

**Rolf Steyer and Werner Nagel**

# **Probability and Conditional Expectation**

**Fundamentals for the Empirical Sciences**

April 4, 2016

University of Jena



# Preface

## Why another book on probability?

This book has two titles. The subtitle, 'Fundamentals for the Empirical Sciences' reflects the intentions and the motivation of the first author for writing this book. He received his academic training in psychology, but considers himself a methodologist. His scientific interest is in explicating fundamental concepts of empirical research (such as causal effects and latent variables) in terms of a language that is *precise* and at the same time is *compatible with the statistical models* used in the analysis of empirical data. Applying statistical models aims at estimating and testing hypotheses about parameters such as expectations, variances, covariances, etc. or of functions of these parameters such as differences between expectations, ratios of variances, or regression coefficients, etc., all of which are terms of probability theory. Precision is necessary for securing logical consistency of theories, whereas compatibility of substantive theories with statistical models is crucial for probing the empirical validity of theoretical propositions via statistical inference.

Much empirical research uses some kind of regression in order to investigate how the expectation of one random variable depends on the values of one or more other random variables. This is true for analysis of variance, regression analysis, applications of the general linear model and the generalized linear model, factor analysis, structural equation modeling, hierarchical linear modeling, and for the analysis of qualitative data. Using these methods we aim at learning about specific regressions. A regression is a synonym for what, in probability theory, is called a factorization of a conditional expectation, provided that the regressor is numerical. This explains the main title of this book: 'Probability and Conditional Expectation'.

## What is it about?

Since the seminal book of Kolmogoroff (1933/1977) the fundamental concepts of probability theory are considered to be special concepts of measure theory. A probability measure is a special finite measure, random variables are special measurable mappings, and expectations of random variables are integrals of measurable mappings with respect to a probability measure. This motivates Part I of this book with three chapters on the measure-theoretical foundations of probability theory. Although at first sight this part seems to be far-off from prac-

tical applications, the contrary is true. This part is indispensable for probability theory and for its applications in empirical sciences. This does not only apply to the concepts of a measure and an integral but in particular to the concept of a measurable mapping, although we concede that the full relevance of this concept will become apparent only in the chapters on conditional expectations. The relevance of measurable mappings is also the reason why this chapter is more detailed than the corresponding chapters in other books on measure theory.

Part II of the book is fairly conventional. The material covered, probability, random variable, expectation, variance, covariance, and some distributions, is found in many books on probability and statistics.

Part III is not only the longest, it is also the core of the book that distinguishes it from other books on probability or on probability and statistics. Only few of these other books contain detailed chapters on conditional expectations. Exceptions are Billingsley (1995), Fristedt and Gray (1997), and Hoffmann-Jørgensen (1994). Our book does not cover any statistical model. However, we treat in much detail *what* we are estimating and *about what* we test or evaluate hypotheses using statistical models. *How* we are estimating is important but *what* we are estimating is of most interest from the empirical scientist point of view and this point is typically neglected in books on statistics and in books on probability theory such as Bauer (1996) or Klenke (2013). A simple example in case is the meaning of the coefficient  $\beta_1$  in the equation  $E(Y|X, Z) = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 ZX$ . Oftentimes, this coefficient is misinterpreted as the ‘main effect’ of  $X$ . However, sometimes  $\beta_1$  has no autonomous meaning at all, e. g. if  $P(Z=0) = 0$ . In general, this coefficient is just a component of the function  $g_1(Z) = \beta_1 + \beta_3 Z$  that can be used to compute the conditional effects of  $X$  on  $Y$  for various values  $z$  of  $Z$  (see chapter 15 for more details). The crucial point is that such concepts can be treated most clearly within probability theory, without referring to a statistical model, sample, estimation or testing. This also includes exemplifying the limitations of conditional expectations. Simple examples show that conditional expectations do not necessarily serve the purpose of the empirical researcher, namely to evaluate the effects of an intervention on an outcome variable. But even in these situations conditional expectations are indispensable for the definition of the parameters of substantive interest (see, e. g., chapter 14).

There is much overlap of Parts II and III with Steyer (2003). However, that book is written in German and the mathematics is considerably less rigorous. Aside from mathematical precision, the two books also differ in the definition of an important concept: In Steyer (2003) the term regression is used as a synonym of a conditional expectation, whereas in this book we use it as a synonym for the factorization of a conditional expectation.

In the first chapter of Part III we gently introduce conditional expectation values and discrete conditional expectations. In the next chapter of this part we then present the general theory of conditional expectations that has been introduced by Kolmogoroff (1933/1977) and is since that time treated in many books on probability theory — although much too briefly in order to be intelligible for researchers in empirical sciences. Our chapter on conditional expectations con-

tains many more details and is supplemented by a number of other chapters on important special aspects and special cases. Such a special aspect is the concept of a residual with respect to a conditional expectation. Residuals have many interesting properties and they are used in order to introduce the concepts of conditional variance and covariance, as well as the notion of a partial correlation. We then turn to specific parameterizations of a conditional expectation, including the concepts of a linear regression and a linear logistic regression. Note that these concepts are introduced as probabilistic concepts. As mentioned above, they are what we aim at estimating in applying the corresponding statistical models. The next two chapters provide the probabilistic foundations of the analysis of conditional and average effects of treatments, interventions, or expositions to potentially harmful or beneficial environments. To our knowledge this material is not found in any other text book. Note, however, that although these two chapters provide important concepts, they do not cover the theory of causal effects, which is another book project of the first author.

Part IV uses conditional expectations in order to introduce conditional independence and conditional distributions. Although these two chapters are more extensive than comparable chapters or sections in other books, the material is found in other books on probability theory as well.

### **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 a proper language, i. e., 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 terms of probability theory what applied statisticians and empirical researchers estimate and test in their research. Both kinds of readers are potential contributors to the methodology of empirical sciences.

Many exercises and their solutions provide extensive material for assignments in courses, but they also facilitate independent learning. At the same time these exercises and their solutions help streamlining the main text.

Note that we do not provide all proofs, in particular in the chapters on measure, integral, and distributions. In these cases we refer to other textbooks, instead. We decided to include only those proofs that may help to increase understanding of the background and to learn important mathematical procedures. Of course, we provide proofs of all propositions for which we did not find an appropriate reference.

### **Prerequisites**

We assume that the reader is familiar with the elementary concepts of logic, sets, functions, sequences, and matrices as presented, e. g., in chapters 1 and 2 of Rosen (2012). We try to stick to his notation as close as possible. One of the exceptions is the symbol for the implication for which we use  $\Rightarrow$  instead of  $\rightarrow$ . Another

exception is the symbol for the equivalence for which we use  $\Leftrightarrow$  instead of  $\leftrightarrow$ . Box 0.1 summarizes the most important notation to start with. The concepts referred to by these symbols are defined, e. g., in Rosen (2012) or in Ellis and Gulick (2006). For a rich collection of mathematical formulae we recommend the handbooks of Harris and Stocker (1998) and Bronshtein, Semendyayev, Musiol, and Mühlig (2015).

### **Acknowledgements**

This book could not have been written without the help of many. First of all we thank 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 supporting us with respect to LaTeX, finding errors or suggesting other improvements we also thank Karoline Bading, Marcel Bauer, Sonja Hahn, Gregor Kappler, Christoph Kiefer, Andreas Neudecker, Axel Mayer, Erik Sengewald, Jan Plötner, and Tom Landes. Thanks are also due to Ernesto San Martin for suggesting section 1.7 and proposition (iv) of Theorem 16.37. The proof of Lemma 12.38 is due to Peter Vogel. Finally, we would like to thank our students who kept us thinking on how to improve the text.

**Box 0.1 A First List of Symbols**

$\neg$	not
$\wedge$	and
$\vee$	or
$\Rightarrow$	implies
$\Leftrightarrow$	equivalent to
$\exists$	there is (there exists)
$\forall$	for all
$a \in A$	$a$ is an element of the set $A$
$\emptyset$	empty set
$I$	nonempty finite, countable, or uncountable index set
$(A_i, i \in I)$	family of sets $A_i, i \in I$
$A \cup B$	union of the sets $A$ and $B$
$\bigcup_{i \in I} A_i$	union of the sets $A_i, i \in I$
$A \cap B$	intersection of the sets $A$ and $B$
$\bigcap_{i \in I} A_i$	intersection of the sets $A_i, i \in I$
$A \setminus B$	set difference of the set $A$ and the set $B$
$A^c := \Omega \setminus A$	complement of a set $A \subset \Omega$ with respect to a set $\Omega$
$A \subset B$	$A$ is a subset of the set $B$ ; $A \subset B$ includes $A = B$
$A \times B$	Cartesian product (product set) of $A$ and $B$
$\prod_{i=1}^n A_i$	Cartesian product of the sets $A_i, i = 1, \dots, n$
$f: A \rightarrow B$	mapping $f$ assigning to each $a \in A$ (the domain) one and only one element $b \in B$ (the codomain)
$\sum_{i=1}^n a_i$	sum of the real numbers $a_1, \dots, a_n$
$\prod_{i=1}^n a_i$	product of the real numbers $a_1, \dots, a_n$
$\lim_{n \rightarrow \infty} a_n$	limit of a sequence $a_1, a_2, \dots$ of real numbers
$\sum_{i=1}^{\infty} a_i$	$\lim_{n \rightarrow \infty} \sum_{i=1}^n a_i$ , where $a_1, a_2, \dots$ and therefore $\sum_{i=1}^1 a_i, \sum_{i=1}^2 a_i, \dots$ are sequences of real numbers



# Contents

## Part I Measure-Theoretical Foundations of Probability Theory

<b>1 Measure</b> .....	3
1.1 Introductory Examples .....	3
1.2 $\sigma$ -Algebra and Measurable Space .....	4
1.2.1 $\sigma$ -Algebra Generated by a Set System .....	9
1.2.2 $\sigma$ -Algebra of Borel Sets on $\mathbb{R}^n$ .....	12
1.2.3 $\sigma$ -Algebra on a Cartesian Product .....	13
1.2.4 $\cap$ -Stable Set Systems That Generate a $\sigma$ -Algebra .....	15
1.3 Measure and Measure Space .....	16
1.3.1 $\sigma$ -Additivity and Related Properties .....	17
1.3.2 Other Properties .....	18
1.4 Specific Measures .....	20
1.4.1 Dirac Measure and Counting Measure .....	21
1.4.2 Lebesgue Measure .....	22
1.4.3 Other Examples of a Measure .....	23
1.4.4 Finite and $\sigma$ -Finite Measures .....	23
1.4.5 Product Measure .....	24
1.5 Continuity of a Measure .....	25
1.6 Specifying a Measure via a Generating System .....	27
1.7 $\sigma$ -Algebra That is Trivial With Respect to a Measure .....	28
1.8 Proofs .....	28
1.9 Exercises .....	31
<b>2 Measurable Mapping</b> .....	41
2.1 Image and Inverse Image .....	41
2.2 Introductory Examples .....	42
2.2.1 Example 1: Rectangles .....	42
2.2.2 Example 2: Flipping two Coins .....	44
2.3 Measurable Mapping .....	46
2.3.1 Measurable Mapping .....	46
2.3.2 $\sigma$ -Algebra Generated by a Mapping .....	51
2.3.3 Final $\sigma$ -Algebra .....	54
2.3.4 Multivariate Mapping .....	54
2.3.5 Projection Mapping .....	56
2.3.6 Measurability With Respect to a Mapping .....	56

2.4	Theorems on Measurable Mappings .....	58
2.4.1	Measurability of a Composition .....	59
2.4.2	Theorems on Measurable Functions .....	61
2.5	Equivalence of Two Mappings With Respect to a Measure .....	64
2.6	Image Measure .....	67
2.7	Proofs .....	70
2.8	Exercises .....	75
<b>3</b>	<b>Integral</b> .....	<b>83</b>
3.1	Definition .....	83
3.1.1	Integral of a Nonnegative Step Function .....	83
3.1.2	Integral of a Nonnegative Measurable Function .....	88
3.1.3	Integral of a Measurable Function .....	93
3.2	Properties .....	96
3.2.1	Integral of $\mu$ -Equivalent Functions .....	98
3.2.2	Integral With Respect to a Weighted Sum of Measures .....	100
3.2.3	Integral With Respect to an Image Measure .....	102
3.2.4	Convergence Theorems .....	103
3.3	Lebesgue and Riemann Integral .....	104
3.4	Density .....	106
3.5	Absolute Continuity and the Radon-Nikodym Theorem .....	108
3.6	Integral With Respect to a Product Measure .....	110
3.7	Proofs .....	111
3.8	Exercises .....	120
<b>Part II</b>	<b>Probability, Random Variable and its Distribution</b>	
<b>4</b>	<b>Probability Measure</b> .....	<b>127</b>
4.1	Probability Measure and Probability Space .....	127
4.1.1	Definition .....	127
4.1.2	Formal and Substantive Meaning of Probabilistic Terms .....	128
4.1.3	Properties of a Probability Measure .....	128
4.1.4	Examples .....	130
4.2	Conditional Probability .....	132
4.2.1	Definition .....	132
4.2.2	Filtration and Time Order Between Events and Sets of Events .....	133
4.2.3	Multiplication Rule .....	135
4.2.4	Examples .....	136
4.2.5	Theorem of Total Probability .....	137
4.2.6	Bayes' Theorem .....	138
4.2.7	Conditional-Probability Measure .....	139
4.3	Independence .....	143
4.3.1	Independence of Events .....	143
4.3.2	Independence of Set Systems .....	144
4.4	Conditional Independence Given an Event .....	145

4.4.1	Conditional Independence of Events Given an Event . . . . .	145
4.4.2	Conditional Independence of Set Systems Given an Event . . . . .	146
4.5	Proofs . . . . .	148
4.6	Exercises . . . . .	150
<b>5</b>	<b>Random Variable, Distribution, Density, and Distribution Function . . . . .</b>	<b>155</b>
5.1	Random Variable and its Distribution . . . . .	155
5.2	Equivalence of Two Random Variables With Respect to a Probability Measure . . . . .	161
5.2.1	Identical and $P$ -Equivalent Random Variables . . . . .	161
5.2.2	$P$ -Equivalence, $P^B$ -Equivalence, and Absolute Continuity . . . . .	164
5.3	Multivariate Random Variable . . . . .	167
5.4	Independence of Random Variables . . . . .	169
5.5	Probability Function of a Discrete Random Variable . . . . .	175
5.6	Probability Density With Respect to a Measure . . . . .	178
5.6.1	General Concepts and Properties . . . . .	178
5.6.2	Density of a Discrete Random Variable . . . . .	180
5.6.3	Density of a Bivariate Random Variable . . . . .	180
5.7	Uni- or Multivariate Real-Valued Random Variable . . . . .	182
5.7.1	Distribution Function of a Univariate Real-Valued Random Variable . . . . .	182
5.7.2	Distribution Function of a Multivariate Real-Valued Random Variable . . . . .	184
5.7.3	Density of a Continuous Univariate Real-Valued Random Variable . . . . .	185
5.7.4	Density of a Continuous Multivariate Real-Valued Random Variable . . . . .	187
5.8	Proofs . . . . .	188
5.9	Exercises . . . . .	196
<b>6</b>	<b>Expectation, Variance, and Other Moments . . . . .</b>	<b>199</b>
6.1	Expectation . . . . .	199
6.1.1	Definition . . . . .	199
6.1.2	Expectation of a Discrete Random Variable . . . . .	200
6.1.3	Computing the Expectation Using a Density . . . . .	202
6.1.4	Transformation Theorem . . . . .	203
6.1.5	Rules of Computation . . . . .	206
6.2	Moments, Variance, and Standard Deviation . . . . .	207
6.3	Proofs . . . . .	212
6.4	Exercises . . . . .	213
<b>7</b>	<b>Linear Quasi-Regression, Covariance, and Correlation . . . . .</b>	<b>217</b>
7.1	Linear Quasi-Regression . . . . .	217
7.2	Covariance . . . . .	220
7.3	Correlation . . . . .	224
7.4	Expectation Vector and Covariance Matrix . . . . .	227

7.4.1	Random Vector and Random Matrix .....	227
7.4.2	Expectation of a Random Vector and a Random Matrix ....	228
7.4.3	Covariance Matrix of two Multivariate Random Variables..	229
7.5	Multiple Linear Quasi-Regression .....	231
7.6	Proofs .....	233
7.7	Exercises .....	237
<b>8</b>	<b>Some Distributions .....</b>	<b>245</b>
8.1	Some Distributions of Discrete Random Variables .....	245
8.1.1	Discrete Uniform Distribution .....	245
8.1.2	Bernoulli Distribution .....	246
8.1.3	Binomial Distribution .....	247
8.1.4	Poisson Distribution .....	250
8.1.5	Geometric Distribution .....	252
8.2	Some Distributions of Continuous Random Variables .....	254
8.2.1	Continuous Uniform Distribution .....	254
8.2.2	Normal Distribution.....	256
8.2.3	Multivariate Normal Distribution.....	259
8.2.4	Central $\chi^2$ -Distribution .....	262
8.2.5	Central $t$ -Distribution .....	264
8.2.6	Central $F$ -Distribution.....	266
8.3	Proofs .....	267
8.4	Exercises .....	271
<b>Part III Conditional Expectation and Regression</b>		
<b>9</b>	<b>Conditional Expectation Value and Discrete Conditional Expectation</b>	<b>277</b>
9.1	Conditional Expectation Value .....	277
9.2	Transformation Theorem .....	280
9.3	Other Properties .....	282
9.4	Discrete Conditional Expectation .....	283
9.5	Discrete Regression .....	285
9.6	Examples .....	287
9.7	Proofs .....	291
9.8	Exercises .....	291
<b>10</b>	<b>Conditional Expectation .....</b>	<b>295</b>
10.1	Assumptions and Definitions .....	295
10.2	Existence and Uniqueness .....	297
10.2.1	Uniqueness With Respect to a Probability Measure .....	298
10.2.2	A Necessary and Sufficient Condition of Uniqueness .....	299
10.2.3	Examples .....	300
10.3	Rules of Computation and Other Properties .....	301
10.3.1	Rules of Computation .....	301
10.3.2	Monotonicity .....	302
10.3.3	Convergence Theorems .....	302

10.4	Factorization, Regression, and Conditional Expectation Value . . . .	306
10.4.1	Existence of a Factorization . . . . .	306
10.4.2	Conditional Expectation and Mean-Squared Error . . . . .	307
10.4.3	Uniqueness of a Factorization . . . . .	308
10.4.4	Conditional Expectation Value . . . . .	309
10.5	Characterizing a Conditional Expectation by the Joint Distribution	312
10.6	Conditional Mean Independence . . . . .	313
10.7	Proofs . . . . .	318
10.8	Exercises . . . . .	321
<b>11</b>	<b>Residual, Conditional Variance, and Conditional Covariance</b> . . . . .	<b>329</b>
11.1	Residual With Respect to a Conditional Expectation . . . . .	329
11.2	Coefficient of Determination and Multiple Correlation . . . . .	333
11.3	Conditional Variance and Covariance Given a $\sigma$ -Algebra . . . . .	338
11.4	Conditional Variance and Covariance Given a Value of a Random Variable . . . . .	339
11.5	Properties of Conditional Variances and Covariances . . . . .	342
11.6	Partial Correlation . . . . .	345
11.7	Proofs . . . . .	347
11.8	Exercises . . . . .	348
<b>12</b>	<b>Linear Regression</b> . . . . .	<b>357</b>
12.1	Basic Ideas . . . . .	357
12.2	Assumptions and Definitions . . . . .	359
12.3	Examples . . . . .	361
12.4	Linear Quasi-Regression . . . . .	366
12.5	Uniqueness and Identification of Regression Coefficients . . . . .	367
12.6	Linear Regression . . . . .	369
12.7	Parametrizations of a Discrete Conditional Expectation . . . . .	370
12.8	Invariance of Regression Coefficients . . . . .	374
12.9	Proofs . . . . .	375
12.10	Exercises . . . . .	377
<b>13</b>	<b>Linear Logistic Regression</b> . . . . .	<b>381</b>
13.1	Logit Transformation of a Conditional Probability . . . . .	381
13.2	Linear Logistic Parametrization . . . . .	383
13.3	A Parametrization of a Discrete Conditional Probability . . . . .	385
13.4	Identification of Coefficients of a Linear Logistic Parametrization . . . . .	387
13.5	Linear Logistic Regression and Linear Logit Regression . . . . .	388
13.6	Proofs . . . . .	394
13.7	Exercises . . . . .	396

<b>14</b>	<b>Conditional Expectation With Respect to a Conditional-Probability Measure</b> .....	399
14.1	Introductory Examples .....	399
14.2	Assumptions and Definitions .....	404
14.3	Properties .....	411
14.4	Partial Conditional Expectation .....	412
14.5	Factorization .....	413
14.5.1	Conditional Expectation Value With Respect to $P^B$ .....	414
14.5.2	Uniqueness of Factorizations .....	415
14.6	Uniqueness .....	415
14.6.1	A Necessary and Sufficient Condition of Uniqueness .....	415
14.6.2	Uniqueness w.r.t. $P$ and Other Probability Measures .....	417
14.6.3	Necessary and Sufficient Conditions of $P$ -Uniqueness .....	418
14.6.4	Properties Related to $P$ -Uniqueness .....	420
14.7	Conditional Mean Independence With Respect to $P^{Z=z}$ .....	424
14.8	Proofs .....	426
14.9	Exercises .....	431
<b>15</b>	<b>Conditional Effect Functions of a Discrete Regressor</b> .....	437
15.1	Assumptions and Definitions .....	437
15.2	Conditional Intercept Function and Effect Functions .....	438
15.3	Implications of Independence of $X$ and $Z$ for Regression Coefficients .....	441
15.4	Adjusted Conditional Effect Functions .....	443
15.5	Conditional Logit Effect Functions .....	447
15.6	Implications of Independence of $X$ and $Z$ for the Logit Regression Coefficients .....	450
15.7	Proofs .....	452
15.8	Exercises .....	454
<b>Part IV Conditional Independence and Conditional Distribution</b>		
<b>16</b>	<b>Conditional Independence</b> .....	459
16.1	Assumptions and Definitions .....	459
16.1.1	Two Events .....	459
16.1.2	Two Sets of Events .....	461
16.1.3	Two Random Variables .....	462
16.2	Properties .....	463
16.3	Conditional Independence and Conditional Mean Independence .....	470
16.4	Families of Events .....	473
16.5	Families of Set Systems .....	473
16.6	Families of Random Variables .....	475
16.7	Proofs .....	478
16.8	Exercises .....	486

**17 Conditional Distribution** ..... 491

17.1 Conditional Distribution Given a  $\sigma$ -Algebra or a Random Variable 491

17.2 Conditional Distribution Given a Value of a Random Variable .... 494

17.3 Existence and Uniqueness ..... 497

    17.3.1 Existence ..... 497

    17.3.2 Uniqueness of the Functions  $P_{Y|G}(\cdot, A')$  ..... 498

    17.3.3 Common Null Set (CNS) Uniqueness of a Conditional  
        Distribution ..... 499

17.4 Conditional-Probability Measure Given a Value of a Random  
    Variable ..... 502

17.5 Decomposing the Joint Distribution of Random Variables ..... 504

17.6 Conditional Independence and Conditional Distributions ..... 506

17.7 Expectations With Respect to a Conditional Distribution ..... 511

17.8 Conditional Distribution Function and Probability Density ..... 513

17.9 Conditional Distribution and Radon-Nikodym Density ..... 516

17.10 Proofs ..... 520

17.11 Exercises ..... 536

**References** ..... 541



## List of Figures

1.1	A Venn diagram of two sets and their intersection	4
1.2	Example of a tree representation of a Cartesian product	7
1.3	A half-open rectangle in the plane $\mathbb{R}^2$	12
1.4	Approximation of an open egg-shaped set $O$ from below	25
1.5	Approximation of an open egg-shaped set $O$ from above	26
2.1	Rectangles and their images under a function	42
2.2	Rectangles and their inverse images under a function	43
2.3	A set and its image under a function	44
2.4	A set and its inverse image under a function	45
2.5	A composition of two mappings	57
2.6	The two step functions $f$ and $h$ in Example 2.55	60
2.7	Positive and negative parts of a function, and its sign function	63
3.1	A partition and a subset of $\Omega$	87
3.2	Increasing nonnegative step functions	90
3.3	Integral of nonnegative step functions w.r.t. the Lebesgue measure	92
3.4	Lebesgue integral of a function from $-5$ to $5$	95
3.5	Illustrating the construction of the Riemann integral 3.22	105
3.6	Integral of a density for two intervals	108
4.1	A filtration with $T = \{1, 2, 3\}$	134
4.2	Probability tree illustrating the multiplication rule	137
4.3	Venn diagram illustrating a partition of a set	138
5.1	Two random variables that are $P^B$ -equivalent if $P(\{\omega_3\}) = 0$	163
6.1	Variance of an indicator of an event as a function of its probability	209
7.1	Linear quasi-regression	218
7.2	The regressor $X$ , the linear quasi-regression $f$ and their composition $Q_{lin}(Y X) = f(X)$	219
7.3	Intercept and slope of a linear function $f: \mathbb{R} \rightarrow \mathbb{R}$	220
8.1	Probability function and distribution function of a binomial distribution	248
8.2	Probability function of the sum of i. i. d. Bernoulli variables	249

