Zfcaaa the value of the distribution P_Y for the interval $(-\infty, y]$, 47
$P_Y(A')$	the value of the distribution of the random variable Y for the set A', 38
$P_{Y\,\|\,X=x}$	$(X{=}x)$-conditional distribution of Y, 62
$P_Z \underset{\mathscr{A}'_Z}{\ll} P_Z^{X=x}$	P_Z is absolutely continuous on the σ-algebra $\mathscr{A}'_Z$ w.r.t. the measure $P_Z^{X=x}$, 111
$P(X{=}x \,\|\, Z) \underset{P}{>} 0$	Z-conditional probability of $\{X{=}x\}$ is greater than 0, P-almost surely, 111
$P \underset{\sigma(Z)}{\ll} P^{X=x}$	P is absolutely continuous on the σ-algebra $\sigma(Z)$ w.r.t. the measure $P^{X=x}$, 111
$P(A \,\|\, B)$	conditional probability of A given B (w.r.t. the probability measure P), 9
$P(A \,\|\, X{=}x, Z)$	a version of the partial $(X, Z{=}z)$-conditional probability of A, 114
$P(Y{=}y \,\|\, X{=}x)$	$(X{=}x)$-conditional probability of the event $\{Y{=}y\}$, 61
$P(A \,\|\, X)$	X-conditional probability of the event A, 83
$P(A \,\|\, X_1, \ldots, X_m)$	version of the conditional probability of the event A given $X = (X_1, \ldots, X_m)$, 84
$P^{X=x}$	$(X{=}x)$-conditional-probability measure, 103
$P^{X=x}(A \,\|\, Z)$	Z-conditional probability of the event A w.r.t. $P^{X=x}$, 105
$P_Z^{X=x}$	distribution of a random variable w.r.t. $P^{X=x}$, 110
$P^{X=x}(Y{=}y)$	probability of the event $\{Y{=}y\}$ w.r.t. $P^{X=x}$, 61
$P^{X=x}(Y{=}y \,\|\, Z)$	Z-conditional probability of the event $\{Y{=}y\}$ w.r.t. $P^{X=x}$, 105
$P(Y{\neq}y \,\|\, X)$	X-conditional probability of the event $\{Y{\neq}y\}$, 83
$P(Y{=}y \,\|\, X)$	X-conditional probability of the event $\{Y{=}y\}$, 83
$P^{X=x} \underset{\mathscr{A}}{\ll} P$	$P^{X=x}$ is absolutely continuous on the σ-algebra $\mathscr{A}$ w.r.t. P, 108
p_Y	probability function of a discrete random variable Y, 46
π_i	the ith projection or coordinate mapping, 34
$\mathscr{P}(\Omega)$	power set of Ω, i.e., the set of all subsets of Ω, 6

$Q(Y\|X)$	composition of X and the linear quasi-regression of Y on X, 70
$Q(Y\|X_1,\ldots,X_m)$	composition of $(X_1,\ldots,X_m)$ and the multiple linear quasi-regression $f\colon \mathbb{R}^m \to \mathbb{R}$, 73
ρ_{XY}, $Corr\,(X,Y)$	correlation of the random variables X and Y, 69
$\mathbb{R}^m$	m-fold Cartesian product of the set $\mathbb{R}$ of real numbers, 86
$(\overline{\mathbb{R}}^m, \overline{\mathscr{B}}_m)$	measurable space consisting of the set $\mathbb{R}^m$ and the Borel σ-algebra on this set, 86
$SD(Y)$, σ_Y	standard deviation of the random variable Y, 66
σ_Y^2, $Var(Y)$	variance of the random variable Y, 66
σ_{XY}, $Cov(X,Y)$	covariance of the random variables X and Y, 67
$\boldsymbol{\Sigma_{xx}}$	variance-covariance matrix of $X = (X_1,\ldots,X_m)$, 72
$\boldsymbol{\Sigma_{xy}}$	covariance matrix of $X = (X_1,\ldots,X_m)$ and $Y = (Y_1,\ldots,Y_n)$, if $n = 1$, then this is a column vector of the covariances $\sigma_{X_iY} = Cov(X_i,Y)$, $i = 1,\ldots,m$, 73
$\sigma(\mathscr{E})$	σ-algebra generated by a set system, 5
$\sigma(Y)$	an alternative notation for $Y^{-1}(\mathscr{A}'_Y)$, the σ-algebra generated by the measurable mapping $Y\colon (\Omega,\mathscr{A}) \to (\Omega'_Y,\mathscr{A}'_Y)$, 33
$\sigma(Y_1,\ldots,Y_m)$	Zpcpca σ-algebra generated by the measurable mapping $(Y_1,\ldots,Y_m)$, 35
$\sigma(Y_i, i\in I)$	σ-algebra generated by the (family of) measurable mappings $Y_i, i \in I$, 35
$\int_B f_Y\, d\lambda$	integral of f_Y over B w.r.t. λ, 49
$\int_{-\infty}^{\infty} y\, f_Y(y)\, dy$	expectation of a continuous real-valued random variable Y in terms of the Riemann integral, 57
$\int Y\, dP$, $\int_\Omega Y\, dP$	the integral of Y (over Ω) w.r.t. P, 55
$\int Y(\omega)\, P(d\omega)$	alternative notation for the integral $\int Y\, dP$, 56
$\int_a^b f_Y(y)\, dy$	Riemann integral of f_Y from a to b, 50
$Var(Y)$, σ_Y^2	variance of the random variable Y, 66
$X(\Omega)$	the image of X, i.e., the set of all values $X(\omega)$, $\omega \in \Omega$, 44
$X = Y$	X and Y are identical, 40
$X \underset{P}{=} Y$	X and Y are P-equivalent, X and Y are identical, P-almost surely, 41
$X \underset{P}{=} x$	X is identical to the constant x, P-almost surely, 61
$X \underset{P^{X=x}}{=} Y$	X and Y are identical, $P^{X=x}$-almost surely, 107
$Y\colon \Omega \to \Omega'_Y$	a mapping Y with domain Ω and co-domain Ω'_Y, 29
Y^+	positive part of the function Y, 55
Y^-	negative part of the function Y, 55
$Y\colon (\Omega,\mathscr{A}) \to (\Omega'_Y,\mathscr{A}'_Y)$	$Y\colon \Omega \to \Omega'_Y$ is an $(\mathscr{A},\mathscr{A}'_Y)$-measurable mapping on $(\Omega,\mathscr{A})$ and $(\Omega'_Y,\mathscr{A}'_Y)$ is its value space, 30
$Y^{-1}(A')$	inverse image of the set A' under the mapping Y, 29
$\{Y\in A'\}$	an alternative notation for $Y^{-1}(A')$. If Y is a random variable, then $\{Y\in A'\}$ is the event that Y takes on a value in the set A', 29

$\{Y{=}y\}$	an alternative notation for $Y^{-1}(\{y\})$. If Y is a random variable, then $\{Y{=}y\}$ is the event that Y takes on the value y, 29
$\{Y{\neq}y\}$	an alternative notation for $Y^{-1}(\Omega'_Y \setminus \{y\})$. If Y is a random variable, then $\{Y{\neq}y\}$ is the event that Y does not take on the value y, 30
$Y \underset{P}{>} \alpha$	Y is greater than α, P-almost surely, 84
$(Y_1, \ldots, Y_m)$	a multivariate mapping, 34
$Y^{-1}(\mathscr{A}'_Y)$, $\sigma(Y)$	the σ-algebra generated by Y, 33
$Z \underset{Q}{\in} B$	Z takes on a value in the set B, Q-almost surely, 109