8.3	Probability functions of binomial and Poisson distributions .....	251
8.4	Probability functions of three geometric distributions .....	254
8.5	Density of a bivariate uniform distribution on a rectangle .....	255
8.6	Densities of three normal distributions .....	257
8.7	Density and distribution function of the standard normal distribution .....	258
8.8	Density of a bivariate normal distribution .....	260
8.9	Densities of three central $\chi^2$ -distributions .....	263
8.10	Densities of the standard normal and three $t$ -distributions .....	265
8.11	Densities of four central $F$ -distributions .....	266
9.1	The conditional expectation $E(Y X)$ as the composition of $X$ and its factorization $g$ .....	286
11.1	Conditional expectation of the residual on its regressor .....	331
12.1	$E(Y X)$ as the composition of $X$ and the linear regression $g$ .....	369
12.2	A regression with a parametrization that is linear in $(X, X^2)$ but not in $X$ .....	374
13.1	Graph of the logit transformation of $p$ .....	383
13.2	Graphs of three logistic functions .....	385
13.3	$P(Y=1 X)$ as the composition of $X$ , the linear logit regression $f$ , and the logistic function $h$ . .....	389
13.4	Graphs of the $\xi$ -conditional probabilities of items $Y_i$ satisfying the Rasch model .....	394
16.1	Likelihood functions for the response patterns of three variables $Y_i$ satisfying the Rasch model .....	478

## List of Tables

2.1	Example of Measurable Sets Represented by a Mapping $X$ . . . . .	45
2.2	Joe and Ann With Random Assignment and Measurable Mappings . .	53
4.1	Joe and Ann With Random Assignment: Probability Measures . . . . .	131
5.1	Tom, Jim, and Kate . . . . .	159
5.2	No Treatment for Joe . . . . .	166
9.1	Joe and Ann With Random Assignment: Conditional Expectations . .	286
9.2	No Treatment for Joe With Conditional Expectations . . . . .	289
10.1	Joe and Ann With no Treatment Effects . . . . .	314
11.1	No Treatment for Joe With Conditional Expectations and Residuals .	333
11.2	Joe and Ann With Self-Selection and Residuals . . . . .	341
12.1	Joe and Ann With Self-Selection: Conditional Expectations . . . . .	359
13.1	Joe and Ann With a Latent Ability Variable . . . . .	392
14.1	Joe and Ann With Random Assignment: Conditional Expectations With Respect to $P^{X=x}$ . . . . .	400
14.2	Joe and Ann With Self-Selection: Conditional Expectations With Respect to $P^{X=x}$ . . . . .	402
14.3	No Treatment for Joe: Conditional Expectations With Respect to $P^{X=x}$	403
15.1	Four Persons With Self-Selection to Treatment . . . . .	445
15.2	Joe and Ann: Reversed Average Logit Effect . . . . .	452
16.1	Joe and Ann With no Individual Treatment Effects . . . . .	466
16.2	$Z$ -Conditional Independence of $X$ and $U$ . . . . .	471
17.1	Joe and Ann With Self-Selection: Conditional Distribution $P_{Y X}$ . . . . .	493



## List of Boxes

0.1	A First List of Symbols	IX
1.1	Rules of Computation for Measures	19
4.1	Rules of Computation for Probabilities	129
4.2	Independence and Conditional Independence of Events	147
6.1	Rules of Computation for Expectations	207
6.2	Rules of Computation for Variances	209
7.1	Rules of Computation for Covariances	222
7.2	Rules of Computation for Expectations of Random Matrices	229
7.3	Rules of Computation for Covariance Matrices	231
9.1	Rules of Computation for $B$ -Conditional Expectation Values	283
9.2	Rules of Computation for $(X=x)$ -Conditional Expectation Values	284
10.1	Rules of Computation for $\mathcal{C}$ -Conditional Expectations	303
10.2	Rules of Computation for $X$ -Conditional Expectations	304
10.3	Monotonicity of Conditional Expectations	305
11.1	Rules of Computation for a Residual	332
11.2	Rules of Computation for $\mathcal{C}$ -Conditional Covariances	343
11.3	Rules of Computation for Conditional Variances	344
14.1	$P$ -Uniqueness of $E^B(Y \mathcal{C})$	421
16.1	Notation	465
16.2	Conditional Independence of Set Systems	469
16.3	Conditional Independence of Random Variables	470



**Part I**  
**Measure-Theoretical Foundations of**  
**Probability Theory**



# Chapter 1

## Measure

In this chapter, we introduce the concept of a measure and other closely related notions. We start with some examples and then introduce the concept of a  $\sigma$ -algebra, which is crucial in measure theory and probability theory. At first glance this concept seems to be a pure technical construction, which is usually not dealt with in textbooks on ‘Probability and Statistics’ for empirical sciences. However, a  $\sigma$ -algebra turned out to be the natural domain for a measure, including probability measures. Moreover, in probability theory, a  $\sigma$ -algebra is not only the domain of probability measures. The  $\sigma$ -algebra generated by a random variable can be interpreted as the set of events that is represented by this random variable. This is treated in more detail in chapter 2 on measurable mappings, which provides the general theory of random variables because random variables are measurable mappings. The virtues of  $\sigma$ -algebras will become fully apparent in chapter 10 on conditional expectations and its subsequent chapters. The pair  $(\Omega, \mathcal{A})$  consisting of a nonempty set  $\Omega$  and a  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$  is called a *measurable space*. Such a measurable space is crucial for the definition of a *measure*. Next, we treat some important examples of measures, including the *counting measure*, the *Dirac measure*, and the *Lebesgue measure*. Finally, we turn to *continuity* and *uniqueness* properties of a measure.

### 1.1 Introductory Examples

Consider Figure 1.1 showing the set  $\Omega$  of all points  $(x, y)$  inside the rectangle and the sets  $A$  and  $B$  of all points  $(x, y)$  inside the two ellipses, respectively. These three sets are subsets of the plane  $\mathbb{R}^2 := \mathbb{R} \times \mathbb{R}$ , where  $\mathbb{R}$  denotes the *set of all real numbers*, and  $\mathbb{R} \times \mathbb{R} := \{(a, b) : a, b \in \mathbb{R}\}$  is the set of all ordered pairs  $(a, b)$  with  $a, b \in \mathbb{R}$ , called the *Cartesian product* or *product set* of  $\mathbb{R}$  with itself. In Figure 1.1, the sets  $A$  and  $B$  have a nonempty intersection. Now let  $area(A)$  and  $area(B)$  denote their areas and  $area(A \cap B)$  the area of their intersection. Inspecting this figure reveals:

$$area(A \cup B) = area(A) + area(B) - area(A \cap B).$$

This example illustrates three important points:

- (a) A measure such as *area* is a function on a *set system on  $\Omega$* , i. e., on a *set of subsets* of a set  $\Omega$  such as  $A$ ,  $B$ , and  $A \cap B$ .

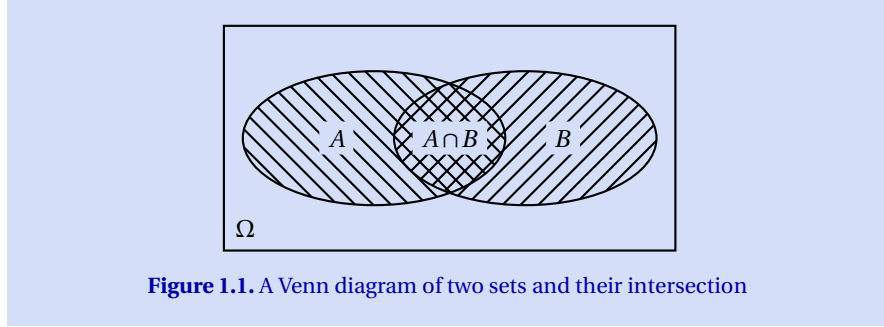


Figure 1.1. A Venn diagram of two sets and their intersection

- (b) If *area* is defined for the subsets  $A, B \subset \Omega$ , then it is also defined for their *intersection*  $A \cap B$  and for their *union*  $A \cup B$ .
- (c) Measures are *additive*. In other words, if  $A$  and  $B$  are *disjoint* subsets of  $\Omega$ , i. e., if  $A \cap B = \emptyset$ , then  $\text{area}(A \cup B) = \text{area}(A) + \text{area}(B)$ .

Note that, in the example presented in Figure 1.1, the sets  $A$  and  $B$  are *not disjoint*, and this is why  $\text{area}(A \cap B)$  has to be subtracted in the equation displayed above. Points (a) to (c) also apply to other measures such as *length* and *volume* as well as to *probability measures*. Therefore, we adopt a more general language and talk about subsets  $A, B$  of a set  $\Omega$  (or *measurable sets*  $A, B$ ) and their *measure*  $\mu$  instead of lines and their lengths, rectangles and their areas, cubes and their volume, or events and their probabilities.

For example, if  $\Omega = \{1, \dots, 6\}$  denotes the set of possible outcomes of tossing a fair dice,  $A = \{1, 6\}$  and  $B = \{2, 4, 6\}$  denote the events of tossing a 1 or a 6 and tossing an *even number*, respectively. Furthermore,  $A \cap B = \{6\}$  and the probability of tossing a 1 or a 6 or an *even number* — the event  $A \cup B$  — is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{4}{6}.$$

In the first example, the measure *area* assigns a real number to a subset of  $\mathbb{R}^2$ . In the second example, the measure  $P$  assigns a real number to a subset of  $\Omega = \{1, \dots, 6\}$ . This suggests that a measure should be defined such that it assigns a real number *to all subsets* of a set, i. e., to all elements of the power set. Unfortunately, this may lead to contradictions (see, e. g., Georgii, 2008, p. 9-10). In contrast, defining a measure on a  $\sigma$ -algebra such contradictions can be avoided.

## 1.2 $\sigma$ -Algebra and Measurable Space

In the following definition, we consider a set system  $\mathcal{A}$  on  $\Omega$ , a sequence  $A_1, A_2, \dots$  of subsets of  $\Omega$ , and their countable union. Remember, a *set system on a set*  $\Omega$  is a set of subsets of  $\Omega$  presuming that  $\Omega$  is not empty. A *sequence of subsets of a set*  $\Omega$  is a function from the set  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$  or  $\mathbb{N} = \{1, 2, \dots\}$  or a subset of these sets

to  $\mathcal{P}(\Omega)$ , the *power set* of  $\Omega$ . Furthermore, the *finite union* of the sets  $A_1, \dots, A_n$  and the *countable union* of the sets  $A_1, A_2, \dots$  are defined by

$$\bigcup_{i=1}^n A_i := \{a \in \Omega : \exists i \in \{1, \dots, n\} : a \in A_i\} \quad (1.1)$$

and

$$\bigcup_{i=1}^{\infty} A_i := \{a \in \Omega : \exists i \in \mathbb{N} : a \in A_i\}, \quad (1.2)$$

respectively. Hence, by definition,  $\bigcup_{i=1}^n A_i$  is the set of all elements that are an element of at least one of the sets  $A_i$ ,  $i = 1, \dots, n$ , and  $\bigcup_{i=1}^{\infty} A_i$  is the set of all elements that are an element of at least one of the sets  $A_i$ ,  $i \in \mathbb{N}$ . Finally,  $A^c := \Omega \setminus A$  denotes the complement of  $A$  (with respect to  $\Omega$ ).

**Definition 1.1 ( $\sigma$ -Algebra)**

A set  $\mathcal{A}$  of subsets of a nonempty set  $\Omega$  is called a  $\sigma$ -algebra (or  $\sigma$ -field) on  $\Omega$ , if the following three conditions hold:

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

An element of a  $\sigma$ -algebra is called a *measurable set*.

**Remark 1.2 (Closure With Respect to Set Operations)** Condition (c) postulates that  $\sigma$ -algebras are closed with respect to *countable* unions of sets  $A_1, A_2, \dots \in \mathcal{A}$ . However, in conjunction with (a) and (b) this implies that a  $\sigma$ -algebra is also closed with respect to *finite* unions of sets  $A_1, \dots, A_n \in \mathcal{A}$ , because every finite union of sets  $A_1, \dots, A_n \in \mathcal{A}$  can be represented as a countable union of the sets that are elements of  $\mathcal{A}$ , e. g.,

$$\bigcup_{i=1}^n A_i = A_1 \cup \dots \cup A_n \cup \emptyset \cup \emptyset \cup \dots \quad (1.3)$$

Note that (a) and (b) imply  $\emptyset \in \mathcal{A}$ , because  $\Omega^c = \emptyset$ .

Furthermore, although condition (c) only requires explicitly that  $\sigma$ -algebras are closed with respect to countable unions, Definition 1.1 implies that a  $\sigma$ -algebra is closed also with respect to intersections such as  $A_1 \cap A_2$  and set differences  $A_1 \setminus A_2$ . In other words, if  $A_1$  and  $A_2$  are elements of  $\mathcal{A}$ , then  $A_1 \cup A_2$ ,  $A_1 \cap A_2$ , and  $A_1 \setminus A_2$  are elements of  $\mathcal{A}$  as well, provided that  $\mathcal{A}$  is a  $\sigma$ -algebra. The same is true for countable intersections  $A_1 \cap A_2 \cap \dots$  of elements of  $\mathcal{A}$ . In more formal terms: If  $\mathcal{A}$  is a  $\sigma$ -algebra, then

$$A_1, A_2, \dots \in \mathcal{A} \quad \Rightarrow \quad \bigcap_{i=1}^{\infty} A_i \in \mathcal{A} \quad (1.4)$$

(see Exercise 1-1), where  $\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \dots$  is defined by

$$\bigcap_{i=1}^{\infty} A_i := \{a \in \Omega : \forall i \in \mathbb{N} : a \in A_i\}. \quad (1.5)$$

Because

$$\bigcap_{i=1}^n A_i = A_1 \cap \dots \cap A_n \cap \Omega \cap \Omega \cap \dots, \quad (1.6)$$

we can also conclude

$$A_1, \dots, A_n \in \mathcal{A} \Rightarrow \bigcap_{i=1}^n A_i \in \mathcal{A}, \quad (1.7)$$

where  $\bigcap_{i=1}^n A_i$ , the *finite intersection* of the sets  $A_1, \dots, A_n$ , is defined by

$$\bigcap_{i=1}^n A_i := \{a \in \Omega : \forall i \in \{1, \dots, n\} : a \in A_i\}. \quad (1.8)$$

◁

**Remark 1.3 (Countable and Uncountable Unions)** Defining a  $\sigma$ -algebra we use the symbol  $\sigma$  in order to emphasize that unions of finitely or countably many sets are considered, *but not other unions of sets*. For example, the *closed interval*  $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b, a, b \in \mathbb{R}\}$  on the real axis is identical to the union of singletons  $\{x\}$  that contain only one single element  $x \in \mathbb{R}$ , i. e.,

$$[a, b] = \bigcup_{a \leq x \leq b} \{x\}. \quad (1.9)$$

This union is neither finite nor countable. Hence, condition (c) of Definition 1.1 does *not* imply that this union is necessarily an element of a  $\sigma$ -algebra  $\mathcal{A}$  on  $\mathbb{R}$ , even if all singletons  $\{x\}$ ,  $x \in \mathbb{R}$ , are elements of  $\mathcal{A}$ . ◁

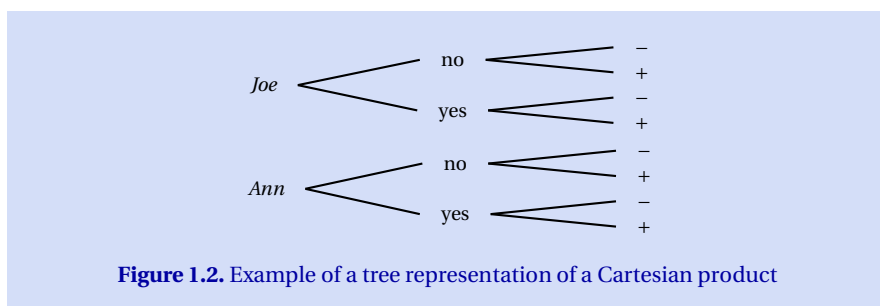
The following notion of a *measurable space* proves to be convenient in measure theory.

**Definition 1.4 (Measurable Space)**

If  $\Omega$  is a nonempty set and  $\mathcal{A}$  a  $\sigma$ -algebra on  $\Omega$ , then the pair  $(\Omega, \mathcal{A})$  is called a *measurable space*.

**Example 1.5 (The Smallest  $\sigma$ -Algebra)** The smallest  $\sigma$ -algebra on a set  $\Omega$  is  $\mathcal{A} = \{\Omega, \emptyset\}$ . It contains only the elements  $\Omega$  and the empty set  $\emptyset$ . As is easily seen,  $\Omega \cup \emptyset = \Omega$ ,  $\Omega^c = \emptyset$ , and  $\emptyset^c = \Omega$  are elements of  $\mathcal{A}$ . This shows that  $\mathcal{A} = \{\Omega, \emptyset\}$  is closed with respect to union and complement. ◁

**Example 1.6 (Power Set)** The *power set*  $\mathcal{P}(\Omega)$  of  $\Omega$ , i. e., the *set of all subsets* of  $\Omega$ , is always a  $\sigma$ -algebra on  $\Omega$ . It is the largest  $\sigma$ -algebra on a set  $\Omega$ . All other  $\sigma$ -algebras on  $\Omega$  are subsets of  $\mathcal{P}(\Omega)$ . ◁



**Example 1.7 (A Small  $\sigma$ -Algebra)** If  $A$  is a subset of  $\Omega$ , then  $\mathcal{A} = \{\Omega, \emptyset, A, A^c\}$  is always a  $\sigma$ -algebra on  $\Omega$  (see Exercise 1-2). Again, it is easily seen that this set system is closed with respect to union and complement.  $\triangleleft$

**Remark 1.8 (Motivation for  $\sigma$ -Algebras)** These examples show that there can be many different  $\sigma$ -algebras on a set  $\Omega$ . Why not simply always use the largest one, the power set  $\mathcal{P}(\Omega)$ ? In fact, this would be possible as long as  $\Omega$  is finite or countable. There are at least three reasons for using  $\sigma$ -algebras. *First*, there are important sets  $\Omega$  (e.g.,  $\Omega = \mathbb{R}$ ) such that measures of interest (e.g., *length* — which is the Lebesgue measure pertaining to  $\Omega = \mathbb{R}$ ) cannot be defined on  $\mathcal{P}(\Omega)$  (see e.g., Wise & Hall, 1993, counterexample 1.25). These measures can be defined, however, on other  $\sigma$ -algebras, such as the Borel- $\sigma$ -algebra [see Eq. (1.18)]. (For an example in which the power set is ‘too large’ see Georgii, 2008, p. 9-10). *Second*, in some sense,  $\sigma$ -algebras contain those elements of a larger set system that are relevant for a particular question. In probability theory, together with  $\Omega$  and a probability measure, each  $\sigma$ -algebra on  $\Omega$  represents a random experiment that is in some sense contained in a (often larger) random experiment. For example, if we consider the random experiment of tossing a dice, then we may focus on whether or not the number is even. Together with  $\Omega$  and the probability measure, the corresponding  $\sigma$ -algebra represents a ‘new’ random experiment contained in the random experiment of tossing a dice (see Exercise 1-3). *Third*, using different  $\sigma$ -algebras is indispensable for introducing conditional expectations, conditional independence, and conditional distributions (see chs. 9 to 17).  $\triangleleft$

**Example 1.9 (Joe and Ann)** Consider the following random experiment: *First*, we sample a unit  $u$  from the set  $\Omega_U := \{Joe, Ann\}$ . *Second*, each unit receives (*yes*) or does not receive a treatment (*no*). *Third*, it is observed whether (+) or not (−) a success criterion is reached (see Fig. 1.2). Defining  $\Omega_X := \{yes, no\}$  and  $\Omega_Y := \{+, -\}$ , the Cartesian product

$$\Omega := \Omega_U \times \Omega_X \times \Omega_Y = \{ (Joe, no, -), (Joe, no, +), \dots, (Ann, yes, +) \}$$

is the set of possible outcomes  $\omega$  of this random experiment. It has eight elements, namely the triples  $(Joe, no, -)$ ,  $(Joe, no, +)$ ,  $\dots$ ,  $(Ann, yes, +)$  (see all eight leaves of Fig. 1.2 for a complete list of these elements).

In this example, a first  $\sigma$ -algebra  $\mathcal{A}$  we may consider is *the set of all subsets of*  $\Omega$ , the power set  $\mathcal{P}(\Omega)$ . This set has  $2^8 = 256$  elements, where 8 is the number of elements, i. e., the cardinality of  $\Omega$  (see Kheyfits, 2010, Th. 1.1.37, p. 22). Among these elements is the set

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

In the context of probability theory it is also called the event that *Joe is drawn*. Other elements of  $\mathcal{A}$  are the sets

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

that the *drawn person is treated*, and

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

that  $\{+\}$  (*success*) occurs, irrespective of which person is drawn and whether or not the person is treated.

Aside from the power set of  $\Omega$  we could also consider the  $\sigma$ -algebras  $\mathcal{A}_1 := \{\Omega, \emptyset, A, A^c\}$ ,  $\mathcal{A}_2 := \{\Omega, \emptyset, B, B^c\}$ , and  $\mathcal{A}_3 := \{\Omega, \emptyset, C, C^c\}$ , to name just three. (For another one see Exercise 1-4). In a sense,  $\mathcal{A}_1$  represents the information which person is drawn. In contrast,  $\mathcal{A}_2$  contains the information whether or not the drawn person is treated, and  $\mathcal{A}_3$  whether or not the drawn person is successful. Of course, all these  $\sigma$ -algebras are subsets of  $\mathcal{P}(\Omega)$ , the power set of  $\Omega$ .  $\triangleleft$

**Example 1.10 (Trace of a Set System and Trace  $\sigma$ -Algebra)** If  $\mathcal{E}$  is a set system on  $\Omega$  and  $\Omega_0 \subset \Omega$ , then

$$\mathcal{E}|_{\Omega_0} := \{\Omega_0 \cap A : A \in \mathcal{E}\}$$

is a set system on  $\Omega_0$ . It is called the *trace of  $\mathcal{E}$  in  $\Omega_0$* . Furthermore, if  $\mathcal{A}$  is a  $\sigma$ -algebra on  $\Omega$  and  $\Omega_0 \subset \Omega$ , then the set system

$$\mathcal{A}|_{\Omega_0} := \{\Omega_0 \cap A : A \in \mathcal{A}\}$$

is a  $\sigma$ -algebra on  $\Omega_0$  (see Exercise 1-5). If  $\Omega \neq \Omega_0$ , then the trace  $\mathcal{A}|_{\Omega_0}$  is a  $\sigma$ -algebra on  $\Omega_0$ , but not on  $\Omega$ , because  $\Omega \notin \mathcal{A}|_{\Omega_0}$ .  $\triangleleft$

**Example 1.11 (Joe and Ann – continued)** In Example 1.9 we defined the event  $A$  that Joe is drawn, the event  $B$  that the drawn person is treated, and the  $\sigma$ -algebra  $\mathcal{A}_2 = \{\Omega, \emptyset, B, B^c\}$ . The trace of  $\mathcal{A}_2$  in  $A$  is

$$\mathcal{A}_2|_A = \{A, \emptyset, A \cap B, A \cap B^c\}.$$

Obviously, just like all elements of  $\mathcal{A}_2$  are subsets of  $\Omega$ , all elements of  $\mathcal{A}_2|_A$  are subsets of  $A$ . From an application point of view, considering  $\mathcal{A}_2|_A$  means to presume that Joe is drawn and consider the events that he is treated or not treated, respectively.  $\triangleleft$

### 1.2.1 $\sigma$ -Algebra Generated by a Set System

The concept of a  $\sigma$ -algebra generated by a set system is useful in order to define important  $\sigma$ -algebras. It is also useful for specifying certain measures (see section 1.6). The following theorem prepares Definition 1.13. Reading this theorem, remember that a  $\sigma$ -algebra on a set  $\Omega$  is itself a set (of subsets of  $\Omega$ ), so that we can consider the intersection of  $\sigma$ -algebras.

**Theorem 1.12 (Intersection of  $\sigma$ -Algebras is a  $\sigma$ -Algebra)**

Let  $I$  be a nonempty (finite, countable, or uncountable) index set and let all  $\mathcal{A}_i, i \in I$ , be  $\sigma$ -algebras on  $\Omega$ . Then  $\bigcap_{i \in I} \mathcal{A}_i$  is also a  $\sigma$ -algebra on  $\Omega$ .

(Proof p. 28)

This theorem allows us to define the  $\sigma$ -algebra generated by a set system on  $\Omega$ .

**Definition 1.13 ( $\sigma$ -Algebra Generated by a Set System)**

Let  $\mathcal{E}$  be a set system on  $\Omega$  and let  $(\mathcal{A}_i, i \in I)$  be the family of all  $\sigma$ -algebras on  $\Omega$  that contain  $\mathcal{E}$  as a subset. Then we define

$$\sigma(\mathcal{E}) := \bigcap_{i \in I} \mathcal{A}_i \quad (1.10)$$

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

**Remark 1.14 (Smallest  $\sigma$ -Algebra Containing  $\mathcal{E}$  as a Subset)** According to Theorem 1.12, every set system  $\mathcal{E}$  on  $\Omega$  generates a uniquely defined  $\sigma$ -algebra  $\sigma(\mathcal{E})$  on  $\Omega$ . Note that the  $\sigma$ -algebra  $\sigma(\mathcal{E})$  is the *smallest*  $\sigma$ -algebra on  $\Omega$  containing  $\mathcal{E}$  as a subset, i. e.,

$$\mathcal{C} \text{ is a } \sigma\text{-algebra on } \Omega \text{ and } \mathcal{E} \subset \mathcal{C} \Rightarrow \sigma(\mathcal{E}) \subset \mathcal{C}. \quad (1.11)$$

Furthermore,

$$\sigma[\sigma(\mathcal{E})] = \sigma(\mathcal{E}). \quad (1.12)$$

◁

The following lemma immediately follows from (1.11). It can be used in proofs of the identity of two  $\sigma$ -algebras.

**Lemma 1.15 (Smallest  $\sigma$ -Algebra Containing  $\mathcal{E}$  as a Subset)**

Let  $(\Omega, \mathcal{A})$  be a measurable space and  $\mathcal{E}$  a set system on  $\Omega$  with  $\sigma(\mathcal{E}) = \mathcal{A}$ . If  $\mathcal{C}$  is a  $\sigma$ -algebra on  $\Omega$  with  $\mathcal{E} \subset \mathcal{C} \subset \mathcal{A}$ , then  $\mathcal{C} = \mathcal{A}$ .

(Proof p. 29)

**Remark 1.16 ( $\sigma$ -Algebra Generated by Unions of Set Systems)** Let  $\mathcal{D}, \mathcal{E}, \mathcal{F}$  be set systems on a nonempty set  $\Omega$ . Then

$$\sigma(\mathcal{D} \cup \mathcal{E} \cup \mathcal{F}) = \sigma[\mathcal{D} \cup \sigma(\mathcal{E} \cup \mathcal{F})] \quad (1.13)$$

(see Exercise 1-6). ◁

**Example 1.17 (Several Set Systems May Generate the Same  $\sigma$ -Algebra)** If  $A$  is a subset of  $\Omega$ , then the set system  $\{A\}$  generates the  $\sigma$ -algebra  $\{\Omega, \emptyset, A, A^c\}$ . Note that  $\{\Omega, \emptyset, A, A^c\}$  is also generated by the set systems  $\{A^c\}$  and  $\{A, A^c\}$ , for instance. Hence,

$$\sigma(\{A\}) = \sigma(\{A^c\}) = \sigma(\{A, A^c\}) = \sigma(\{\Omega, \emptyset, A, A^c\}) = \{\Omega, \emptyset, A, A^c\}.$$

In contrast, if  $\emptyset \neq A \neq \Omega$ , then the  $\sigma$ -algebra  $\{\Omega, \emptyset, A, A^c\}$  is neither generated by the set system  $\{A\}$  nor by  $\{\Omega, \emptyset\}$ . Instead,

$$\sigma(\{\emptyset\}) = \sigma(\{\Omega\}) = \sigma(\{\Omega, \emptyset\}) = \{\Omega, \emptyset\},$$

i. e.,  $\{\Omega\}$ ,  $\{\emptyset\}$ , and  $\{\Omega, \emptyset\}$  generate the  $\sigma$ -algebra  $\{\Omega, \emptyset\}$ . ◁

**Example 1.18 (A Generator of the Power Set)** Let  $\Omega$  be finite or countable and let  $\mathcal{E} := \{\{\omega\} : \omega \in \Omega\}$ . Then  $\sigma(\mathcal{E}) = \mathcal{P}(\Omega)$  (see Exercise 1-7). ◁

This example is generalized in Lemma 1.20.

**Remark 1.19 (Partition)** Reading the following lemma, remember that a set system  $\mathcal{E}$  on  $\Omega$  is called a *partition* of  $\Omega$  if

- (a)  $\forall B \in \mathcal{E} : B \neq \emptyset$ .
  - (b)  $\forall B, C \in \mathcal{E} : B \neq C \Rightarrow B \cap C = \emptyset$ .
  - (c)  $\bigcup_{B \in \mathcal{E}} B = \Omega$ .
- ◁

**Lemma 1.20 (An Element of a  $\sigma$ -Algebra Generated by a Partition)**

Let  $\mathcal{E} := \{B_1, \dots, B_n\}$  or  $\mathcal{E} := \{B_1, B_2, \dots\}$  be a finite or countable partition of  $\Omega$ , respectively. Then for all  $C \in \sigma(\mathcal{E})$  there is an  $I(C) \subset \mathbb{N}$  such that

$$C = \bigcup_{i \in I(C)} B_i = \bigcup_{B_i \subset C} B_i, \quad (1.14)$$

where, by convention,  $\bigcup_{i \in \emptyset} B_i := \emptyset$ .

(Proof p. 29)

**Remark 1.21 (Constructing a  $\sigma$ -Algebra)** If  $\mathcal{E} = \{A_1, \dots, A_m\}$  is a finite set of subsets of  $\Omega$ , then there is a finite partition  $\mathcal{F} = \{B_1, \dots, B_n\}$  of  $\Omega$  with  $\sigma(\mathcal{E}) = \sigma(\mathcal{F})$ . Furthermore, if  $\mathcal{E}$  is a *finite* set of subsets of  $\Omega$ , then each element of  $\sigma(\mathcal{E})$  is obtained by finitely many unions, intersections, or complements of elements of  $\mathcal{E}$  (see Exercise 1-8). ◁

**Example 1.22 (Joe and Ann – continued)** In Example 1.11, we already considered the event  $A$  that Joe is drawn and noted that the trace of the  $\sigma$ -algebra  $\mathcal{A}_2 = \{\Omega, \emptyset, B, B^c\}$  in  $A$  is  $\mathcal{A}_2|_A = \{A, \emptyset, A \cap B, A \cap B^c\}$ . In contrast, the  $\sigma$ -algebra on  $\Omega$  generated by the trace  $\mathcal{A}_2|_A$  is

$$\sigma(\mathcal{A}_2|_A) = \{\Omega, \emptyset, A, A^c, A \cap B, A \cap B^c, (A \cap B) \cup A^c, (A \cap B^c) \cup A^c\},$$

where  $(A \cap B) \cup A^c = A^c \cup B$  and  $(A \cap B^c) \cup A^c = A^c \cup B^c$ .  $\triangleleft$

**Remark 1.23 (Monotonicity of Generated  $\sigma$ -Algebras)**

Let  $\mathcal{E}_1, \mathcal{E}_2$  be set systems on  $\Omega$  with  $\mathcal{E}_1 \subset \mathcal{E}_2$ . Then  $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$  (see Exercise 1-9).  $\triangleleft$

An important kind of  $\sigma$ -algebras are those for which there is a countable set system that generates them.

**Definition 1.24 (Countably Generated  $\sigma$ -Algebra)**

Let  $(\Omega, \mathcal{A})$  be a measurable space. Then  $\mathcal{A}$  is called countably generated if there is a finite or countable set  $\mathcal{E} \subset \mathcal{A}$  such that  $\sigma(\mathcal{E}) = \mathcal{A}$ .

**Example 1.25 (Some Countably Generated  $\sigma$ -Algebras)** Examples of countably generated  $\sigma$ -algebras are:

- (a) All  $\sigma$ -algebras on a finite set  $\Omega$ .
- (b)  $\mathcal{P}(\mathbb{N}_0^n)$ ,  $n \in \mathbb{N}$ .

(For a proof see Exercise 1-10). For another example, see Remark 1.28.  $\triangleleft$

**Remark 1.26 (A Caveat)** Note that there are countably generated  $\sigma$ -algebras for which not all of their elements can be constructed by countably many unions, intersections, or complements of elements of the generating system. An example in case are Borel  $\sigma$ -algebras on  $\mathbb{R}$  or  $\mathbb{R}^n$  (see Rem. 1.28 and Michel, 1978, sect. I.4).  $\triangleleft$

**Lemma 1.27 ( $\sigma$ -Algebra Generated by the Trace of a Set System)**

Let  $A \subset \Omega$  be nonempty,  $\mathcal{E} \subset \mathcal{P}(\Omega)$ , and  $\mathcal{E}|_A := \{C \cap A : C \in \mathcal{E}\}$ . Then

$$\sigma(\mathcal{E}|_A) = \sigma(\mathcal{E})|_A, \quad (1.15)$$

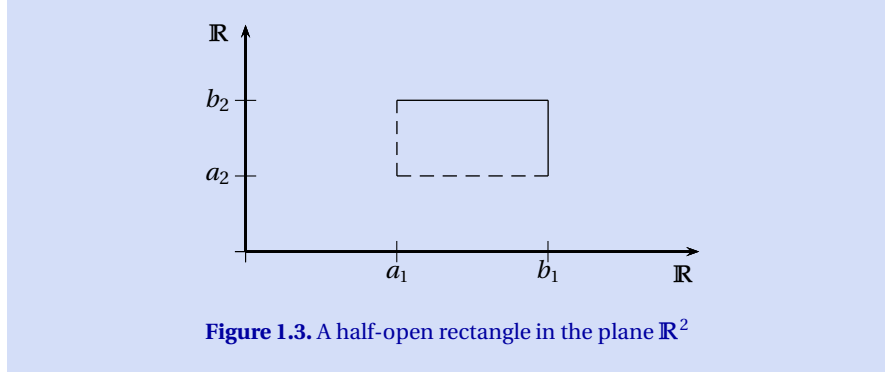
where  $\sigma(\mathcal{E}|_A)$  denotes the  $\sigma$ -algebra generated on  $A$ , whereas  $\sigma(\mathcal{E})$  is a  $\sigma$ -algebra on  $\Omega$ . Furthermore, if  $\mathcal{C}$  is a  $\sigma$ -algebra on  $\Omega$  and  $A \in \mathcal{C}$  such that

$$\forall C \in \mathcal{C}: C \neq A \Rightarrow A \cap C = \emptyset, \quad (1.16)$$

i. e.,  $A$  does not intersect with any other element of  $\mathcal{C}$ , then

$$\sigma(\mathcal{C} \cup \mathcal{E})|_A = \mathcal{C}|_A. \quad (1.17)$$

(Proof p. 30)



Hence, according to Equation (1.15), the  $\sigma$ -algebra generated by the trace of a set system  $\mathcal{E}$  is the trace of the  $\sigma$ -algebra generated by  $\mathcal{E}$  and according to Equation (1.17), the trace of the  $\sigma$ -algebra  $\sigma(\mathcal{C} \cup \mathcal{E})$  in the set  $A$  is identical to the trace of the  $\sigma$ -algebra  $\mathcal{C}$  in  $A$ , if 1.16 holds.

### 1.2.2 $\sigma$ -Algebra of Borel Sets on $\mathbb{R}^n$

For  $a, b \in \mathbb{R}$  with  $a < b$ , let us consider a *half-open interval*  $]a, b]$  in  $\mathbb{R}$ , which is defined by

$$]a, b] := \{x \in \mathbb{R} : a < x \leq b\},$$

and the *set system*

$$\mathcal{I}_1 := \{]a, b] : a, b \in \mathbb{R} \text{ and } a < b\}$$

of all half-open intervals in  $\mathbb{R}$ . The  $\sigma$ -algebra generated by this set system is called the *Borel  $\sigma$ -algebra* on  $\mathbb{R}$ . It is denoted by  $\mathcal{B}$ . The elements of  $\mathcal{B}$  are called the *Borel sets* of  $\mathbb{R}$ . In formal terms,

$$\mathcal{B} := \mathcal{B}_1 := \sigma(\mathcal{I}_1). \quad (1.18)$$

Note that there are several sets systems generating the Borel  $\sigma$ -algebra (see, e.g., Klenke, 2013, Th. 1.23, p. 9). In particular,

$$\mathcal{B}_1 = \sigma(\{]-\infty, b] : b \in \mathbb{R}\}) \quad (1.19)$$

(see Georgii, 2008, p. 12). Similarly, we define the *Borel  $\sigma$ -algebra on  $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$*  to be the  $\sigma$ -algebra generated by the set system  $\mathcal{I}_2$  of all half-open *rectangles* in  $\mathbb{R}^2$ , whose sides are parallel to the axes (see Fig. 1.3). These rectangles are defined by

$$]a_1, b_1] \times ]a_2, b_2] = \{(x_1, x_2) \in \mathbb{R}^2 : a_1 < x_1 \leq b_1, a_2 < x_2 \leq b_2\}.$$

The  $\sigma$ -algebra  $\sigma(\mathcal{I}_2)$  is denoted by  $\mathcal{B}_2$ , i. e.,  $\mathcal{B}_2 := \sigma(\mathcal{I}_2)$ , and its elements are called the *Borel sets of  $\mathbb{R}^2$* .

This definition is easily generalized: The *Borel  $\sigma$ -algebra on  $\mathbb{R}^n$*  is defined by  $\mathcal{B}_n := \sigma(\mathcal{I}_n)$ ,  $n \in \mathbb{N}$ , where  $\mathcal{I}_n$  is the system of all half-open *cuboids* in  $\mathbb{R}^n$ , whose sides are parallel to the axes. Such a cuboid is a set

$$\begin{aligned} & ] a_1, b_1] \times \dots \times ] a_n, b_n] \\ & = \{(x_1, \dots, x_n) \in \mathbb{R}^n : a_1 < x_1 \leq b_1, \dots, a_n < x_n \leq b_n\}, \end{aligned} \quad (1.20)$$

where  $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$ . Just like  $\mathcal{B}_1$ , the  $\sigma$ -algebra  $\mathcal{B}_n$  has several generating systems, one of which is

$$\mathcal{B}_n = \sigma(\{ ] -\infty, b_1] \times \dots \times ] -\infty, b_n] : b_1, \dots, b_n \in \mathbb{R} \}) \quad (1.21)$$

(see Exercise 1-11).

Note that not every subset of  $\mathbb{R}^n$  is a Borel set. In other words,  $\mathcal{B}_n$  is not the power set of  $\mathbb{R}^n$  (see Rem. 1.60). However, for each  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ , the singleton  $\{x\}$  is a Borel set of  $\mathbb{R}^n$ , i. e.,

$$\{x\} \in \mathcal{B}_n, \quad \forall x \in \mathbb{R}^n \quad (1.22)$$

(see Exercise 1-12).

Furthermore, if  $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$  denotes the *extended set of real numbers*, then

$$\bar{\mathcal{B}} := \sigma(\mathcal{B} \cup \{-\infty, +\infty\})$$

is a  $\sigma$ -algebra on  $\bar{\mathbb{R}}$  and it is called the *Borel  $\sigma$ -algebra on  $\bar{\mathbb{R}}$* . Similarly,  $\bar{\mathcal{B}}_n$  is called the *Borel  $\sigma$ -algebra on  $\bar{\mathbb{R}}^n$* . It is defined as the product of the  $\sigma$ -algebra  $\mathcal{B}$  with itself ( $n$  times) (see Def. 1.31). Finally, we may sometimes consider  $\bar{\mathcal{B}}_n|_{\Omega_0}$ , the *trace of the Borel  $\sigma$ -algebra on  $\bar{\mathbb{R}}^n$  in  $\Omega_0 \subset \bar{\mathbb{R}}^n$* .

**Remark 1.28 (The Borel  $\sigma$ -Algebra is Countably Generated)** Note that

$$\mathcal{B} = \sigma(\{ ] a, b] : a, b \in \mathbb{Q}, a < b \}),$$

where  $\mathbb{Q}$  is the set of rational numbers. Because  $\mathbb{Q}$  is countable, the set of intervals  $\{ ] a, b] : a, b \in \mathbb{Q}, a < b \}$  is countable as well. Therefore, the Borel  $\sigma$ -algebra  $\mathcal{B}$  is countably generated. This also holds for  $\mathcal{B}_n$ ,  $n \in \mathbb{N}$  (see Klenke, 2013, Th. 1.23, p. 9).  $\triangleleft$

**Remark 1.29 (Trace of the Borel  $\sigma$ -Algebra in a Countable Subset of  $\mathbb{R}$ )** Let  $\mathcal{B}$  denote the Borel  $\sigma$ -algebra on  $\mathbb{R}$ . If  $\Omega_0 \subset \mathbb{R}$  is finite or countable, then  $\mathcal{B}|_{\Omega_0} = \mathcal{P}(\Omega_0)$ , where  $\mathcal{B}|_{\Omega_0}$  is the trace of the Borel  $\sigma$ -algebra on  $\mathbb{R}$  in  $\Omega_0 \subset \mathbb{R}$  (see Exercise 1-13).  $\triangleleft$

### 1.2.3 $\sigma$ -Algebra on a Cartesian Product

In section 1.2.2 we defined a  $\sigma$ -algebra on  $\mathbb{R}^n = \mathbb{R} \times \dots \times \mathbb{R}$  ( $n$ -times). Now we consider  $\sigma$ -algebras on general Cartesian products. We start with an example.

**Example 1.30 (Joe and Ann – continued)** In Example 1.9 we already considered the Cartesian product

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

which consists of the eight triples  $(Joe, no, -), (Joe, no, +), \dots, (Ann, yes, +)$  (see again Fig. 1.2). Now consider the  $\sigma$ -algebras  $\mathcal{A}_1 := \mathcal{P}(\Omega_U)$ ,  $\mathcal{A}_2 := \mathcal{P}(\Omega_X)$ , and  $\mathcal{A}_3 := \mathcal{P}(\Omega_Y)$ , as well as the set

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

which is a set system on  $\Omega$  consisting of  $4 \cdot 4 \cdot 4 = 64$  elements. For example, the set system  $\mathcal{E}$  contains the elements

$$A := \{Joe\} \times \{no\} \times \{-\} = \{(Joe, no, -)\}$$

and

$$B := \{Ann\} \times \{yes\} \times \{+\} = \{(Ann, yes, +)\}.$$

However,  $\mathcal{E}$  does not contain

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

as an element. The only product set  $A_1 \times A_2 \times A_3$  with  $A_1 \in \mathcal{A}_1$ ,  $A_2 \in \mathcal{A}_2$ ,  $A_3 \in \mathcal{A}_3$  that contains  $A \cup B$  as a subset is  $\Omega_U \times \Omega_X \times \Omega_Y = \Omega$ . However,  $A \cup B \neq \Omega$ . Therefore,  $\mathcal{E}$  is not a  $\sigma$ -algebra [cf. condition (c) of Rem. 1.2]. In this example, the  $\sigma$ -algebra generated by  $\mathcal{E}$  is the power set of  $\Omega$ , i. e.,  $\sigma(\mathcal{E}) = \mathcal{P}(\Omega)$ . It consists of  $2^8 = 256$  elements. According to the following definition,  $\sigma(\mathcal{E})$  is denoted by  $\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \mathcal{A}_3$  and called the *product  $\sigma$ -algebra of  $\mathcal{A}_1$ ,  $\mathcal{A}_2$ , and  $\mathcal{A}_3$* .  $\triangleleft$

**Definition 1.31 (Product  $\sigma$ -Algebra)**

Let  $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_n, \mathcal{A}_n)$  be measurable spaces and  $\Omega := \Omega_1 \times \dots \times \Omega_n$ . Then

$$\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n := \bigotimes_{i=1}^n \mathcal{A}_i := \sigma\left(\left\{\prod_{i=1}^n A_i : A_i \in \mathcal{A}_i, i = 1, \dots, n\right\}\right) \quad (1.23)$$

is called the *product  $\sigma$ -algebra of the  $\sigma$ -algebras  $\mathcal{A}_i, i = 1, \dots, n$* .

To emphasize, the product  $\sigma$ -algebra  $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$  is *not* the Cartesian product  $\mathcal{A}_1 \times \dots \times \mathcal{A}_n$ . Instead, the product  $\sigma$ -algebra is generated by the set system of all Cartesian products of elements of the  $\sigma$ -algebras  $\mathcal{A}_1, \dots, \mathcal{A}_n$ . In Lemma 2.42 we give an equivalent specification of a product  $\sigma$ -algebra, using projection mappings.

The following lemma provides a relationship between the generating systems of the  $\sigma$ -algebras  $\mathcal{A}_i, i = 1, \dots, n$ , and the generating system of the product  $\sigma$ -algebra.

**Lemma 1.32 (Generating System of a Product  $\sigma$ -Algebra)**

For  $i = 1, \dots, n$ , let  $(\Omega_i, \mathcal{A}_i)$  be measurable spaces and  $\mathcal{E}_i \subset \mathcal{A}_i$  with  $\sigma(\mathcal{E}_i) = \mathcal{A}_i$ .

Then

$$\bigotimes_{i=1}^n \mathcal{A}_i = \sigma\left(\left\{\prod_{i=1}^n A_i : A_i \in \mathcal{E}_i, i = 1, \dots, n\right\}\right). \quad (1.24)$$

For a proof see Klenke (2013, Th. 14.12 (i), p. 276).

This lemma implies

$$\mathcal{B}_n = \bigotimes_{i=1}^n \mathcal{B} = \mathcal{B} \otimes \dots \otimes \mathcal{B} \quad (n\text{-times})$$

for the Borel  $\sigma$ -algebra on  $\mathbb{R}^n$ . This lemma also implies the following corollary.

**Corollary 1.33 (Countable Generating System of a Product  $\sigma$ -Algebra)**

Let  $(\Omega_i, \mathcal{A}_i)$ ,  $i = 1, \dots, n$ , be measurable spaces, where all  $\mathcal{A}_i$  are countably generated. Then  $\bigotimes_{i=1}^n \mathcal{A}_i$  is countably generated as well.

**Example 1.34 (Countable Sets and Product  $\sigma$ -Algebra)** Let  $\Omega_1, \dots, \Omega_n$  be finite or countable sets and  $\mathcal{A}_1, \dots, \mathcal{A}_n$  be their power sets. Then

$$\bigotimes_{i=1}^n \mathcal{A}_i = \mathcal{P}\left(\prod_{i=1}^n \Omega_i\right),$$

i. e.,  $\bigotimes_{i=1}^n \mathcal{A}_i$  is the power set on  $\Omega := \Omega_1 \times \dots \times \Omega_n$  (see Exercise 1-14).  $\triangleleft$

**Remark 1.35 (Complement of a Cartesian Product)** Let  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$  be a measurable space,  $A \in \mathcal{A}_1$ , and  $B \in \mathcal{A}_2$ . Then  $(A \times B)^c \in \mathcal{A}_1 \otimes \mathcal{A}_2$ , and this set can be written

$$(A \times B)^c = (A^c \times B) \cup (\Omega_1 \times B^c), \quad (1.25)$$

which is a union of disjoint sets (see Exercise 1-15).  $\triangleleft$

**1.2.4  $\cap$ -Stable Set Systems That Generate a  $\sigma$ -Algebra**

For many proofs, generating set systems are useful, which have the property of  $\cap$ -stability.

**Definition 1.36 ( $\cap$ -Stability)**

Let  $\Omega$  denote a nonempty set. A set  $\mathcal{E}$  of subsets of  $\Omega$  is called  $\cap$ -stable (or  $\cap$ -closed) if  $A \cap B \in \mathcal{E}$  for all  $A, B \in \mathcal{E}$ .

**Example 1.37 (Set System With One Single Element)** A set system  $\{A\}$  that has only a single element  $A \subset \Omega$  is  $\cap$ -stable (cf. Example 1.17).  $\triangleleft$

**Example 1.38 (Partition and  $\cap$ -Stability)** If  $\mathcal{E}$  is a partition of the set  $\Omega$ , then  $\mathcal{D} := \mathcal{E} \cup \{\emptyset\}$  is  $\cap$ -stable.  $\triangleleft$

**Example 1.39 (A  $\cap$ -Stable Generating System of a Product  $\sigma$ -Algebra)** Consider the measurable spaces  $(\Omega_i, \mathcal{A}_i)$ ,  $i = 1, \dots, n$ . The set

$$\{A_1 \times \dots \times A_n : A_i \in \mathcal{A}_i, i = 1, \dots, n\},$$

is a  $\cap$ -stable generating system of  $\bigotimes_{i=1}^n \mathcal{A}_i$  (see Exercise 1-16).  $\triangleleft$

Another type of a set system is a Dynkin system. It can be used in order to show that a specific set system is a  $\sigma$ -algebra.

**Definition 1.40 (Dynkin System)**

A set  $\mathcal{D}$  of subsets of a set  $\Omega$  is called a Dynkin system on  $\Omega$ , if the following three conditions hold:

- (a)  $\Omega \in \mathcal{D}$ .
- (b) If  $A \in \mathcal{D}$ , then  $A^c \in \mathcal{D}$ .
- (c) If  $A_1, A_2, \dots \in \mathcal{D}$  and they are pairwise disjoint, then  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{D}$ .

In the definition of a  $\sigma$ -algebra  $\mathcal{A}$  we require  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$  for all sequences  $A_1, A_2, \dots \in \mathcal{A}$ , whereas for a Dynkin system the corresponding requirement is only made for all sequences  $A_1, A_2, \dots \in \mathcal{D}$  of pairwise disjoint sets. Analogously to Definition 1.13, for a set system  $\mathcal{E}$  on  $\Omega$ ,  $\delta(\mathcal{E})$  is defined as the Dynkin system generated by  $\mathcal{E}$ , i. e., as the intersection of all Dynkin systems containing  $\mathcal{E}$ . According to the following theorem a Dynkin system is also a  $\sigma$ -algebra if and only if it is  $\cap$ -stable.

**Theorem 1.41 (Dynkin System and  $\sigma$ -Algebra)**

Let  $\Omega$  be a nonempty set.

- (i) A Dynkin system  $\mathcal{D}$  on  $\Omega$  is a  $\sigma$ -algebra if and only if it is  $\cap$ -stable.
- (ii) If  $\mathcal{E}$  is a  $\cap$ -stable set of subsets of  $\Omega$ , then  $\delta(\mathcal{E}) = \sigma(\mathcal{E})$ .

For a proof see Bauer (2001, Ths. 2.3 and 2.4, p. 6, 7). According to proposition (i) of this theorem we can prove that a set system is a  $\sigma$ -algebra by showing that it is a  $\cap$ -stable Dynkin system, and proposition (ii) can be applied to show that the Dynkin system generated by a  $\cap$ -stable set system is a  $\sigma$ -algebra.

### 1.3 Measure and Measure Space

A measure assigns to all elements of a  $\sigma$ -algebra an element of the closed interval

$$[0, \infty] := \{x \in \mathbb{R} : 0 \leq x\} \cup \{\infty\},$$

i. e., a nonnegative real number or the element  $\infty$ .

**Example 1.42 (A First Example)** Let  $\Omega = \mathbb{R}$  and assume that the closed interval  $[3, 9] = \{x \in \mathbb{R} : 3 \leq x \leq 9\}$  as well as the union  $[3, 9] \cup [10, 12]$  are elements of a  $\sigma$ -algebra on  $\Omega$ . If the measure is *length*, then

$$\text{length}([3, 9]) = 9 - 3 = 6$$

and

$$\begin{aligned} \text{length}([3, 9] \cup [10, 12]) &= \text{length}([3, 9]) + \text{length}([10, 12]) \\ &= (9 - 3) + (12 - 10) = 6 + 2 = 8, \end{aligned}$$

because the two intervals are disjoint, i. e., their intersection is the empty set  $\emptyset$ . In this case the lengths of the intervals  $[3, 9]$  and  $[10, 12]$  add up to the length of their union  $[3, 9] \cup [10, 12]$ . In Definition 1.43 (c) we do not only require additivity but  $\sigma$ -additivity.  $\triangleleft$

Reading the following definition, remember that  $\sum_{i=1}^{\infty} a_i$  is defined by

$$\sum_{i=1}^{\infty} a_i := \lim_{n \rightarrow \infty} \sum_{i=1}^n a_i.$$

**Definition 1.43 (Measure and Measure Space)**

Let  $(\Omega, \mathcal{A})$  be a measurable space. A function  $\mu: \mathcal{A} \rightarrow \bar{\mathbb{R}}$  is called a *measure* and the triple  $(\Omega, \mathcal{A}, \mu)$  is called a *measure space*, if

- (a)  $\mu(\emptyset) = 0$ .
- (b)  $\mu(A) \geq 0, \forall A \in \mathcal{A}$ . (*nonnegativity*)
- (c) If  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint, then  $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ . ( *$\sigma$ -additivity*)

### 1.3.1 $\sigma$ -Additivity and Related Properties

**Remark 1.44 ( $\sigma$ -Additivity Implies Finite Additivity)** Note that  $\sigma$ -additivity of a measure implies finite additivity, i. e., it implies

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i), \quad \text{if } A_1, \dots, A_n \in \mathcal{A} \text{ are pairwise disjoint} \quad (1.26)$$

[see Rule (ii) of Box 1.1 and its proof in Exercise 1-18].  $\triangleleft$

**Remark 1.45 ( $\sigma$ -Additivity)** Using the term  *$\sigma$ -additivity* signalsizes that unions of finitely or countably many sets are considered, but not other unions of sets. If, instead of  $\sigma$ -additivity, we would require additivity for *any kind of unions*, including uncountable unions, then the Lebesgue measure  $\lambda$  on  $(\mathbb{R}, \mathcal{B})$  — the measure of *length* — could not be constructed any more. This is explained in more detail in Remark 1.71.  $\triangleleft$

**Remark 1.46 (Representation of a Union as a Union of Pairwise Disjoint Sets)**

Let  $(\Omega, \mathcal{A})$  be a measurable space. If  $A_1, A_2, \dots \in \mathcal{A}$  is a sequence of subsets of  $\Omega$ , then there is a sequence  $B_1, B_2, \dots \in \mathcal{A}$  of pairwise disjoint sets with

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i. \quad (1.27)$$

One way to construct  $B_1, B_2, \dots$  is to define  $B_1 := A_1$  and

$$B_i := A_i \setminus \left( \bigcup_{j=1}^{i-1} A_j \right), \quad \text{for } i > 1, \quad (1.28)$$

(see Exercise 1-17). ◁

**Remark 1.47 (Additivity of Measures for Partitions)** Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space,  $B \in \mathcal{A}$ , and assume

- (a)  $A_1, \dots, A_n \in \mathcal{A}$  are pairwise disjoint,
- (b)  $B \subset \bigcup_{i=1}^n A_i$ .

Then

$$\mu(B) = \sum_{i=1}^n \mu(B \cap A_i). \quad (1.29)$$

Analogously, if

- (c)  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint,
- (d)  $B \subset \bigcup_{i=1}^{\infty} A_i$ ,

then

$$\mu(B) = \sum_{i=1}^{\infty} \mu(B \cap A_i). \quad (1.30)$$

(see Exercise 1-19). ◁

**1.3.2 Other Properties**

Other important properties of a measure are displayed in Box 1.1. Some of these properties can intuitively be understood inspecting the Venn diagram presented in Figure 1.1. These properties always hold with the conventions  $+\infty + \infty = +\infty$  and  $\alpha + \infty = +\infty$ , for  $\alpha \in \mathbb{R}$ . However, note that the term  $+\infty - \infty$  cannot meaningfully be defined. Therefore, properties (vi) and (vii) only hold if we assume  $\mu(A \cap B) < \infty$ . For proofs of all these properties see Exercise 1-18.

**Remark 1.48 (Finite Additivity and  $\sigma$ -Additivity Applied to Singletons)** If  $\Omega$  is finite or countable, then each  $A \subset \Omega$  is finite or countable as well. Hence, for any measure  $\mu$  on the measurable space  $(\Omega, \mathcal{P}(\Omega))$ ,

$$\mu(A) = \mu\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} \mu(\{\omega\}), \quad \forall A \subset \Omega. \quad (1.31)$$

**Box 1.1 Rules of Computation for Measures**

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space.

If  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \quad (\sigma\text{-additivity}) \quad (\text{i})$$

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i), \quad \forall n \in \mathbb{N}. \quad (\text{finite additivity}) \quad (\text{ii})$$

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

$$\mu(A) = \mu(A \cap B) + \mu(A \setminus B). \quad (\text{iii})$$

$$\mu(\Omega) = \mu(B) + \mu(B^c). \quad (\text{iv})$$

$$\mu(A) \leq \mu(B), \quad \text{if } A \subset B. \quad (\text{monotonicity}) \quad (\text{v})$$

$$\mu(A \setminus B) = \mu(A) - \mu(A \cap B), \quad \text{if } \mu(A \cap B) < \infty. \quad (\text{vi})$$

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B), \quad \text{if } \mu(A \cap B) < \infty. \quad (\text{vii})$$

$$\mu(A) = \mu(\Omega) < \infty \Rightarrow \mu(A \cap B) = \mu(B). \quad (\text{viii})$$

$$\mu(A) = 0 \Rightarrow \mu(A \cup B) = \mu(B). \quad (\text{ix})$$

Let  $A \in \mathcal{A}$  and let  $\Omega_0 \subset \Omega$  and be finite or countable with  $\mu(\Omega \setminus \Omega_0) = 0$ .

If, for all  $\omega \in \Omega_0$ ,  $\{\omega\} \in \mathcal{A}$ , then

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

If  $A_1, A_2, \dots \in \mathcal{A}$ , then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i). \quad (\sigma\text{-subadditivity}) \quad (\text{xi})$$

This means that a measure on  $(\Omega, \mathcal{P}(\Omega))$  is already uniquely defined if its values  $\mu(\{\omega\})$  are uniquely defined for all  $\omega \in \Omega$ , provided that  $\Omega$  is finite or countable. Rule (x) of Box 1.1 extends this result to a more general measure space  $(\Omega, \mathcal{A}, \mu)$ . This rule shows that a measure on  $(\Omega, \mathcal{A})$  is already uniquely defined if its values  $\mu(\{\omega\})$  are uniquely defined for all  $\omega \in \Omega_0$ , provided that  $\Omega_0$  is finite or countable with  $\mu(\Omega \setminus \Omega_0) = 0$  and  $\{\omega\} \in \mathcal{A}$  for all  $\omega \in \Omega_0$ .  $\triangleleft$

## 1.4 Specific Measures

Now we consider some examples of measures, all of which are used later on in order to introduce still other measures. For some of these examples we use the *indicator* of a set  $A$ .

**Definition 1.49 (Indicator)**

Let  $\Omega$  be a set and  $A \subset \Omega$ . Then the function  $1_A: \Omega \rightarrow \mathbb{R}$  defined by

$$1_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases} \quad (1.32)$$

is called the *indicator* of  $A$ .

**Remark 1.50 (Sums and Products of Indicators)** If  $1_A, 1_B: \Omega \rightarrow \mathbb{R}$  are the indicators of two sets  $A, B \subset \Omega$ , then

$$1_A \cdot 1_B = 1_{A \cap B} \quad (1.33)$$

and

$$1_A + 1_B - 1_{A \cap B} = 1_A + 1_B - 1_A \cdot 1_B = 1_{A \cup B}. \quad (1.34)$$

Equation (1.33) immediately implies

$$1_A + 1_B = 1_{A \cup B}, \quad \text{if } A \cap B = \emptyset. \quad (1.35)$$

More generally, if  $A_1, \dots, A_n$  is a finite sequence of pairwise disjoint subsets of  $\Omega$ , then

$$\sum_{i=1}^n 1_{A_i} = 1_{\bigcup_{i=1}^n A_i}, \quad (1.36)$$

i. e., then the sum of the indicators of the sets  $A_1, \dots, A_n$  is the indicator of the union  $\bigcup_{i=1}^n A_i$ . Finally, if  $A_1, A_2, \dots$  is a sequence of pairwise disjoint subsets of  $\Omega$ , then

$$\sum_{i=1}^{\infty} 1_{A_i} = 1_{\bigcup_{i=1}^{\infty} A_i}. \quad (1.37)$$

◁

**Remark 1.51 (Indicators of Products Sets)** Let  $\Omega_1, \Omega_2$  be nonempty sets,  $A \subset \Omega_1$  and  $B \subset \Omega_2$ . Then

$$1_A(\omega_1) \cdot 1_B(\omega_2) = 1_{A \times B}(\omega_1, \omega_2), \quad \forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2. \quad (1.38)$$

This equation follows from the definitions of the product set and the indicator. ◁

### 1.4.1 Dirac Measure and Counting Measure

**Example 1.52 (Dirac Measure)** Let  $(\Omega, \mathcal{A})$  be a measurable space, let  $\omega \in \Omega$ , and consider the function  $\delta_\omega: \mathcal{A} \rightarrow \{0, 1\}$  defined by

$$\delta_\omega(A) := 1_A(\omega), \quad \forall A \in \mathcal{A}. \quad (1.39)$$

Then  $\delta_\omega$  is a measure on  $(\Omega, \mathcal{A})$  (see Exercise 1-20).  $\triangleleft$

**Definition 1.53 (Dirac Measure)**

The function  $\delta_\omega$  defined by Equation (1.39) is called the *Dirac measure at (point)  $\omega$* .

**Example 1.54 (Counting Measure)** Let  $(\Omega, \mathcal{A})$  be a measurable space and define the function  $\mu_\#: \mathcal{A} \rightarrow \bar{\mathbb{R}}$  by

$$\mu_\#(A) := \begin{cases} \sum_{\omega \in \Omega} 1_A(\omega), & \text{if } A \text{ is finite,} \\ \infty, & \text{if } A \text{ is infinite,} \end{cases} \quad \forall A \in \mathcal{A}. \quad (1.40)$$

Then  $\mu_\#$  is a measure on  $(\Omega, \mathcal{A})$  (see Exercise 1-21).  $\triangleleft$

**Definition 1.55 (Counting Measure)**

The function  $\mu_\#$  defined by Equation (1.40) is called the *counting measure on  $(\Omega, \mathcal{A})$* .

**Remark 1.56 (Cardinality of a Set)** If  $A$  is finite, then  $\mu_\#(A)$  is called the *cardinality* of  $A$ , i. e.,  $\mu_\#(A)$  simply counts the number of elements  $\omega$  of the set  $A$ . Furthermore, for finite or countable  $\Omega$  and  $A \subset \Omega$ ,

$$\mu_\#(A) = \sum_{\omega \in \Omega} 1_A(\omega) = \sum_{\omega \in \Omega} \delta_\omega(A). \quad (1.41)$$

$\triangleleft$

**Example 1.57 (Sum of Dirac Measures)** Let  $(\Omega, \mathcal{A})$  be a measurable space. If  $B \subset \Omega$  is finite or countable and  $\delta_\omega$  is the Dirac measure on  $(\Omega, \mathcal{A})$  at point  $\omega$ , then  $\sum_{\omega \in B} \delta_\omega: \mathcal{A} \rightarrow [0, \infty]$  defined by

$$\left( \sum_{\omega \in B} \delta_\omega \right)(A) := \sum_{\omega \in B} \delta_\omega(A), \quad \forall A \in \mathcal{A}, \quad (1.42)$$

is a measure on  $(\Omega, \mathcal{A})$  (see Exercise 1-22). Hence, if  $\Omega$  itself is finite or countable, then  $\sum_{\omega \in \Omega} \delta_\omega$  is a measure on  $(\Omega, \mathcal{A})$ , and it is identical to the counting measure defined in Example 1.54, because, for  $A \in \mathcal{A}$ ,

$$\begin{aligned}
\left( \sum_{\omega \in \Omega} \delta_{\omega} \right) (A) &= \sum_{\omega \in \Omega} \delta_{\omega}(A) && [(1.42)] \\
&= \sum_{\omega \in \Omega} 1_A(\omega) && [(1.39)] \\
&= \mu_{\#}(A). && [(1.41)]
\end{aligned} \tag{1.43}$$

&lt;

### 1.4.2 Lebesgue Measure

Consider the *half-open interval*  $]a, b]$ . Then

$$\lambda_1(]a, b]) = b - a \tag{1.44}$$

is the *length* of the interval  $]a, b]$ . Next consider a *rectangle*  $]a_1, b_1] \times ]a_2, b_2]$  in  $\mathbb{R}^2$  with  $a_1 < b_1$  and  $a_2 < b_2$ . This set can be visualized by the set of all points inside the rectangle presented in Figure 1.3 (excluding the lower and left boundary). Obviously,

$$\lambda_2(]a_1, b_1] \times ]a_2, b_2]) = (b_1 - a_1) \cdot (b_2 - a_2) \tag{1.45}$$

is the *area* of this rectangle.

According to the following theorem, there is one and only one measure on  $(\mathbb{R}, \mathcal{B})$  satisfying (1.44) for all such intervals. This measure is called the *Lebesgue measure on  $(\mathbb{R}, \mathcal{B})$*  and is denoted by  $\lambda$  or  $\lambda_1$ . Similarly, there is one and only one measure on  $(\mathbb{R}^2, \mathcal{B}_2)$  satisfying (1.45) for all such rectangles. It is called the *Lebesgue measure on  $(\mathbb{R}^2, \mathcal{B}_2)$*  and is denoted by  $\lambda_2$ . The following theorem deals with the general case.

**Theorem 1.58 (Existence and Uniqueness of the Lebesgue Measure)**

For all  $n \in \mathbb{N}$ , there is a uniquely defined measure  $\lambda_n$  on  $(\mathbb{R}^n, \mathcal{B}_n)$  satisfying

$$\begin{aligned}
\lambda_n(]a_1, b_1] \times \dots \times ]a_n, b_n]) &= \prod_{i=1}^n (b_i - a_i), \\
\forall a_i, b_i \in \mathbb{R} \text{ with } a_i < b_i, \quad i &= 1, \dots, n.
\end{aligned} \tag{1.46}$$

For a proof see Klenke (2013, Th. 1.55, p. 24).

**Definition 1.59 (Lebesgue Measure)**

The measure  $\lambda_n$  satisfying Equation (1.46) is called the *Lebesgue measure on  $(\mathbb{R}^n, \mathcal{B}_n)$* .

**Remark 1.60 (Sets of Real Numbers That are not Lebesgue Measurable)** Hence, the Lebesgue measure  $\lambda_n$  is defined on  $(\mathbb{R}^n, \mathcal{B}_n)$ . Note, however, that this measure

space  $(\mathbb{R}^n, \mathcal{B}_n, \lambda_n)$  can be completed by extending the Borel  $\sigma$ -algebra by including all the subsets of sets  $A \in \mathcal{B}_n$  with  $\lambda_n(A) = 0$ . In Wise and Hall (1993, counterexample 1.25) it is shown for  $n = 1$  that there are subsets  $B \subset \mathbb{R}$  that are not elements of the completed  $\sigma$ -algebra. Therefore,  $B \notin \mathcal{B}$ , and this implies  $\mathcal{B} \neq \mathcal{P}(\mathbb{R})$ .  $\triangleleft$

### 1.4.3 Other Examples of a Measure

**Example 1.61 (Restriction of a Measure to a Sub- $\sigma$ -Algebra)** Suppose  $(\Omega, \mathcal{A}, \mu)$  is a measure space and  $\mathcal{C} \subset \mathcal{A}$  a  $\sigma$ -algebra. Then the function  $\nu: \mathcal{C} \rightarrow \bar{\mathbb{R}}$  defined by

$$\nu(A) := \mu(A), \quad \forall A \in \mathcal{C}, \quad (1.47)$$

is a measure on  $(\Omega, \mathcal{C})$  (see Exercise 1-23).  $\triangleleft$

**Example 1.62 (Weighted Sum of Measures)** If  $\mu_1, \mu_2, \dots$  are measures on  $(\Omega, \mathcal{A})$  and  $0 \leq \alpha_1, \alpha_2, \dots \in \mathbb{R}$ , then  $\sum_{i=1}^{\infty} \alpha_i \mu_i: \mathcal{A} \rightarrow [0, \infty]$  defined by

$$\left( \sum_{i=1}^{\infty} \alpha_i \mu_i \right) (A) := \sum_{i=1}^{\infty} \alpha_i \mu_i(A), \quad \forall A \in \mathcal{A}, \quad (1.48)$$

is again a measure on  $(\Omega, \mathcal{A})$  (see Exercise 1-24). For  $0 = \alpha_{n+1} = \alpha_{n+2} = \dots$  this implies: If  $\mu_1, \dots, \mu_n$  are measures on  $(\Omega, \mathcal{A})$  and  $\alpha_1, \dots, \alpha_n$  are nonnegative, then the function  $\sum_{i=1}^n \alpha_i \mu_i$  defined by

$$\left( \sum_{i=1}^n \alpha_i \mu_i \right) (A) := \sum_{i=1}^n \alpha_i \mu_i(A), \quad \forall A \in \mathcal{A}, \quad (1.49)$$

is also a measure on  $(\Omega, \mathcal{A})$ .  $\triangleleft$

### 1.4.4 Finite and $\sigma$ -Finite Measures

A measure  $\mu$  on a measurable space  $(\Omega, \mathcal{A})$  is called *finite* if  $\mu(\Omega) < \infty$ . Otherwise it is called *infinite*. Within the class of infinite measures there is a subclass with an important property, called  *$\sigma$ -finiteness*. Many fundamental propositions of measure and integration theory only hold for measures that are  $\sigma$ -finite.

#### Definition 1.63 ( $\sigma$ -Finite Measure)

Let  $\mu$  be a measure on a measurable space  $(\Omega, \mathcal{A})$ . Then  $\mu$  is called  *$\sigma$ -finite* if there is a sequence  $A_1, A_2, \dots \in \mathcal{A}$  with  $\bigcup_{i=1}^{\infty} A_i = \Omega$  and, for all  $i = 1, 2, \dots$ ,  $\mu(A_i) < \infty$ .

To emphasize, even if  $\mu(\Omega) = \infty$ , the measure  $\mu$  can be  $\sigma$ -finite (see Examples 1.64 and 1.65). Note that any finite measure is also  $\sigma$ -finite.

**Example 1.64 ( $\sigma$ -Finiteness of the Lebesgue-Measure)** The Lebesgue measure  $\lambda$  on  $(\mathbb{R}, \mathcal{B})$  is  $\sigma$ -finite, because  $\mathbb{R} = \bigcup_{i=1}^{\infty} [-i, i]$  and  $\lambda([-i, i]) = 2 \cdot i < \infty$ , for all  $i \in \mathbb{N}$ .  $\triangleleft$

**Example 1.65 (A  $\sigma$ -Finite Counting Measure)** Consider the measurable space  $(\mathbb{R}, \mathcal{B})$  and the measure  $\mu: \mathcal{B} \rightarrow [0, \infty]$ , where  $\mu = \sum_{i=0}^{\infty} \delta_i$  and  $\delta_i$  denotes the Dirac measure at  $i$  on  $(\mathbb{R}, \mathcal{B})$  with  $\delta_i(A) = 1_A(i)$ ,  $A \in \mathcal{B}$ ,  $i \in \mathbb{N}_0$  (see Example 1.57). Then  $\mu$  is  $\sigma$ -finite because  $\mathbb{R} = \bigcup_{n=1}^{\infty} [-n, n]$  and  $\mu([-n, n]) = n + 1$ , for all  $n \in \mathbb{N}_0$ . This measure simply counts the number of elements  $i \in \mathbb{N}_0$  in a Borel set  $A$ . In other words, for all finite  $A \in \mathcal{B}$ ,  $\mu(A)$  is the cardinality of the set  $A \cap \mathbb{N}_0$ .  $\triangleleft$

### 1.4.5 Product Measure

In section 1.4.2 we considered the Lebesgue measure on  $(\mathbb{R}^n, \mathcal{B}_n)$  that is specified for  $n$ -dimensional cuboids by Equation (1.46) using the *product* of one-dimensional Lebesgue measures on  $(\mathbb{R}, \mathcal{B})$ . Now we introduce the general concept of a product measure. The following lemma shows that  $\sigma$ -finiteness of measures is sufficient for the existence and uniqueness of such a measure. Hence, this lemma shows that presuming finite measures is sufficient but not necessary for the definition of the product measure.

#### Lemma 1.66 (Existence and Uniqueness)

Let  $(\Omega_i, \mathcal{A}_i, \mu_i)$  be measure spaces with  $\sigma$ -finite measures  $\mu_i$ ,  $i = 1, \dots, n$ . Then there is a uniquely defined measure, denoted  $\mu_1 \otimes \dots \otimes \mu_n$ , on the product space

$$\left( \prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i \right),$$

satisfying

$$\begin{aligned} \forall (A_1, \dots, A_n) \in \mathcal{A}_1 \times \dots \times \mathcal{A}_n: \\ \mu_1 \otimes \dots \otimes \mu_n(A_1 \times \dots \times A_n) = \mu_1(A_1) \cdot \dots \cdot \mu_n(A_n). \end{aligned} \quad (1.50)$$

This measure is  $\sigma$ -finite as well.

For a proof see Bauer (2001, Th. 23.9, p. 143). Hence,  $\mu := \mu_1 \otimes \dots \otimes \mu_n$  is a measure on the product space  $(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i)$  with

$$\mu(A_1 \times \dots \times A_n) := \mu_1(A_1) \cdot \dots \cdot \mu_n(A_n), \quad \forall (A_1, \dots, A_n) \in (\mathcal{A}_1 \times \dots \times \mathcal{A}_n). \quad (1.51)$$

#### Definition 1.67 (Product Measure)

The measure  $\mu_1 \otimes \dots \otimes \mu_n$  defined by Equation (1.50) is called the *product measure* of  $\mu_1, \dots, \mu_n$ .

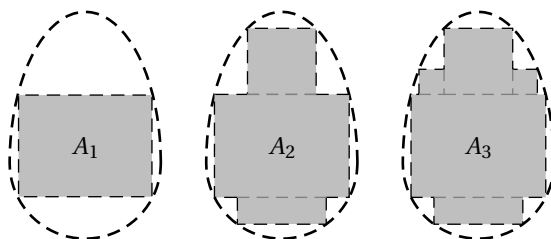


Figure 1.4. Approximation of an open egg-shaped set  $O$  from below

## 1.5 Continuity of a Measure

The term  $\sigma$ -additivity refers to *countable* unions of pairwise disjoint sets and it implies finite additivity, which involves *finite* unions of pairwise disjoint sets. Furthermore,  $\sigma$ -additivity implies the following continuity properties of a measure, which is essential for the definition of the integral (see ch. 3).

### Theorem 1.68 (Continuity of a Measure)

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space and let  $A_1, A_2, \dots \in \mathcal{A}$ .

(i) If  $A_1 \subset A_2 \subset \dots$ , then

$$\lim_{i \rightarrow \infty} \mu(A_i) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right). \quad (\text{continuity from below})$$

(ii) If  $A_1 \supset A_2 \supset \dots$  and there is an  $n \in \mathbb{N}$  with  $\mu(A_i) < \infty$ , then

$$\lim_{i \rightarrow \infty} \mu(A_i) = \mu\left(\bigcap_{i=1}^{\infty} A_i\right). \quad (\text{continuity from above})$$

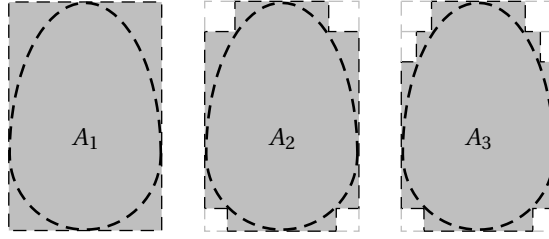
For a proof see Klenke (2013, Theorem 1.36, p. 15, 16).

**Remark 1.69 (Finite Case)** If  $A_1, \dots, A_n \in \mathcal{A}$  is a finite sequence with  $A_1 \subset \dots \subset A_n$ , then  $\bigcup_{i=1}^n A_i = A_n$  and

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \mu(A_n). \quad (1.52)$$

This is a trivial case of Theorem 1.68 (with  $A_n = A_{n+1} = A_{n+2} = \dots$ ).  $\triangleleft$

**Example 1.70 (Geometric Examples)** Figures 1.4 and 1.5 illustrate this theorem for the Lebesgue measure  $\lambda_2$  on  $(\mathbb{R}^2, \mathcal{B}_2)$ , the *area* of a set  $O$  and the sets  $A_i, i \in \mathbb{N}$ . In this example,  $A_1$  is the open rectangle in the open (i.e., the set without its boundary) egg-shaped set  $O$  displayed in Figure 1.4,  $A_2$  the union of  $A_1$  with two



**Figure 1.5.** Approximation of an open egg-shaped set  $O$  from above

other rectangles in the middle figure, and  $A_3$  the union of  $A_2$  with two additional rectangles in the right figure. Adding more and more rectangles it is plausible that  $A_1 \subset A_2 \subset \dots \subset O$  and that their union approximates  $O$ , i.e.,  $\bigcup_{i=1}^{\infty} A_i = O$ . Under these premises Theorem 1.68 (i) yields the conclusion  $\lim_{i \rightarrow \infty} \lambda_2(A_i) = \lambda_2(\bigcup_{i=1}^{\infty} A_i) = \lambda_2(O)$ . Figure 1.5 illustrates the same principle. However, now the area of the egg-shaped set  $O$  is approximated from above by subtracting the areas of appropriate rectangles.

As a second example consider the Lebesgue measure  $\lambda$  on  $(\mathbb{R}, \mathcal{B})$  and the intervals  $A_i = ]x - \frac{1}{i}, x]$ ,  $i \in \mathbb{N}$ . Obviously,  $A_1 \supset A_2 \supset \dots$  and  $\lambda(A_i) = \frac{1}{i} < \infty$ , for all  $i \in \mathbb{N}$  (see also Exercise 1-12). Hence, for all  $x \in \mathbb{R}$ ,

$$\lambda(\{x\}) = \lambda\left(\bigcap_{i=1}^{\infty} ]x - \frac{1}{i}, x]\right) = \lim_{i \rightarrow \infty} \lambda(]x - \frac{1}{i}, x]) = \lim_{i \rightarrow \infty} \frac{1}{i} = 0. \quad (1.53)$$

This is an implication of continuity from above, and it implies

$$\forall a, b \in \mathbb{R}: a < b \Rightarrow \lambda(]a, b]) = \lambda([a, b]) = \lambda([a, b[) = \lambda(]a, b[) = b - a. \quad (1.54)$$

◁

**Remark 1.71 (A Motivation for  $\sigma$ -Additivity)** As already mentioned in Remark 1.45,  $\sigma$ -additivity refers to unions of finitely or countably many sets. Now consider  $\bigcup_{1 \leq x \leq 2} \{x\} = [1, 2] \in \mathcal{B}$  [see Eq. (1.9)]. According to Equation (1.53),  $\lambda(\{x\}) = 0$ , for all  $x \in [1, 2]$ , and hence  $\lambda(\{x \in [1, 2]: x \in \mathbb{Q}\}) = 0$ , because the set of rational numbers is countable. In other words, the Lebesgue measure  $\lambda$  of the set of all rational number in the closed interval  $[1, 2]$  is zero, and this is not a contradiction to

$$\lambda\left(\bigcup_{1 \leq x \leq 2} \{x\}\right) = \lambda([1, 2]) = 2 - 1 = 1,$$

because  $\bigcup_{1 \leq x \leq 2} \{x\}$  is an uncountable union. This illustrates that additivity for uncountable unions can be meaningless. ◁

## 1.6 Specifying a Measure via a Generating System

Given a measurable space  $(\Omega, \mathcal{A})$ , a measure is a function that is defined on  $\mathcal{A}$ . In many situations, e. g., when  $\mathcal{A} = \sigma(\mathcal{E})$  can only be described by a generating set system  $\mathcal{E}$  (such as the set system  $\mathcal{I}_1$  generating the Borel  $\sigma$ -algebra on  $\mathbb{R}$ ), it is important to answer the following questions:

- (a) *Existence*: If there is a set function  $\tilde{\mu}: \mathcal{E} \rightarrow \bar{\mathbb{R}}$ , is there also a measure  $\mu: \sigma(\mathcal{E}) \rightarrow \bar{\mathbb{R}}$  such that  $\mu(A) = \tilde{\mu}(A), \forall A \in \mathcal{E}$ ?
- (b) *Uniqueness*: Is a measure  $\mu$  on  $(\Omega, \sigma(\mathcal{E}))$  already uniquely defined by its values  $\mu(A), A \in \mathcal{E}$ ?

(Sufficient conditions for the existence of such a measure  $\mu$  are formulated in Klenke, 2013, Theorem 1.53, p. 23.)

The following uniqueness theorem for finite measures provides an answer to these questions, which suffices for our purposes. (A more general formulation for  $\sigma$ -finite measures with additional assumptions and a proof of Theorem 1.72 is found in Klenke, 2013, Lemma 1.42, p. 18.)

### Theorem 1.72 (Generating System and Uniqueness of a Measure)

Let  $(\Omega, \mathcal{A})$  be a measurable space and let  $\mathcal{E} \subset \mathcal{A}$ , where  $\mathcal{E}$  is  $\cap$ -stable and  $\sigma(\mathcal{E}) = \mathcal{A}$ . If  $\mu_1$  and  $\mu_2$  are finite measures on  $(\Omega, \mathcal{A})$ , i. e., measures with  $\mu_1(\Omega), \mu_2(\Omega) < \infty$ , then

$$\forall A \in \mathcal{E}: \mu_1(A) = \mu_2(A) \quad \Rightarrow \quad \forall A \in \mathcal{A}: \mu_1(A) = \mu_2(A).$$

**Example 1.73 (Countable  $\Omega$ )** Let  $\Omega$  be a finite or countable set and let  $\mathcal{A} = \mathcal{P}(\Omega)$ . Then the set system

$$\mathcal{E}_1 = \{\emptyset\} \cup \{\{\omega\}: \omega \in \Omega\}$$

is  $\cap$ -stable and  $\sigma(\mathcal{E}_1) = \mathcal{A}$ . As already noted in Remark 1.48, a finite measure  $\mu$  on  $(\Omega, \mathcal{A})$  is uniquely defined by its values  $\mu(\{\omega\}), \omega \in \Omega$ .  $\triangleleft$

**Example 1.74 (Measures on  $(\mathbb{R}, \mathcal{B})$ )** The set system

$$\mathcal{E}_2 = \{]a, b[: a < b, a, b \in \mathbb{R}\} \cup \{\emptyset\}$$

is  $\cap$ -stable and  $\sigma(\mathcal{E}_2) = \mathcal{B}$  [see Eq. (1.18) and section 1.2.4]. Another  $\cap$ -stable set system  $\mathcal{E}_3$  with  $\sigma(\mathcal{E}_3) = \mathcal{B}$  is

$$\mathcal{E}_3 = \{]-\infty, b]: b \in \mathbb{R}\}$$

(cf. Klenke, 2013, p. 9). This set system is crucial for the definition of a cumulative distribution function (see section 5.7.1).  $\triangleleft$

## 1.7 $\sigma$ -Algebra That is Trivial With Respect to a Measure

All  $\sigma$ -algebras treated in section 1.2 have been defined without reference to a measure. Now we define the concept of a *trivial  $\sigma$ -algebra*, which is defined referring to a measure. We start with a lemma about the set of all subsets of a set  $\Omega$  with  $\mu(A) = 0$  or  $\mu(A) = \mu(\Omega)$ , i. e., the set of all sets that are *trivial* with respect to the measure  $\mu$ . Hence, the set of  $\mu$ -trivial sets includes all *null sets*, i. e., all sets  $A \subset \Omega$  with  $\mu(A) = 0$ , and all sets  $A \subset \Omega$  with  $\mu(A) = \mu(\Omega)$ .

### Lemma 1.75 (The Set of all Trivial Sets is a $\sigma$ -Algebra)

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space and assume that  $\mu$  is finite. Then

$$\mathcal{T}_\mu := \{A \in \mathcal{A} : \mu(A) = 0 \text{ or } \mu(A) = \mu(\Omega)\} \quad (1.55)$$

is a  $\sigma$ -algebra.

(Proof p. 30)

This lemma allows for the following definition:

### Definition 1.76 (Trivial $\sigma$ -Algebra With Respect to a Measure)

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space, assume that  $\mu$  is finite, and let  $\mathcal{T}_\mu$  be defined by (1.55). Then each  $\sigma$ -algebra  $\mathcal{C} \subset \mathcal{T}_\mu$  is called a  $\mu$ -trivial  $\sigma$ -algebra and its elements  $\mu$ -trivial sets.

Obviously,  $\{\Omega, \emptyset\}$  is a trivial  $\sigma$ -algebra with respect to all measures on  $\{\Omega, \emptyset\}$ . Hence, we can call it a *trivial  $\sigma$ -algebra*.

## 1.8 Proofs

### Proof of Theorem 1.12

(a)

$$\begin{aligned} \forall i \in I: \mathcal{A}_i \text{ is a } \sigma\text{-algebra on } \Omega &\Rightarrow \forall i \in I: \Omega \in \mathcal{A}_i && [\text{Def. 1.1 (a)}] \\ &\Rightarrow \Omega \in \bigcap_{i \in I} \mathcal{A}_i. \end{aligned}$$

(b)

$$\begin{aligned} A \in \bigcap_{i \in I} \mathcal{A}_i &\Rightarrow \forall i \in I: A \in \mathcal{A}_i \\ &\Rightarrow \forall i \in I: A^c \in \mathcal{A}_i && [\text{Def. 1.1 (b)}] \\ &\Rightarrow A^c \in \bigcap_{i \in I} \mathcal{A}_i. \end{aligned}$$

(c)

$$\begin{aligned}
A_1, A_2, \dots \in \bigcap_{i \in I} \mathcal{A}_i &\Rightarrow \forall i \in I: A_1, A_2, \dots \in \mathcal{A}_i \\
&\Rightarrow \forall i \in I: \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}_i \quad [\text{Def. 1.1 (c)}] \\
&\Rightarrow \bigcup_{j=1}^{\infty} A_j \in \bigcap_{i \in I} \mathcal{A}_i.
\end{aligned}$$

### **Proof of Lemma 1.15**

If  $\mathcal{C}$  is a  $\sigma$ -algebra with  $\mathcal{E} \subset \mathcal{C}$  and  $\mathcal{A} = \sigma(\mathcal{E})$ , then (1.11) and the assumption  $\mathcal{C} \subset \mathcal{A}$  imply  $\mathcal{A} = \sigma(\mathcal{E}) \subset \mathcal{C} \subset \mathcal{A}$ . Hence,  $\mathcal{C} = \mathcal{A}$ .

### **Proof of Lemma 1.20**

Define  $\mathcal{D} := \{C = \bigcup_{i \in I(C)} B_i : I(C) \subset \mathbb{N}\}$ .

$\mathcal{E} \subset \mathcal{D}$ : For  $B_j \in \mathcal{E}$  choose  $I(B_j) = \{j\}$ . Then  $B_j = \bigcup_{i \in I(B_j)} B_i$ .

$\mathcal{D} \subset \mathcal{E}$ : Because  $\mathbb{N}$  is countable, any  $I(C) \subset \mathbb{N}$  is finite or countable, and this implies that  $C = \bigcup_{i \in C} B_i$  is an element of  $\sigma(\mathcal{E})$  [see Def. 1.1 (c), (1.3)].

Checking the three conditions defining a  $\sigma$ -algebra (see Def. 1.1), we show that  $\mathcal{D}$  is a  $\sigma$ -algebra.

(a)

$$\Omega = \begin{cases} \bigcup_{i=1}^n B_i, & \text{if } \mathcal{E} = \{B_1, \dots, B_n\} \\ \bigcup_{i=1}^{\infty} B_i, & \text{if } \mathcal{E} = \{B_1, B_2, \dots\}, \end{cases}$$

because  $\mathcal{E}$  is assumed to be a partition. This shows that  $\Omega \in \mathcal{D}$ .

(b) The equation for  $\Omega$  in (a) also implies  $I(C^c) = I(C)^c$ . Therefore,  $C^c \in \mathcal{D}$  if  $C \in \mathcal{D}$ .

(c) If  $C_1, C_2, \dots \in \mathcal{D}$ , then

$$\bigcup_{j=1}^{\infty} C_j = \bigcup_{j=1}^{\infty} \bigcup_{i \in I(C_j)} B_i = \bigcup_{i \in \bigcup_{j=1}^{\infty} I(C_j)} B_i \in \mathcal{D},$$

because  $\bigcup_{j=1}^{\infty} I(C_j) \subset \mathbb{N}$ .

Finally, we prove the second equation in (1.14). If  $j \in I(C)$  and  $C = \bigcup_{i \in I(C)} B_i$ , then  $B_j \subset C$ , which implies

$$\bigcup_{i \in I(C)} B_i \subset \bigcup_{B_i \subset C} B_i.$$

Vice versa, if  $B_j \subset C$ , then  $j \in I(C)$ , because for any  $\omega \in B_j$ , there is no  $i \neq j$  such that  $\omega \in B_i$  [see condition (b) of Rem. 1.19]. Hence,

$$\bigcup_{B_i \subset C} B_i \subset \bigcup_{i \in I(C)} B_i,$$

which proves the second equation in (1.14).

### **Proof of Lemma 1.27**

In this proof we use  $\sigma_\Omega(\mathcal{E})$  to denote the  $\sigma$ -algebra on  $\Omega$  generated by  $\mathcal{E} \subset \mathcal{P}(\Omega)$ . Similarly,  $\sigma_A(\mathcal{D})$  denotes the  $\sigma$ -algebra on  $A$  generated by  $\mathcal{D} \subset \mathcal{P}(A)$ .

(1.15).  $\sigma_\Omega(\mathcal{E})$  is a  $\sigma$ -algebra on  $\Omega$  and  $\mathcal{E} \subset \sigma_\Omega(\mathcal{E})$ , by definition of  $\sigma_\Omega(\mathcal{E})$ . Hence,  $\mathcal{E}|_A \subset \sigma_\Omega(\mathcal{E})|_A$ , and  $\sigma_\Omega(\mathcal{E})|_A$  is a  $\sigma$ -algebra on  $A$  (see Exercise 1-5). Therefore, the definition (1.10) yields

$$\sigma_A(\mathcal{E}|_A) \subset \sigma_\Omega(\mathcal{E})|_A.$$

Furthermore,  $\mathcal{E} \subset \sigma_\Omega(\mathcal{E}|_A \cup \mathcal{E}|_{A^c})$ , which implies

$$\sigma_\Omega(\mathcal{E}) \subset \sigma_\Omega(\mathcal{E}|_A \cup \mathcal{E}|_{A^c}) \quad [\text{Rem. 1.23}]$$

$$\subset \sigma_\Omega(\sigma_A(\mathcal{E}|_A) \cup \sigma_{A^c}(\mathcal{E}|_{A^c})) \quad [\text{Rem. 1.23}]$$

$$= \{C \cup D: C \in \sigma_A(\mathcal{E}|_A), D \in \sigma_{A^c}(\mathcal{E}|_{A^c})\}. \quad [\text{this set system is a } \sigma\text{-algebra}]$$

Therefore,

$$\begin{aligned} \sigma_\Omega(\mathcal{E})|_A &\subset \{C \cup D: C \in \sigma_A(\mathcal{E}|_A), D \in \sigma_{A^c}(\mathcal{E}|_{A^c})\}|_A \\ &= \{(C \cup D) \cap A: C \in \sigma_A(\mathcal{E}|_A), D \in \sigma_{A^c}(\mathcal{E}|_{A^c})\} \\ &= \{C \cap A: C \in \sigma_A(\mathcal{E}|_A)\} \quad [D \subset A^c] \\ &= \sigma_A(\mathcal{E}|_A). \quad [C \subset A] \end{aligned}$$

Hence, we have shown  $\sigma_A(\mathcal{E}|_A) \subset \sigma_\Omega(\mathcal{E})|_A$  and  $\sigma_\Omega(\mathcal{E})|_A \subset \sigma_A(\mathcal{E}|_A)$ , which is equivalent to  $\sigma_A(\mathcal{E}|_A) = \sigma_\Omega(\mathcal{E})|_A$ .

(1.17).

$$\begin{aligned} \sigma_\Omega(\mathcal{C} \cup \mathcal{E})|_A &= \sigma_A(\mathcal{C} \cup \mathcal{E}|_A) \quad [(1.15)] \\ &= \sigma_A(\mathcal{C}|_A \cup \mathcal{E}|_A) \quad [\text{see def. of the trace in Example 1.10}] \\ &= \sigma_A(\mathcal{C}|_A \cup \{\emptyset, A\}) \quad [(1.16)] \\ &= \sigma_A(\mathcal{C}|_A) \quad [\{\emptyset, A\} \subset \mathcal{C}|_A] \\ &= \mathcal{C}|_A. \quad [\text{Exercise 1-5, (1.12)}] \end{aligned}$$

### **Proof of Lemma 1.75**

(a)  $\Omega \in \mathcal{T}_\mu$  by definition of  $\mathcal{T}_\mu$ .

(b) If  $A \in \mathcal{T}_\mu$ , then Rules (iv) and (v) of Box 1.1 and finiteness of  $\mu$  yield

$$\mu(A^c) = \mu(\Omega) - \mu(A) = \begin{cases} \mu(\Omega), & \text{if } \mu(A) = 0, \\ 0, & \text{if } \mu(A) = \mu(\Omega), \end{cases}$$

which implies  $A^c \in \mathcal{T}_\mu$ .

(c) Let  $A_1, A_2, \dots \in \mathcal{A}$ . We consider two cases. *First*, if  $\mu(A_i) = 0$ , for all  $A_i, i \in \mathbb{N}$ , then Rule (xi) of Box 1.1 yields  $\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i) = 0$ , i. e.,  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{T}_\mu$ . *Second*, if there is a  $j \in \mathbb{N}$  such that  $\mu(A_j) = \mu(\Omega)$ , then Rule (v) of Box 1.1 yields

$$\mu(\Omega) = \mu(A_j) \leq \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \mu(\Omega),$$

which implies  $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu(\Omega)$ . Therefore,  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_\mu$ .

## 1.9 Exercises

▷ **Exercise 1-1** Let  $\mathcal{A}$  be a  $\sigma$ -algebra of subsets of a nonempty set  $\Omega$  and let  $A_1, A_2, \dots \in \mathcal{A}$ . Show: (a)  $A_1 \cap A_2 \cap \dots \in \mathcal{A}$ , (b)  $A_1 \cap A_2 \in \mathcal{A}$ , and (c)  $A_1 \setminus A_2 \in \mathcal{A}$ .

▷ **Exercise 1-2** Show that the set system  $\mathcal{A} = \{\Omega, \emptyset, A, A^c\}$  is stable (closed) with respect to union of elements of  $\mathcal{A}$ .

▷ **Exercise 1-3** Consider the set  $\Omega = \{\omega_1, \dots, \omega_6\}$  representing the set of all possible outcomes of tossing a dice and the power set  $\mathcal{P}(\Omega)$ , which, in probability theory, represents the set of all possible events (including the ‘impossible’ event  $\emptyset$ ) in this random experiment. Specify the  $\sigma$ -algebra on  $\Omega$  that represents all possible events if we only distinguish between even and uneven number of points.

▷ **Exercise 1-4** Consider the random experiment that has been described in Example 1.9. Aside from the power set of  $\Omega$  we already considered the  $\sigma$ -algebras  $\mathcal{A}_1 = \{\Omega, \emptyset, A, A^c\}$ ,  $\mathcal{A}_2 = \{\Omega, \emptyset, B, B^c\}$ , and  $\mathcal{A}_3 = \{\Omega, \emptyset, C, C^c\}$ . Define another  $\sigma$ -algebra not yet mentioned.

▷ **Exercise 1-5** Prove: If  $\mathcal{A}$  is a  $\sigma$ -algebra on  $\Omega$  and  $\Omega_0 \subset \Omega$ , then  $\mathcal{A}|_{\Omega_0} = \{\Omega_0 \cap A : A \in \mathcal{A}\}$  is a  $\sigma$ -algebra on  $\Omega_0$ .

▷ **Exercise 1-6** Prove the proposition of Remark 1.16.

▷ **Exercise 1-7** Show that  $\sigma(\mathcal{E}) = \mathcal{P}(\Omega)$  if  $\Omega$  is finite or countable and  $\mathcal{E} := \{\{\omega\} : \omega \in \Omega\}$ .

▷ **Exercise 1-8** Prove the proposition of Remark 1.21.

▷ **Exercise 1-9** Let  $\mathcal{E}_1, \mathcal{E}_2$  be set systems on  $\Omega$  with  $\mathcal{E}_1 \subset \mathcal{E}_2$ . Show that  $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$ .

▷ **Exercise 1-10** Prove propositions (a) and (b) of Example 1.25.

▷ **Exercise 1-11** Prove Equation (1.21).

▷ **Exercise 1-12** Show that  $\{x\} \in \mathcal{B}_n$  for all  $x \in \mathbb{R}^n$ , where  $\mathcal{B}_n$  is the Borel  $\sigma$ -algebra on  $\mathbb{R}^n$ .

▷ **Exercise 1-13** Let  $\mathcal{B}$  be the Borel  $\sigma$ -algebra on  $\mathbb{R}$  and let  $\Omega_0 \subset \mathbb{R}$  be finite or countable. Show that  $\mathcal{B}|_{\Omega_0} = \mathcal{P}(\Omega_0)$ .

▷ **Exercise 1-14** Prove the proposition of Example 1.34.

▷ **Exercise 1-15** Prove the proposition of Remark 1.35.

▷ **Exercise 1-16** Let  $(\Omega_i, \mathcal{A}_i)$ ,  $i = 1, \dots, n$ , be measurable spaces. Show that the set system  $\mathcal{E} := \{A_1 \times \dots \times A_n : A_i \in \mathcal{A}_i, i = 1, \dots, n\}$  is  $\cap$ -stable.

▷ **Exercise 1-17** Prove the proposition of Remark 1.46.

▷ **Exercise 1-18** Prove the rules of Box 1.1.

- ▷ **Exercise 1-19** Prove the propositions of Remark 1.47.
- ▷ **Exercise 1-20** Show that  $\delta_\omega: \mathcal{A} \rightarrow \{0, 1\}$  in Example 1.52 is a measure.
- ▷ **Exercise 1-21** Prove that the function defined by Equation (1.40) is a measure on  $(\Omega, \mathcal{A})$ .
- ▷ **Exercise 1-22** Show that  $\sum_{\omega \in B} \delta_\omega$  in Example 1.57 is a measure.
- ▷ **Exercise 1-23** Show that  $\nu: \mathcal{C} \rightarrow \bar{\mathbf{R}}$  defined in Example 1.61 is a measure on  $(\Omega, \mathcal{C})$ .
- ▷ **Exercise 1-24** Prove that the function  $\sum_{i=1}^{\infty} \alpha_i \mu_i$  defined in Example 1.62 is a measure on  $(\Omega, \mathcal{A})$ .

## Solutions

- ▷ **Solution 1-1** (a) If  $A_1, A_2, \dots \in \mathcal{A}$ , then  $A_1^c, A_2^c, \dots \in \mathcal{A}$  [see Def. 1.1 (b)]. Hence,

$$\bigcap_{i=1}^{\infty} A_i = \left[ \left( \bigcap_{i=1}^{\infty} A_i \right)^c \right]^c = \left[ \bigcup_{i=1}^{\infty} A_i^c \right]^c \quad \begin{array}{l} \text{[de Morgan]} \\ \text{[Def. 1.1 (c), (b)]} \end{array}$$

$\in \mathcal{A}.$

- (b) Let  $A_1, A_2 \in \mathcal{A}$  and choose  $A_3, A_4, \dots$  such that  $\Omega = A_i$ , for all  $i \geq 3$ ,  $i \in \mathbf{N}$ . Then, according to Definition 1.1 (a),

$$A_1 \cap A_2 = A_1 \cap A_2 \cap \Omega = \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}.$$

- (c)  $A_1 \setminus A_2 = A_1 \cap A_2^c \in \mathcal{A}$  [see (b) and Def. 1.1 (b)].

- ▷ **Solution 1-2** The unions  $\Omega \cup A = \Omega$ ,  $\Omega \cup A^c = \Omega$ , and  $\Omega \cup \emptyset = \Omega$  are all elements of  $\mathcal{A}$  and the same is true for  $\emptyset \cup A = A$ ,  $\emptyset \cup A^c = A^c$ , and  $A \cup A^c = \Omega$ . Furthermore,  $B \cup B = B$  for all  $B \in \mathcal{A}$ .

- ▷ **Solution 1-3** The  $\sigma$ -algebra on  $\Omega$  that only distinguishes between even and uneven number of points is  $\mathcal{A}_1 := \{\{\omega_1, \omega_3, \omega_5\}, \{\omega_2, \omega_4, \omega_6\}, \Omega, \emptyset\}$ . This is a sub- $\sigma$ -algebra of  $\mathcal{P}(\Omega)$ . Therefore,  $\mathcal{A}_1$  represents the set of all possible events of a random experiment that is, in a sense, contained in the original random experiment.

- ▷ **Solution 1-4** Consider the set system that contains as elements  $A, A^c, B, B^c, \Omega, \emptyset$ , all unions and all intersections of these sets as well as the unions and intersections of the resulting sets such as  $(A^c \cup B^c) \cap (A \cup B)$  and  $(A^c \cup B^c) \cup (A \cup B)$ . Altogether these are 16 sets. This is  $\sigma(\mathcal{A}_1 \cup \mathcal{A}_2)$ , the  $\sigma$ -algebra generated by  $\mathcal{A}_1 \cup \mathcal{A}_2 = \{A, A^c, B, B^c, \Omega, \emptyset\}$  (see Def. 1.13 and Rem. 1.21).

- ▷ **Solution 1-5** (a)  $\Omega_0 \cap \Omega = \Omega_0$ . This implies  $\Omega_0 \in \mathcal{A}|_{\Omega_0}$ .
- (b)

$$A^* \in \mathcal{A}|_{\Omega_0} \Rightarrow \exists A \in \mathcal{A}: A^* = \Omega_0 \cap A.$$

With this set  $A$  and using  $B^c$  for the complement of a set  $B$  with respect to  $\Omega$ ,

$$\Omega_0 \setminus A^* = \Omega_0 \setminus (\Omega_0 \cap A)$$

$$\begin{aligned}
&= \Omega_0 \cap (\Omega_0 \cap A)^c \\
&= \Omega_0 \cap (\Omega_0^c \cup A^c) \\
&= (\Omega_0 \cap \Omega_0^c) \cup (\Omega_0 \cap A^c) \\
&= \Omega_0 \cap A^c \in \mathcal{A}|_{\Omega_0}.
\end{aligned}$$

(c)

$$A_1^*, A_2^*, \dots \in \mathcal{A}|_{\Omega_0} \Rightarrow \exists A_1, A_2, \dots \in \mathcal{A}: A_i^* = \Omega_0 \cap A_i, i \in \mathbf{N}.$$

Hence,

$$A_1^* \cup A_2^* \cup \dots = (\Omega_0 \cap A_1) \cup (\Omega_0 \cap A_2) \cup \dots = \Omega_0 \cap (A_1 \cup A_2 \cup \dots) \in \mathcal{A}|_{\Omega_0}.$$

▷ **Solution 1-6** If  $\mathcal{G}$  is a  $\sigma$ -algebra on  $\Omega$ , then

$$\mathcal{E} \cup \mathcal{F} \subset \mathcal{G} \Leftrightarrow \sigma(\mathcal{E} \cup \mathcal{F}) \subset \mathcal{G}. \quad [(1.11)] \quad (1.56)$$

Furthermore, for three sets  $A, B, C$ ,

$$A \cup B \subset C \Leftrightarrow A \subset C \wedge B \subset C. \quad (1.57)$$

Hence,

$$\mathcal{D} \cup \mathcal{E} \cup \mathcal{F} \subset \mathcal{G} \Leftrightarrow (\mathcal{D} \subset \mathcal{G}) \wedge (\mathcal{E} \cup \mathcal{F} \subset \mathcal{G}) \quad [(1.57)]$$

$$\Leftrightarrow (\mathcal{D} \subset \mathcal{G}) \wedge (\sigma(\mathcal{E} \cup \mathcal{F}) \subset \mathcal{G}) \quad [(1.56)]$$

$$\Leftrightarrow \mathcal{D} \cup \sigma(\mathcal{E} \cup \mathcal{F}) \subset \mathcal{G}. \quad [(1.57)]$$

Now Definition 1.13 yields the proposition.

▷ **Solution 1-7** If  $\Omega$  is finite or countable, then each of its subsets  $A$  is finite or countable as well. Therefore,

$$\forall A \subset \Omega: A = \bigcup_{\omega \in A} \{\omega\} \in \sigma(\mathcal{E}). \quad [\text{Def. 1.1 (c), Rem. 1.2}]$$

Because each element  $A$  of  $\mathcal{P}(\Omega)$  is a union  $\bigcup_{\omega \in A} \{\omega\}$  of singletons  $\{\omega\}$ ,  $\omega \in A$ , this implies  $\mathcal{P}(\Omega) \subset \sigma(\mathcal{E})$ . Hence,  $\mathcal{E} \subset \mathcal{P}(\Omega) \subset \sigma(\mathcal{E})$ . Therefore, Lemma 1.15 implies  $\sigma(\mathcal{E}) = \mathcal{P}(\Omega)$ .▷ **Solution 1-8** Suppose that  $\mathcal{E} = \{A_1, \dots, A_m\}$  and  $A_j^1 := A_j$  and let  $A_j^c$  denote the complement of  $A_j$ . Then, for all  $(k_1, \dots, k_m) \in \{1, c\}^m$  define

$$B_{(k_1, \dots, k_m)} := \bigcap_{j=1}^m A_j^{k_j}.$$

Then

$$\mathcal{F} := \{B_{(k_1, \dots, k_m)} : (k_1, \dots, k_m) \in \{1, c\}^m, B_{(k_1, \dots, k_m)} \neq \emptyset\}$$

is a finite partition of  $\Omega$ . Note that  $\mathcal{F}$  contains all nonempty intersections of sets  $A_j$  or their complements, respectively, where  $j = 1, \dots, m$ . Now Lemma 1.20 implies the proposition.▷ **Solution 1-9** If  $\mathcal{E}_1 \subset \mathcal{E}_2 \subset \mathcal{P}(\Omega)$ , then for any  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$  with  $\mathcal{E}_2 \subset \mathcal{A}$  also  $\mathcal{E}_1 \subset \mathcal{A}$ . Remember, if  $J \subset I$ , then  $\bigcap_{i \in J} B_i \subset \bigcap_{i \in I} B_i$ , for any sets  $B_i, i \in I$ . Therefore,  $\sigma(\mathcal{E}_1)$ , which is the intersection of all  $\sigma$ -algebras containing  $\mathcal{E}_1$ , is a subset of the intersection of all  $\sigma$ -algebras containing  $\mathcal{E}_2$ , which is  $\sigma(\mathcal{E}_2)$ .

▷ **Solution 1-10** (a) If  $\Omega$  is finite, then  $\mathcal{P}(\Omega)$  is a finite set system. Therefore, each  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$  is a finite set system. Because  $\mathcal{A} = \sigma(\mathcal{A})$ , this  $\sigma$ -algebra is countably generated.

(b) The set  $\mathbb{N}_0$  is countable and therefore also  $\mathbb{N}_0^n$  for  $n \in \mathbb{N}$ . Example 1.18 then implies that  $\mathcal{P}(\mathbb{N}_0^n)$  is countably generated.

▷ **Solution 1-11** Let  $\mathcal{H}_n = \{ ]-\infty, b_1] \times \dots \times ]-\infty, b_n] : b_1, \dots, b_n \in \mathbb{R} \}$ .

(i) For all  $(b_1, \dots, b_n) \in \mathbb{R}^n$  and all  $m \in \mathbb{N}$  with  $m < b_i, i = 1, \dots, n$ ,

$$B_m := ]-m, b_1] \times \dots \times ]-m, b_n] \in \mathcal{H}_n.$$

According to Definition 1.1 (c) this implies

$$\bigcup_{\substack{m \in \mathbb{N} \\ m < b_i, i=1, \dots, n}} B_m = ]-\infty, b_1] \times \dots \times ]-\infty, b_n] \in \sigma(\mathcal{H}_n).$$

Hence,  $\mathcal{H}_n \subset \sigma(\mathcal{H}_n)$ , which, according to (1.11) and (1.12), implies

$$\sigma(\mathcal{H}_n) \subset \sigma(\mathcal{H}_n) = \mathcal{B}_n.$$

(ii) For all  $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$ , with  $a_i < b_i, i = 1, \dots, n$ ,

$$]a_1, b_1] \times \dots \times ]a_n, b_n] = ]-\infty, b_1] \times \dots \times ]-\infty, b_n] \setminus \left( \bigcup_{j=1}^n H_j \right),$$

where  $H_j := ]-\infty, b_1] \times \dots \times ]-\infty, b_{j-1}] \times ]-\infty, a_j] \times ]-\infty, b_{j+1}] \times \dots \times ]-\infty, b_n]$ . Hence, according to Remark 1.2,  $]a_1, b_1] \times \dots \times ]a_n, b_n] \in \sigma(\mathcal{H}_n)$  and  $\mathcal{H}_n \subset \sigma(\mathcal{H}_n)$ , which, according to (1.11) and (1.12), implies

$$\mathcal{B}_n = \sigma(\mathcal{H}_n) \subset \sigma(\mathcal{H}_n).$$

▷ **Solution 1-12** If  $x \in \mathbb{R}$ , then  $\{x\} = \bigcap_{i=1}^{\infty} ]x - 1/i, x]$ . According to Equation (1.18), the intervals  $]x - 1/i, x]$  are elements of the generating set system of  $\mathcal{B}$ , the Borel  $\sigma$ -algebra on  $\mathbb{R}$ . Therefore, their countable intersection is an element of  $\mathcal{B}$ . If  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ , then

$$\{x\} = \bigcap_{i=1}^{\infty} \left( \bigtimes_{j=1}^n ]x_j - \frac{1}{i}, x_j] \right).$$

According to Equation (1.20), the cuboids  $\bigtimes_{j=1}^n ]x_j - \frac{1}{i}, x_j]$  are elements of the set system  $\mathcal{H}_n$  and  $\sigma(\mathcal{H}_n) = \mathcal{B}_n$ .

▷ **Solution 1-13** Because  $\{x\} \in \mathcal{B}$  for all  $x \in \mathbb{R}$  (see Exercise 1-12) we can conclude:  $\{x\} \in \mathcal{B}|_{\Omega_0}$  for all  $x \in \Omega_0$ . Hence, if  $\Omega_0$  is finite or countable, Example 1.18 implies  $\mathcal{B}|_{\Omega_0} = \mathcal{P}(\Omega_0)$ .

▷ **Solution 1-14** Let  $\Omega_1, \dots, \Omega_n$  be finite or countable sets and let  $\mathcal{A}_1, \dots, \mathcal{A}_n$  be their power sets. Then  $\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n$  implies  $\{\omega_1\} \in \mathcal{A}_1, \dots, \{\omega_n\} \in \mathcal{A}_n$ . Therefore,

$$\{(\omega_1, \dots, \omega_n)\} = \{\omega_1\} \times \dots \times \{\omega_n\} \in \left\{ \bigtimes_{i=1}^n A_i : A_i \in \mathcal{A}_i, i \in \{1, \dots, n\} \right\}.$$

Hence,

$$\sigma(\{(\omega_1, \dots, \omega_n)\} : \omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n) \subset \bigotimes_{i=1}^n \mathcal{A}_i.$$

With  $\Omega_i$  being finite or countable,  $\Omega = \Omega_1 \times \dots \times \Omega_n$  is finite or countable. Therefore,

$$\sigma(\{(\omega_1, \dots, \omega_n) : \omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n\}) = \mathcal{P}(\Omega)$$

(see Example 1.18). Because  $\bigotimes_{i=1}^n \mathcal{A}_i \subset \mathcal{P}(\Omega)$ , we can conclude

$$\bigotimes_{i=1}^n \mathcal{A}_i = \mathcal{P}(\Omega_1 \times \dots \times \Omega_n) = \mathcal{P}\left(\prod_{i=1}^n \Omega_i\right).$$

▷ **Solution 1-15**

$$\begin{aligned} (A \times B)^c &= \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \omega_1 \notin A \text{ or } \omega_2 \notin B\} \\ &= \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : (\omega_1 \notin A, \omega_2 \in B) \text{ or } \omega_2 \notin B\} \\ &= (A^c \times B) \cup (\Omega_1 \times B^c) \end{aligned}$$

and

$$\begin{aligned} (A^c \times B) \cap (\Omega_1 \times B^c) &= \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \omega_1 \notin A, \omega_2 \in B, \omega_2 \notin B\} \\ &= \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \omega_1 \notin A, \omega_2 \in B \cap B^c = \emptyset\} \\ &= \emptyset. \end{aligned}$$

▷ **Solution 1-16** Remember that  $(a \in A, b \in B)$  means  $(a \in A \text{ and } b \in B)$  and that  $(a \in A \text{ and } b \in B)$  and  $(b \in B \text{ and } a \in A)$  are equivalent. Let  $A_1, B_1 \in \mathcal{A}_1, \dots, A_n, B_n \in \mathcal{A}_n$ . Then  $A_1 \cap B_1 \in \mathcal{A}_1, \dots, A_n \cap B_n \in \mathcal{A}_n$ . Hence,  $A_1 \times \dots \times A_n \in \mathcal{E}, B_1 \times \dots \times B_n \in \mathcal{E}$  and  $(A_1 \cap B_1) \times \dots \times (A_n \cap B_n) \in \mathcal{E}$ . Furthermore,

$$\begin{aligned} (A_1 \times \dots \times A_n) \cap (B_1 \times \dots \times B_n) &= \{(\omega_1, \dots, \omega_n) : \omega_1 \in A_1, \dots, \omega_n \in A_n, \omega_1 \in B_1, \dots, \omega_n \in B_n\} \\ &= \{(\omega_1, \dots, \omega_n) : \omega_1 \in (A_1 \cap B_1), \dots, \omega_n \in (A_n \cap B_n)\} \\ &= (A_1 \cap B_1) \times \dots \times (A_n \cap B_n) \in \mathcal{E}. \end{aligned}$$

▷ **Solution 1-17** Let  $B_i$  denote the sets defined in Remark 1.46.

(i)  $B_1 = A_1 \in \mathcal{A}$ . For all  $i \in \mathbb{N}, i > 1, B_i \in \mathcal{A}$ :

$$B_i = A_i \setminus \left( \bigcup_{j=1}^{i-1} A_j \right) = A_i \cap \left( \bigcup_{j=1}^{i-1} A_j \right)^c \in \mathcal{A}. \quad [\text{Def. 1.1 (b), Rem. 1.2}]$$

(ii) For any sequence  $C_1, C_2, \dots \subset \Omega$ , define

$$\bigcup_{j=m}^n C_j := \emptyset, \quad \text{if } m > n, \quad \text{and} \quad \bigcap_{j=m}^n C_j := \Omega, \quad \text{if } m > n.$$

Then, using associativity and commutativity of the intersection, for  $1 \leq k < l$ ,

$$\begin{aligned} B_k \cap B_l &= \left[ A_k \setminus \left( \bigcup_{j=1}^{k-1} A_j \right) \right] \cap \left[ A_l \setminus \left( \bigcup_{j=1}^{l-1} A_j \right) \right] \\ &= A_k \cap \left( \bigcup_{j=1}^{k-1} A_j \right)^c \cap A_l \cap \left( \bigcup_{j=1}^{l-1} A_j \right)^c && [A \setminus B = A \cap B^c] \\ &= A_k \cap \left( \bigcap_{j=1}^{k-1} A_j^c \right) \cap A_l \cap \left( \bigcap_{j=1}^{l-1} A_j^c \right) && [\text{de Morgan}] \\ &= A_k \cap A_l \cap \left( \bigcap_{j=1}^{k-1} A_j^c \right) \cap \left( \bigcap_{j=1}^{l-1} A_j^c \right) \cap A_k^c \cap \left( \bigcap_{j=k+1}^{l-1} A_j^c \right) \\ &= \emptyset. && [A_k \cap A_k^c = \emptyset] \end{aligned}$$

(iii) The sets  $B_i$  are defined such that  $B_i \subset A_i$ , for all  $i \in I$ . Therefore,  $\bigcup_{i=1}^{\infty} B_i \subset \bigcup_{i=1}^{\infty} A_i$ . Furthermore, for all  $\omega \in \Omega$ ,

$$\begin{aligned} \omega \in \bigcup_{i=1}^{\infty} A_i &\Rightarrow \exists i \in \mathbf{N}: \omega \in A_i \wedge (\forall j < i: \omega \notin A_j) \\ &\Rightarrow \exists i \in \mathbf{N}: \omega \in A_1^c \cap \dots \cap A_{i-1}^c \cap A_i = B_i \\ &\Rightarrow \omega \in \bigcup_{i=1}^{\infty} B_i. \end{aligned}$$

Hence,  $\bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i=1}^{\infty} B_i$ , and this implies  $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ .

▷ **Solution 1-18** (i) This is condition (c) of Definition 1.43.

(ii) If  $A_1, \dots, A_n \in \mathcal{A}$  are pairwise disjoint, then  $A_1, A_2, \dots$  with  $\emptyset = A_{n+1} = A_{n+2} = \dots$  is a sequence of pairwise disjoint measurable sets. Therefore, conditions (a) and (c) of Def. 1.43 imply

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) = \sum_{i=1}^n \mu(A_i) + \sum_{i=n+1}^{\infty} \mu(\emptyset) = \sum_{i=1}^n \mu(A_i).$$

(iii) For  $A, B \subset \Omega$ ,

$$A = (A \cap B) \cup (A \cap B^c) = (A \cap B) \cup (A \setminus B)$$

and

$$(A \cap B) \cap (A \cap B^c) = A \cap B \cap B^c = \emptyset.$$

Hence, for sets  $A, B \in \mathcal{A}$ , Rule (ii) (finite additivity of  $\mu$ ) implies proposition (iii).

(iv) This proposition is a special case of (iii) with  $A = \Omega$ .

(v) Exchanging the roles of  $A$  and  $B$  in (iii) we obtain

$$\mu(B) = \mu(A \cap B) + \mu(B \setminus A).$$

If  $A \subset B$ , then  $A \cap B = A$  and, because  $\mu(B \setminus A) \geq 0$ ,

$$\mu(A) = \mu(A \cap B) \leq \mu(A \cap B) + \mu(B \setminus A) = \mu(B).$$

(vi) This rule immediately follows from proposition (iv) for  $\mu(A \cap B) < \infty$ . [Note that  $\mu(A) - \mu(A \cap B)$  is not defined if  $\mu(A) = \mu(A \cap B) = \infty$ .]

(vii) For  $A, B \subset \Omega$ ,

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).$$

Because the right-hand side is a union of pairwise disjoint sets, finite additivity of  $\mu$  yields

$$\begin{aligned} \mu(A \cup B) + \mu(A \cap B) &= \mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A) + \mu(A \cap B) \\ &= \mu(A) + \mu(B). \end{aligned} \quad [\text{Box 1.1 (iii)}]$$

(viii)  $\mu(\Omega) = \mu(A \cup A^c) = \mu(A) + \mu(A^c)$ . Hence, if  $\mu(\Omega) = \mu(A) < \infty$ , then  $\mu(A^c) = 0$ . Therefore, for all  $B \in \mathcal{A}$ , (v) implies  $\mu(A^c \cap B) = 0$ . Furthermore,  $B = (A \cap B) \cup (A^c \cap B)$  and  $(A \cap B) \cap (A^c \cap B) = \emptyset$ . Hence,  $\mu(B) = \mu(A \cap B) + \mu(A^c \cap B) = \mu(A \cap B)$ . Note that, in general,  $\mu(A) = \mu(\Omega)$  does not imply  $A = \Omega$ .

(ix)  $\mu(A) = 0$  implies

$$\begin{aligned} \mu(B) &= \mu(A) + \mu(B) \\ &\geq \mu(A \cup B) \quad [(xi)] \\ &\geq \mu(B). \quad [(v)] \end{aligned}$$

Note that, in general,  $\mu(A) = 0$  does not imply  $A = \emptyset$ .

(x) Let  $B := \Omega \setminus \Omega_0$ . Then  $\mu(B) = 0$  as well as  $\mu(A \cap B) = 0$  for all  $A \in \mathcal{A}$  [see Box 1.1 (v)]. Furthermore, for  $A \in \mathcal{A}$ :  $A = (A \cap \Omega_0) \cup (A \cap B)$ , where  $A \cap \Omega_0$  and  $A \cap B$  are disjoint. Now, the sets  $A \cap \Omega_0$ ,  $A \in \mathcal{A}$ , are the elements of the trace  $\sigma$ -algebra and  $(\Omega_0, \mathcal{A}|_{\Omega_0}) = (\Omega_0, \mathcal{P}(\Omega_0))$ . Therefore we can apply Equation (1.31). Hence, for all  $A \in \mathcal{A}$ ,

$$\begin{aligned} \mu(A) &= \mu(A \cap \Omega_0) + \mu(A \cap B) && \text{[Box 1.1 (ii)]} \\ &= \sum_{\omega \in A \cap \Omega_0} \mu(\{\omega\}) + \mu(A \cap B) && \text{[(1.31)]} \\ &= \sum_{\omega \in A \cap \Omega_0} \mu(\{\omega\}). && [\mu(A \cap B) = 0] \end{aligned}$$

(xi) Let  $A_1, A_2, \dots \in \mathcal{A}$  and define  $B_1, B_2, \dots \in \mathcal{A}$  by  $B_1 = A_1$ , and  $B_i = A_i \setminus \bigcup_{j=1}^{i-1} B_j$  for  $i > 1$  (see Rem. 1.46). Then  $B_1, B_2, \dots$  is a sequence of pairwise disjoint sets with  $B_i \subset A_i$  for all  $i \in \mathbb{N}$  and  $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ . Hence,

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mu\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum_{i=1}^{\infty} \mu(B_i) && \text{[Def. 1.43 (c)]} \\ &\leq \sum_{i=1}^{\infty} \mu(A_i). && \text{[Box 1.1 (v)]} \end{aligned}$$

▷ **Solution 1-19** If the  $A_1, \dots, A_n \in \mathcal{A}$  are pairwise disjoint and  $B \in \mathcal{A}$ , then, for  $i \neq j$ ,  $i, j = 1, \dots, n$ ,

$$(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j) = B \cap \emptyset = \emptyset.$$

Hence, the sets  $B \cap A_1, \dots, B \cap A_n$  are pairwise disjoint. Furthermore, condition (b) of Remark 1.47 implies

$$\bigcup_{i=1}^n (B \cap A_i) = B \cap \bigcup_{i=1}^n A_i = B.$$

Therefore, additivity of  $\mu$  yields

$$\mu(B) = \mu\left(\bigcup_{i=1}^n (B \cap A_i)\right) = \sum_{i=1}^n \mu(B \cap A_i),$$

which is Equation (1.29). The proof of Equation (1.30) is literally the same except for replacing  $\bigcup_{i=1}^n$  by  $\bigcup_{i=1}^{\infty}$ ,  $\sum_{i=1}^n$  by  $\sum_{i=1}^{\infty}$ , and additivity of  $\mu$  by  $\sigma$ -additivity.

▷ **Solution 1-20** Let  $\omega \in \Omega$ .

(a) According to Equation (1.32),  $\delta_{\omega}(\emptyset) = 1_{\emptyset}(\omega) = 0$ .

(b) According to Equation (1.32),  $\delta_{\omega}(A) = 1_A(\omega) \in \{0, 1\}$ , for all  $A \in \mathcal{A}$ , and this implies  $\delta_{\omega}(A) \geq 0$ , for all  $A \in \mathcal{A}$ .

(c) If  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint, then

$$\begin{aligned} \delta_{\omega}\left(\bigcup_{i=1}^{\infty} A_i\right) &= 1_{\bigcup_{i=1}^{\infty} A_i}(\omega) && \text{[(1.32)]} \\ &= \sum_{i=1}^{\infty} 1_{A_i}(\omega) && \text{[(1.37)]} \\ &= \sum_{i=1}^{\infty} \delta_{\omega}(A_i). && \text{[(1.32)]} \end{aligned}$$

- ▷ **Solution 1-21** (a) According to Equation (1.40),  $\mu_{\#}(\emptyset) = \sum_{\omega \in \Omega} 1_{\emptyset}(\omega) = 0$ .  
 (b) According to Equation (1.40),  $\mu_{\#}(A) = \sum_{\omega \in \Omega} 1_A(\omega)$ , for all finite  $A \in \mathcal{A}$ , and  $\mu_{\#}(A) = \infty$ , if  $A$  is infinite. This implies  $\mu_{\#}(A) \geq 0$ , for all  $A \in \mathcal{A}$ .  
 (c) If  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint and all  $A_i$  are finite, then

$$\begin{aligned} \mu_{\#}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{\omega \in \Omega} 1_{\bigcup_{i=1}^{\infty} A_i}(\omega) \quad [(1.40)] \\ &= \sum_{\omega \in \Omega} \sum_{i=1}^{\infty} 1_{A_i}(\omega) \quad [(1.37)] \\ &= \sum_{i=1}^{\infty} \sum_{\omega \in \Omega} 1_{A_i}(\omega) \\ &= \sum_{i=1}^{\infty} \mu_{\#}(A_i). \quad [(1.40)] \end{aligned}$$

Note that the set  $\bigcup_{i=1}^{\infty} A_i$  can be countably infinite even if all  $A_i$  are finite. In this case  $\mu_{\#}\left(\bigcup_{i=1}^{\infty} A_i\right) = \infty = \sum_{i=1}^{\infty} \mu_{\#}(A_i)$ . If at least one of the  $A_i$  is infinite, then  $\bigcup_{j=1}^{\infty} A_j \supset A_i$  is an infinite set and  $\mu_{\#}\left(\bigcup_{j=1}^{\infty} A_j\right) \geq \mu_{\#}(A_i)$  is infinite as well.

- ▷ **Solution 1-22** (a) Using Equations (1.42) and (1.39),

$$\left(\sum_{\omega \in B} \delta_{\omega}\right)(\emptyset) = \sum_{\omega \in B} \delta_{\omega}(\emptyset) = \sum_{\omega \in B} 1_{\emptyset}(\omega) = \sum_{\omega \in B} 0 = 0.$$

- (b) Using Equations (1.42) and (1.39),

$$\forall A \in \mathcal{A}: \left(\sum_{\omega \in B} \delta_{\omega}\right)(A) = \sum_{\omega \in B} \delta_{\omega}(A) = \sum_{\omega \in B} 1_A(\omega) \geq 0.$$

- (c) If  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint, then

$$\begin{aligned} \left(\sum_{\omega \in B} \delta_{\omega}\right)\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{\omega \in B} \delta_{\omega}\left(\bigcup_{i=1}^{\infty} A_i\right) \quad [(1.42)] \\ &= \sum_{\omega \in B} 1_{\bigcup_{i=1}^{\infty} A_i}(\omega) \quad [(1.39)] \\ &= \sum_{\omega \in B} \sum_{i=1}^{\infty} 1_{A_i}(\omega) \quad [(1.37)] \\ &= \sum_{\omega \in B} \sum_{i=1}^{\infty} \delta_{\omega}(A_i) \quad [(1.39)] \\ &= \sum_{i=1}^{\infty} \left(\sum_{\omega \in B} \delta_{\omega}\right)(A_i). \quad [(1.42)] \end{aligned}$$

- ▷ **Solution 1-23** (a) Equation (1.47) yields:  $\nu(\emptyset) = \mu(\emptyset) = 0$ .  
 (b) Equation (1.47) also yields:  $\nu(A) = \mu(A) \geq 0$ , for all  $A \in \mathcal{C}$ .  
 (c) If  $A_1, A_2, \dots \in \mathcal{C}$  are pairwise disjoint, then

$$\begin{aligned} \nu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \quad [\text{Def. 1.1 (c), (1.47)}] \\ &= \sum_{i=1}^{\infty} \mu(A_i) \quad [\text{Def. 1.43 (c)}] \\ &= \sum_{i=1}^{\infty} \nu(A_i). \quad [(1.47)] \end{aligned}$$

▷ **Solution 1-24** (a) Using Equation (1.48) and Definition Def. 1.43 (a) yields

$$\left( \sum_{i=1}^{\infty} \alpha_i \mu_i \right) (\emptyset) = \sum_{i=1}^{\infty} \alpha_i \mu_i (\emptyset) = \sum_{i=1}^{\infty} 0 = 0.$$

(b) Similarly, using Equation (1.48) yields, for all  $A \in \mathcal{A}$ ,

$$\left( \sum_{i=1}^{\infty} \alpha_i \mu_i \right) (A) = \sum_{i=1}^{\infty} \alpha_i \mu_i (A) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \alpha_i \mu_i (A) \geq 0,$$

because  $\mu_i(A) \geq 0$  and we assume  $\alpha_i \geq 0$ .

(c) If  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint, then

$$\begin{aligned} \left( \sum_{i=1}^{\infty} \alpha_i \mu_i \right) \left( \bigcup_{j=1}^{\infty} A_j \right) &= \sum_{i=1}^{\infty} \alpha_i \mu_i \left( \bigcup_{j=1}^{\infty} A_j \right) && [(1.48)] \\ &= \sum_{i=1}^{\infty} \alpha_i \sum_{j=1}^{\infty} \mu_i (A_j) && [\text{Def. 1.43 (c)}] \\ &= \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \alpha_i \mu_i (A_j) \\ &= \sum_{j=1}^{\infty} \left( \sum_{i=1}^{\infty} \alpha_i \mu_i \right) (A_j). && [(1.48)] \end{aligned}$$

Note that the last but one equation holds, because rearranging summands does not change the sum if the terms  $\alpha_i$  and  $\mu_i(A_j)$  are nonnegative.



## Chapter 2

# Measurable Mapping

In chapter 1 we treated the concepts of a  $\sigma$ -algebra and a  $\sigma$ -algebra generated by a set system on a set  $\Omega$ . An element  $A$  of a  $\sigma$ -algebra  $\mathcal{A}$  has been called a *measurable set*. We also introduced the concept of a *measure*, which assigns a nonnegative real number or  $\infty$  to all elements of a  $\sigma$ -algebra. This chapter is devoted to the concept of a *measurable mapping*, related concepts such as the  $\sigma$ -algebra generated by a mapping, and the *image measure* of  $\mu$  under  $f$ , the measure induced by a measurable mapping  $f$  on its codomain space. All these concepts play an important role in integration and probability theory. In probability theory, a measurable set is called an *event*, a measurable mapping  $f$  is called a *random variable* and the image measure of the probability measure  $P$  under  $f$  is called the *distribution of  $f$* .

### 2.1 Image and Inverse Image

Two key concepts of this chapter are the *image* of a set  $A \subset \Omega$  and the *inverse image* of a set  $A' \subset \Omega'$  under a mapping  $f: \Omega \rightarrow \Omega'$ . We start with the formal definitions and then illustrate these concepts in section 2.2.

**Definition 2.1 (Image and Inverse Image)**

Let  $\Omega, \Omega'$  denote two sets and  $f: \Omega \rightarrow \Omega'$  a mapping. Then we call

$$f(A) := \{f(\omega) : \omega \in A\}, \quad A \subset \Omega, \quad (2.1)$$

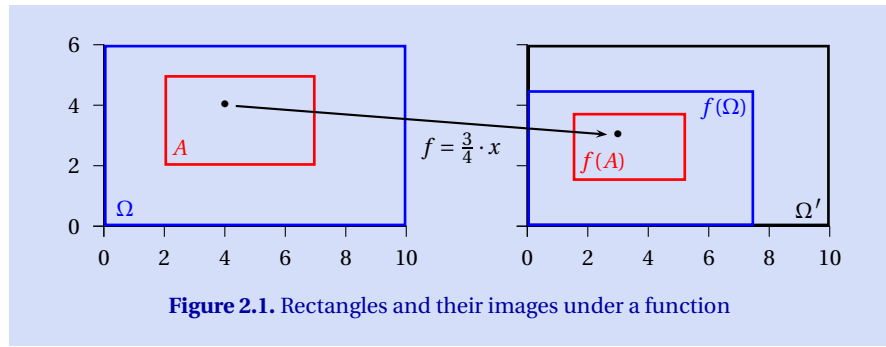
the *image of  $A$  under  $f$* , and

$$f^{-1}(A') := \{\omega \in \Omega : f(\omega) \in A'\}, \quad A' \subset \Omega', \quad (2.2)$$

the *inverse image of  $A'$  under  $f$* .

Whereas the image  $f(A)$  is a subset of  $\Omega'$ , the inverse image  $f^{-1}(A')$  is the set of all elements of the *domain*  $\Omega$  for which  $f$  takes on a value in the subset  $A'$  of its *codomain*  $\Omega'$ . For convenience, we also use the notation

$$\{f \in A'\} := f^{-1}(A') \quad \text{and} \quad \{f = \omega'\} := f^{-1}(\{\omega'\}). \quad (2.3)$$



**Remark 2.2 (Properties of Inverse Images)** Let  $f: \Omega \rightarrow \Omega'$  be a mapping,  $I$  be an index set,  $A' \subset \Omega'$ , and  $(A'_i, i \in I)$  a family of subsets  $A'_i$  of  $\Omega'$ . Then

$$f^{-1}[(A')^c] = [f^{-1}(A')]^c, \quad (2.4)$$

$$f^{-1}\left(\bigcap_{i \in I} A'_i\right) = \bigcap_{i \in I} f^{-1}(A'_i), \quad (2.5)$$

$$f^{-1}\left(\bigcup_{i \in I} A'_i\right) = \bigcup_{i \in I} f^{-1}(A'_i) \quad (2.6)$$

(see Exercise 2-1). Note that, in general, the corresponding properties do not necessarily hold for the image  $f(A)$ ,  $A \subset \Omega$ .  $\triangleleft$

## 2.2 Introductory Examples

### 2.2.1 Example 1: Rectangles

Our first example deals with rectangles, their *images*, and their *inverse images* under a mapping  $f$ .

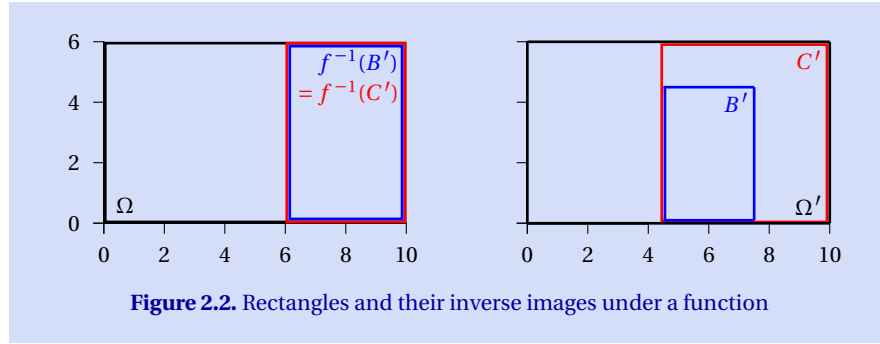
#### The Measurable Space

Let  $[a, b]$ ,  $a, b \in \mathbb{R}$ , denote the closed interval between  $a$  and  $b$ , inclusively, and consider the two rectangles

$$\Omega = [0, 10] \times [0, 6] \quad \text{and} \quad A = [2, 7] \times [2, 5]$$

depicted on the left-hand side of Figure 2.1. The elements of  $\Omega$  and  $A$  are points  $x = (x_1, x_2)$  in these rectangles with coordinates  $x_1$  on the horizontal axis and  $x_2$  on the vertical axis. Furthermore, let us consider a  $\sigma$ -algebra on  $\Omega$ ,

$$\mathcal{A} = \{\Omega, \emptyset, A, A^c\}.$$



### The Mapping and the Image

Consider the set  $\Omega' = \Omega$  and the function  $f: \Omega \rightarrow \Omega'$  defined by

$$f(x) = \frac{3}{4} \cdot x = \left( \frac{3}{4} \cdot x_1, \frac{3}{4} \cdot x_2 \right), \quad \forall x \in \Omega. \quad (2.7)$$

Hence,  $f$  maps all points  $x = (x_1, x_2) \in \Omega$  to the points  $f(x_1, x_2) \in \Omega'$ . This is illustrated by Figure 2.1 for the point  $x = (4, 4)$ , which is mapped to  $f(x) = (3, 3)$ . The right-hand side of Figure 2.1 also depicts the *image of A under f*, i. e.,  $f(A) = \{f(x): x \in A\}$ , as well as the image  $f(\Omega)$  of  $\Omega$  under  $f$ .

### The Inverse Images

We specify the  $\sigma$ -algebra

$$\mathcal{A}' = \{\Omega', \emptyset, B', (B')^c\}$$

on  $\Omega'$ , where

$$B' = ]4.5, 7.5] \times ]0, 4.5]$$

is the rectangle depicted on the right-hand side of Figure 2.2, and  $(B')^c = \Omega' \setminus B'$  is its complement.

Now we consider the *inverse image* of  $B'$  under  $f$  [see Eq. (2.7)], i. e.,

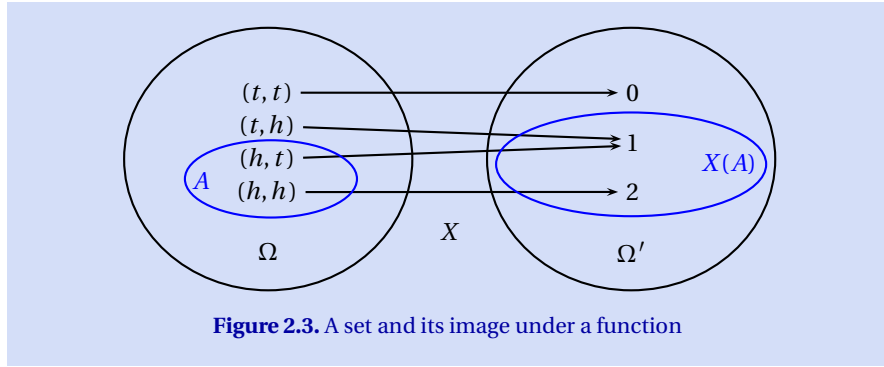
$$f^{-1}(B') = ]6, 10] \times ]0, 6]$$

(see Fig. 2.2). It is the rectangle on the right side of  $\Omega$ . For further examples see Exercises 2-2 and 2-3.

Also consider the inverse image of the rectangle

$$C' = ]4.5, 10] \times ]0, 6]$$

(see Fig. 2.2). Its inverse image under  $f$  is identical to the inverse image  $f^{-1}(B')$ , i. e.,



$$f^{-1}(C') = f^{-1}(B'),$$

which follows from

$$\begin{aligned} f^{-1}(C') &= f^{-1}[B' \cup (C' \setminus B')] && [B' \subset C', \text{ Fig. 2.2}] \\ &= f^{-1}(B') \cup f^{-1}(C' \setminus B') && [(2.6)] \\ &= f^{-1}(B') \cup \emptyset = f^{-1}(B'). \end{aligned}$$

Note that  $f^{-1}(C' \setminus B') = \emptyset$ , because  $f$  has been defined on  $\Omega = [0, 10] \times [0, 6]$ . If we would define  $f$  on  $\Omega = \mathbb{R}^2$ , then the set  $f^{-1}(C' \setminus B')$  would *not* be empty. (See also Exercise 2-4.)

### 2.2.2 Example 2: Flipping two Coins

Now we consider the random experiment of *flipping two coins*.

#### The Measurable Space

In this random experiment, the set of possible outcomes is

$$\Omega = \{(h, h), (h, t), (t, h), (t, t)\}.$$

This set consists of four elements (pairs). For example, the first component of the pair  $(h, t)$  represents the outcome of flipping  $h = \text{heads}$  with the first coin and the second component represents the outcome of flipping  $t = \text{tails}$  with the second coin. As a  $\sigma$ -algebra on  $\Omega$  we consider the power set  $\mathcal{A} = \mathcal{P}(\Omega)$ .

#### The Mapping

Consider the function  $X: \Omega \rightarrow \Omega' = \{0, 1, 2\}$  defined by

$$X[(t, t)] = 0, \quad X[(t, h)] = 1, \quad X[(h, t)] = 1, \quad \text{and} \quad X[(h, h)] = 2.$$

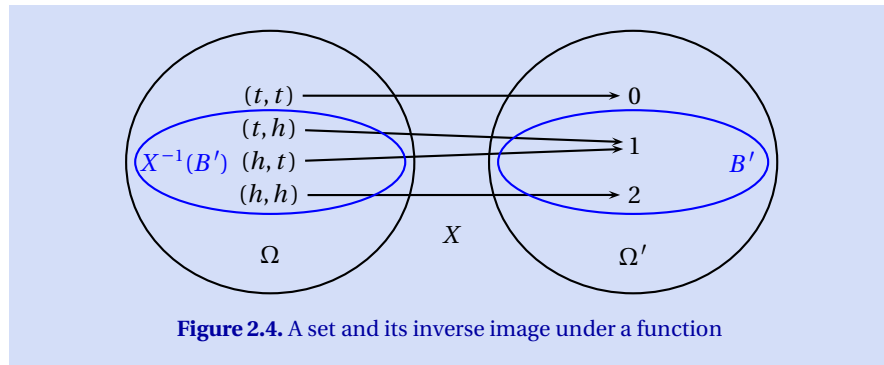


Figure 2.4. A set and its inverse image under a function

Table 2.1. Example of Measurable Sets Represented by a Mapping  $X$

$\{X \in \Omega'\}$	$= X^{-1}(\Omega')$	$= \Omega$	0, 1, or 2 heads are flipped
$\{X \in \emptyset\}$	$= X^{-1}(\emptyset)$	$= \emptyset$	neither 0, 1, nor 2 heads are flipped
$\{X = 0\}$	$= X^{-1}(\{0\})$	$= \{(t, t)\}$	no heads are flipped.
$\{X = 1\}$	$= X^{-1}(\{1\})$	$= \{(h, t), (t, h)\}$	heads are flipped exactly once
$\{X = 2\}$	$= X^{-1}(\{2\})$	$= \{(h, h)\}$	two heads are flipped
$\{X \in \{0, 1\}\}$	$= X^{-1}(\{0, 1\})$	$= \{(h, t), (t, h), (t, t)\}$	not more than one heads are flipped.
$\{X \in \{0, 2\}\}$	$= X^{-1}(\{0, 2\})$	$= \{(h, h), (t, t)\}$	either two heads or no heads at all are flipped
$\{X \in \{1, 2\}\}$	$= X^{-1}(\{1, 2\})$	$= \{(h, h), (h, t), (t, h)\}$	at least one heads is flipped

Looking at this assignment rule shows that this function may be called *number of flipping heads*. Again, we consider the *image of a set*  $A \subset \Omega$  under  $X$ , i. e.,  $X(A) = \{X(\omega) : \omega \in A\}$ ,  $A \subset \Omega$ . For example, for  $A = \{(h, h), (h, t)\}$ , the image under  $X$  is  $X(A) = \{1, 2\}$  (see Fig. 2.3).

### The Inverse Images

Suppose  $\mathcal{A}' = \mathcal{P}(\Omega')$  is the power set of  $\Omega' = \{0, 1, 2\}$ . In this example there are also  $2^3 = 8$  inverse images  $X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\}$ ,  $A' \in \mathcal{A}'$ . Three of these eight inverse images are:

$$X^{-1}(\{0\}) = \{(t, t)\}, \quad X^{-1}(\{1\}) = \{(h, t), (t, h)\}, \quad X^{-1}(\{2\}) = \{(h, h)\}.$$

These are the events that  $X$  takes on the value 0, 1, and 2, respectively. (In order to identify the inverse images listed above, trace back the arrows from right to left in Figure 2.4.) Furthermore, consider the inverse images

$$\begin{aligned} X^{-1}(\{0,1\}) &= \{(t,t), (h,t), (t,h)\}, \\ X^{-1}(\{0,2\}) &= \{(t,t), (h,h)\}, \\ X^{-1}(\{1,2\}) &= \{(h,t), (t,h), (h,h)\}. \end{aligned}$$

These are the events that  $X$  takes on a value in the sets  $\{0,1\}$ ,  $\{0,2\}$ , and  $\{1,2\}$ , respectively. One of these inverse images, namely  $X^{-1}(B')$ , with  $B' := \{1,2\}$ , is represented in Figure 2.4. Finally,

$$X^{-1}(\Omega') = \Omega \quad \text{and} \quad X^{-1}(\emptyset) = \emptyset.$$

Hence, we listed all eight inverse images  $X^{-1}(A')$ ,  $A' \in \mathcal{A}'$ . They are the eight measurable sets that can be represented by the mapping  $X$  and the  $\sigma$ -algebra  $\mathcal{A}' = \mathcal{P}(\Omega')$ . These sets are listed in Table 2.1, using the notation  $\{X \in A'\} := X^{-1}(A')$ ,  $A' \in \mathcal{A}'$ , and  $\{X = x\} := X^{-1}(\{x\})$ ,  $\{x\} \in \mathcal{A}'$  [see Eq. (2.3)].

### 2.3 Measurable Mapping

Now we define the concept of a *measurable mapping* and related concepts such as the  *$\sigma$ -algebra generated by a mapping* and *measurability of a mapping with respect to a mapping*.

**Remark 2.3 (Mapping)** Remember, a mapping  $f: \Omega \rightarrow \Omega'$  assigns to *all*  $\omega \in \Omega$  a unique  $f(\omega) \in \Omega'$ . Hence,  $f$  is, by definition, a subset of the Cartesian product  $\Omega \times \Omega'$ , i. e.,  $f = \{(\omega, f(\omega)) : \omega \in \Omega\}$ . This implies that, instead of  $f: \Omega \rightarrow \Omega'$ , we can also write  $f: \Omega \rightarrow \Omega''$  for the *same mapping*, provided that  $f(\Omega) \subset \Omega''$ .  $\triangleleft$

**Remark 2.4 (Identical Mappings)** If  $f, g: \Omega \rightarrow \Omega'$  are two mappings, then

$$f = g \quad \Leftrightarrow \quad \{(\omega, f(\omega)) : \omega \in \Omega\} = \{(\omega, g(\omega)) : \omega \in \Omega\}. \quad (2.8)$$

If  $f = g$  we say that the two mappings are *identical*. Hence, even if  $f: \Omega \rightarrow \Omega'$  and  $g: \Omega \rightarrow \Omega''$  are mappings with  $\Omega' \neq \Omega''$ , it is still possible that  $f$  and  $g$  are identical. Note that (2.8) also implies: If, for  $f: \Omega \rightarrow \Omega'$  and  $g: \Omega \rightarrow \Omega''$ , we write  $f, g: \Omega \rightarrow \Omega'''$  with  $\Omega''' := \Omega' \cup \Omega''$ , then  $f$  and  $g$  remain unchanged.  $\triangleleft$

#### 2.3.1 Measurable Mapping

Now the core concept of this chapter is defined as follows:

**Definition 2.5 (Measurable Mapping)**

Let  $(\Omega, \mathcal{A})$ ,  $(\Omega', \mathcal{A}')$  be measurable spaces and let  $f: \Omega \rightarrow \Omega'$  be a mapping. Then  $f$  is called  $(\mathcal{A}, \mathcal{A}')$ -measurable if

$$f^{-1}(A') \in \mathcal{A}, \quad \forall A' \in \mathcal{A}'.$$

**Remark 2.6 (Notation)** We use the notation

$$f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$$

to express that the mapping  $f: \Omega \rightarrow \Omega'$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable. If there is no ambiguity about  $\mathcal{A}'$ , we also say that  $f$  is  $\mathcal{A}$ -measurable or measurable with respect to  $\mathcal{A}$ .  $\triangleleft$

### Examples

**Example 2.7 (Rectangles – continued)** In Example 2.2.1, we considered the mapping  $f: \Omega \rightarrow \Omega' = \Omega$  defined by  $f(x) = \frac{3}{4}x$ . Furthermore, we considered the rectangle  $B' = ]4.5, 7.5] \times [0, 4.5]$  and the inverse image

$$f^{-1}(B') = ]6, 10] \times [0, 6].$$

If  $A = [2, 7] \times [2, 5]$ , then the inverse image  $f^{-1}(B')$  is not an element of the  $\sigma$ -algebra  $\mathcal{A} = \{\Omega, \emptyset, A, A^c\}$ . In this example, we also specified the  $\sigma$ -algebra  $\mathcal{A}' = \{\Omega', \emptyset, B', (B')^c\}$ . Hence,  $f$  is not  $(\mathcal{A}, \mathcal{A}')$ -measurable. However, if we specify a  $\sigma$ -algebra  $\mathcal{C}$  such that  $f^{-1}(B') \in \mathcal{C}$ , then  $f$  is  $(\mathcal{C}, \mathcal{A}')$ -measurable. As we see later on, this condition is sufficient for  $f$  to be  $(\mathcal{C}, \mathcal{A}')$ -measurable (see Th. 2.20).  $\triangleleft$

**Example 2.8 (Flipping two Coins – continued)** In Example 2.2.2, we considered the mapping  $X = \text{number of flipping heads}$  and in Table 2.1 (p. 45) we listed all inverse images  $X^{-1}(A')$ ,  $A' \in \mathcal{A}' = \mathcal{P}(\{0, 1, 2\})$ . Of course,  $\mathcal{A} = \mathcal{P}(\Omega)$  ensures that all inverse images  $X^{-1}(A')$ ,  $A' \in \mathcal{A}'$ , are elements of  $\mathcal{A}$ .

However, instead of  $\mathcal{A} = \mathcal{P}(\Omega)$ , we might consider the  $\sigma$ -algebra

$$\mathcal{A}_0 = \{\Omega, \emptyset, \{(h, h), (h, t)\}, \{(t, h), (t, t)\}\}.$$

The element  $\{(h, h), (h, t)\}$  represents the event that *heads* are flipped in the first flip and  $\{(t, h), (t, t)\}$  is the event that *tails* are flipped in the first flip. Hence, the  $\sigma$ -algebra  $\mathcal{A}_0$  contains the events that refer to the outcome of the *first flip* only, whereas  $X$  represents the number of heads in *both coin flips*. If we choose  $\mathcal{A}'$  to be the power set of  $\Omega' = \{0, 1, 2\}$ , then it is *not* true that all eight inverse images  $X^{-1}(A')$ ,  $A' \in \mathcal{A}'$ , are elements of  $\mathcal{A}_0$ . The inverse image  $X^{-1}(\{2\}) = \{(h, h)\}$ , e. g., is not an element of  $\mathcal{A}_0$ . Hence, if we consider the measurable spaces  $(\Omega, \mathcal{A}_0)$  and  $(\Omega', \mathcal{P}(\Omega'))$ , then the mapping  $X$  is *not*  $(\mathcal{A}_0, \mathcal{P}(\Omega'))$ -measurable. Hence, in some sense  $\mathcal{A}_0$  is ‘not well-adapted’ to  $X$ .  $\triangleleft$

**Example 2.9 (Two Trivial Cases)** If (a)  $\mathcal{A} = \mathcal{P}(\Omega)$  is the power set of  $\Omega$  or if (b)  $\mathcal{A}' = \{\Omega', \emptyset\}$ , then every mapping  $f: \Omega \rightarrow \Omega'$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable. This is easily seen as follows: (a) If  $\mathcal{A} = \mathcal{P}(\Omega)$  is the power set of  $\Omega$ , then all inverse images  $f^{-1}(A')$ ,  $A' \subset \Omega'$ , are elements in  $\mathcal{A} = \mathcal{P}(\Omega)$ , because it is the set of *all* subsets of  $\Omega$ . (b) If  $\mathcal{A}' = \{\Omega', \emptyset\}$ , then every mapping  $f: \Omega \rightarrow \Omega'$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable, because  $f^{-1}(\Omega') = \Omega$  and  $f^{-1}(\emptyset) = \emptyset$ . Again, the inverse images  $\Omega$  and  $\emptyset$  are both elements in every  $\sigma$ -algebra on  $\Omega$ . Hence, in both cases, (a) and (b), every mapping  $f: \Omega \rightarrow \Omega'$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable.  $\triangleleft$

**Example 2.10 (Constant Mapping)** A constant mapping  $f: \Omega \rightarrow \Omega'$  is defined by

$$f(\omega) = \omega', \quad \forall \omega \in \Omega,$$

where  $\omega'$  is a fixed element of  $\Omega'$ . Such a constant mapping is  $(\mathcal{A}, \mathcal{A}')$ -measurable for any  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$  and any  $\sigma$ -algebra  $\mathcal{A}'$  on  $\Omega'$ . This is true, because for all subsets  $A'$  of  $\Omega'$ : If  $\omega' \in A'$ , then  $f^{-1}(A') = \Omega$ . If, in contrast,  $\omega' \notin A'$ , then  $f^{-1}(A') = \emptyset$ . However,  $\Omega$  and  $\emptyset$  are elements of *all*  $\sigma$ -algebras on  $\Omega$ .  $\triangleleft$

**Example 2.11 (Identity Mapping)** The identity mapping  $id: \Omega \rightarrow \Omega$  defined by

$$id(\omega) = \omega, \quad \forall \omega \in \Omega,$$

is  $(\mathcal{A}, \mathcal{A}_0)$ -measurable for any pair of  $\sigma$ -algebras on  $\Omega$  with  $\mathcal{A}_0 \subset \mathcal{A}$ . This is easily seen as follows:

$$id^{-1}(A) = A, \quad \forall A \in \mathcal{A}_0.$$

Because we assume  $\mathcal{A}_0 \subset \mathcal{A}$ , we can conclude that  $id$  is  $(\mathcal{A}, \mathcal{A}_0)$ -measurable.  $\triangleleft$

**Example 2.12 (Indicator of a Measurable Set)** Let  $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$  be two measurable spaces, where  $\mathcal{A}'$  is any  $\sigma$ -algebra on  $\Omega' \subset \mathbb{R}$  with  $\{0\}, \{1\} \in \mathcal{A}'$ . Then the indicator  $1_A: \Omega \rightarrow \Omega'$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable *if and only if*  $A \in \mathcal{A}$ . Note that the requirement  $\{0\}, \{1\} \in \mathcal{A}'$  is not only satisfied by  $(\Omega', \mathcal{A}') := (\{0, 1\}, \mathcal{P}(\{0, 1\}))$ , but also by  $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B})$  and by  $(\Omega', \mathcal{A}') = (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ , where  $\mathcal{B}$  denotes the Borel  $\sigma$ -algebra on  $\mathbb{R}$  and  $\overline{\mathcal{B}}$  the Borel  $\sigma$ -algebra on  $\overline{\mathbb{R}}$ .  $\triangleleft$

**Example 2.13 (Indicators of Unions and Intersections)** If  $(\Omega, \mathcal{A})$  is a measurable space and  $A, B \in \mathcal{A}$ , then  $1_{A \cap B}$  and  $1_{A \cup B}$  are  $(\mathcal{A}, \mathcal{B})$ -measurable. This follows from the fact that  $A \cap B \in \mathcal{A}$  and  $A \cup B \in \mathcal{A}$ . For the same reason,  $A_1, A_2, \dots \in \mathcal{A}$  implies that  $1_{\bigcup_{i=1}^{\infty} A_i}$  is  $(\mathcal{A}, \mathcal{B})$ -measurable.  $\triangleleft$

**Example 2.14 (Constant Function)** Assume that  $(\Omega, \mathcal{A})$  and  $(\Omega', \mathcal{A}')$  are measurable spaces such that  $\{\omega'\} \in \mathcal{A}'$ , for all  $\omega' \in \Omega'$ . Furthermore, let  $f: \Omega \rightarrow \Omega'$ . If  $\mathcal{A} = \{\Omega, \emptyset\}$ , then  $f$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable if and only if  $f$  is a constant function, i. e., if and only if there is an  $\omega' \in \Omega'$  such that  $f(\omega) = \omega'$ , for all  $\omega \in \Omega$  (see Exercise 2-5). Note that for  $(\Omega', \mathcal{A}') = (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ ,  $\{x\} \in \overline{\mathcal{B}}$ , for all  $x \in \overline{\mathbb{R}}$ .  $\triangleleft$

**Example 2.15 (Dichotomous Function)** If  $\mathcal{A} = \{\Omega, \emptyset, A, A^c\}$  with  $A \subset \Omega$ , then  $f: \Omega \rightarrow \mathbb{R}$  is  $(\mathcal{A}, \mathcal{B})$ -measurable if and only if  $f = \alpha_1 1_A + \alpha_2 1_{A^c}$  for  $\alpha_1, \alpha_2 \in \mathbb{R}$  (see Exercise 2-6).  $\triangleleft$

### Step Function

Another important example of a measurable function is a step function, which is defined as follows:

**Definition 2.16 (Step Function)**

Let  $A_1, \dots, A_n$ ,  $n \in \mathbb{N}$ , be a finite sequence of subsets of a set  $\Omega$ . Then a finite linear combination

$$f = \sum_{i=1}^n \alpha_i 1_{A_i}, \quad \alpha_1, \dots, \alpha_n \in \mathbb{R}, \quad (2.9)$$

is called a step function.

**Remark 2.17 (Step Function and a Partition of  $\Omega$ )** If the sets  $A_1, \dots, A_n$  are pairwise disjoint, if we define  $A_{n+1} := \Omega \setminus (\bigcup_{i=1}^n A_i)$ , then  $\{A_1, \dots, A_n, A_{n+1}\}$  is a finite partition of  $\Omega$ . Furthermore, if  $f$  satisfies (2.9), and  $\alpha_{n+1} := 0$ , then, for all  $A' \subset \mathbb{R}$ ,

$$f^{-1}(A') = \bigcup_{\substack{i=1, \dots, n+1, \\ \alpha_i \in A'}} A_i \quad (2.10)$$

(see Exercise 2-7). ◁

**Remark 2.18 (Measurability of a Step Function)** If  $(\Omega, \mathcal{A})$  is a measurable space and  $A_1, \dots, A_n \in \mathcal{A}$ , then the step function  $f: \Omega \rightarrow \mathbb{R}$  defined by Equation (2.9) is  $(\mathcal{A}, \mathcal{B})$ -measurable (see Exercise 2-8). ◁

**Lemma 2.19 (Measurability if  $\mathcal{A}$  is Countably Generated)**

Let  $(\Omega, \mathcal{A})$  be a measurable space and let  $\mathcal{A} = \sigma(\mathcal{E})$ , where  $\mathcal{E}$  is a finite (i. e.,  $\mathcal{E} = \{A_1, \dots, A_n\}$ ) or countable (i. e.,  $\mathcal{E} = \{A_1, A_2, \dots\}$ ) partition of  $\Omega$ . Then  $f: \Omega \rightarrow \overline{\mathbb{R}}$  is  $(\mathcal{A}, \overline{\mathcal{B}})$ -measurable if and only if there are  $\alpha_1, \alpha_2, \dots \in \overline{\mathbb{R}}$  such that  $f = \sum_i \alpha_i 1_{A_i}$ , where  $i \in \{i = 1, \dots, n\}$  or  $i \in \mathbb{N}$ , respectively.

(Proof p. 70)

**A Necessary and Sufficient Condition of Measurability**

Let  $(\Omega', \mathcal{A}')$  be a measurable space and  $\mathcal{E}' \subset \mathcal{A}'$ . Then we denote

$$f^{-1}(\mathcal{E}') := \{f^{-1}(A') : A' \in \mathcal{E}'\}. \quad (2.11)$$

This notation is used in the following theorem, which can be utilized for proving  $(\mathcal{A}, \mathcal{A}')$ -measurability of a mapping  $f: \Omega \rightarrow \Omega'$ .

**Theorem 2.20 (Measurable Mapping and Generating Systems)**

Let  $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$  denote measurable spaces, let  $\mathcal{E}' \subset \mathcal{A}'$ , and  $f: \Omega \rightarrow \Omega'$ . Then

$$\sigma[f^{-1}(\mathcal{E}')] = f^{-1}[\sigma(\mathcal{E}')]. \quad (2.12)$$

Furthermore, if  $\sigma(\mathcal{E}') = \mathcal{A}'$ , then  $f$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable if and only if  $f^{-1}(A') \in \mathcal{A}$ , for all  $A' \in \mathcal{E}'$ .

For a proof see Klenke (2013, Theorem 1.81, p. 33).

Now consider a finite or countable set  $\Omega'$ . Then Theorem 2.20 and Example 1.18 immediately imply the following corollary:

**Corollary 2.21 (Finite or Countable Generating Systems)**

Let  $(\Omega, \mathcal{A})$ ,  $(\Omega', \mathcal{P}(\Omega'))$  be measurable spaces, where  $\Omega'$  is finite or countable, and let  $\mathcal{E}' = \{\{\omega'\}: \omega' \in \Omega'\}$ . Then a mapping  $f: \Omega \rightarrow \Omega'$  is  $(\mathcal{A}, \mathcal{P}(\Omega'))$ -measurable if and only if  $f^{-1}(\{\omega'\}) \in \mathcal{A}$ , for all  $\omega' \in \Omega'$ .

**Example 2.22 (Rectangles – continued)** In Example 2.2.1 we considered the mapping  $f: \Omega \rightarrow \Omega' = \Omega$  defined by  $f(x) = \frac{3}{4}x$ . Furthermore, we considered the rectangle  $B' = ]4.5, 7.5] \times [0, 4.5]$ . The set system

$$\mathcal{E}' = \{B'\},$$

which contains  $B'$  as the only element, generates the  $\sigma$ -algebra

$$\mathcal{A}' = \{\Omega', \emptyset, B', (B')^c\}.$$

Hence according to Theorem 2.20, the mapping  $f$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable provided that  $f^{-1}(B') \in \mathcal{A}$ .  $\triangleleft$

**Example 2.23 (Flipping two Coins – continued)** In Example 2.2.2, we defined the mapping  $X = \text{number of flipping heads}$  with codomain  $\Omega' = \{0, 1, 2\}$ . Now consider the system

$$\mathcal{E}' = \{\{0\}, \{1\}\}$$

of subsets of  $\Omega'$ . First, note that  $\sigma(\mathcal{E}') = \mathcal{P}(\Omega')$ . Therefore, Theorem 2.20 implies that  $X$  is  $(\mathcal{A}, \mathcal{P}(\Omega'))$ -measurable for each  $\sigma$ -algebra  $\mathcal{A}$  on

$$\Omega = \{(h, h), (h, t), (t, h), (t, t)\}$$

for which

$$X^{-1}(\{0\}) = \{(t, t)\} \in \mathcal{A} \quad \text{and} \quad X^{-1}(\{1\}) = \{(h, t), (t, h)\} \in \mathcal{A}.$$

This does not only hold for  $\mathcal{A}_1 = \mathcal{P}(\Omega)$ , but also for the  $\sigma$ -algebra

$$\mathcal{A}_2 = \{\Omega, \emptyset, \{(t, t)\}, \{(h, t), (t, h)\}, \{(h, h)\}, \\ \{(h, h), (h, t), (t, h)\}, \{(h, h), (t, t)\}, \{(h, t), (t, h), (t, t)\}\}.$$

As mentioned before,  $\mathcal{A}_2$  contains all events that can be represented by  $X$  (see Table 2.1, p. 45). In contrast, this does not hold for the  $\sigma$ -algebra

$$\mathcal{A}_0 = \{\Omega, \emptyset, \{(h, h), (h, t)\}, \{(t, h), (t, t)\}\}$$

(see Example 2.8). Hence,  $X$  is measurable with respect to  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , but it is not measurable with respect to  $\mathcal{A}_0$ . In this application, this means that the events  $\{(h, h), (h, t)\}$  and  $\{(t, h), (t, t)\}$  cannot be formulated in terms of  $X$ . Furthermore, some of the events that *can* be formulated in terms of  $X$  are not elements of  $\mathcal{A}_0$ . For example,  $X^{-1}(\{0\}) = \{(t, t)\}$  is *not* an element of  $\mathcal{A}_0$ .  $\triangleleft$

### 2.3.2 $\sigma$ -Algebra Generated by a Mapping

Let us consider again Example 2.2.2 and the mapping  $X = \text{number of flipping heads}$ . The set that consists of the eight inverse images  $X^{-1}(A')$ ,  $A' \in \mathcal{A}'$ , is again a  $\sigma$ -algebra on  $\Omega$ . In a sense, this  $\sigma$ -algebra carries the information associated with the mapping  $X$ ; it contains all events that can be represented by  $X$  (see Table 2.1). In the following theorem we formulate the general proposition.

**Theorem 2.24 ( $\sigma$ -Algebra Generated by a Mapping)**

Let  $f: \Omega \rightarrow \Omega'$  be a mapping and let  $(\Omega', \mathcal{A}')$  be a measurable space. Then

$$f^{-1}(\mathcal{A}') := \{f^{-1}(A') : A' \in \mathcal{A}'\} \quad (2.13)$$

is a  $\sigma$ -algebra on  $\Omega$ .

For a proof see Klenke (2013, Theorem 1.81, p. 33)).

**Remark 2.25 (Smallest  $\sigma$ -Algebra)** Note that  $f^{-1}(\mathcal{A}')$  is the smallest  $\sigma$ -algebra  $\mathcal{C}$  on  $\Omega$  such that  $f$  is  $(\mathcal{C}, \mathcal{A}')$ -measurable, i. e.,

$$\mathcal{C} \text{ is a } \sigma\text{-algebra on } \Omega \text{ and } f \text{ is } (\mathcal{C}, \mathcal{A}')\text{-measurable} \Rightarrow f^{-1}(\mathcal{A}') \subset \mathcal{C}.$$

◁

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

**Definition 2.26 ( $\sigma$ -Algebra Generated by a Mapping)**

The set  $f^{-1}(\mathcal{A}')$  defined by Equation (2.13) is called the  $\sigma$ -algebra generated by  $f$  and  $\mathcal{A}'$ . If there is no ambiguity about  $\mathcal{A}'$ , then we also say that  $f^{-1}(\mathcal{A}')$  is generated by  $f$  and use the notation

$$\sigma(f) := f^{-1}(\mathcal{A}'). \quad (2.14)$$

**Remark 2.27 (Monotonicity)** Note that, for two set systems  $\mathcal{C}' \subset \mathcal{A}'$ ,

$$f^{-1}(\mathcal{C}') \subset f^{-1}(\mathcal{A}'), \quad (2.15)$$

because  $f^{-1}(\mathcal{C}') = \{f^{-1}(A') : A' \in \mathcal{C}'\} \subset \{f^{-1}(A') : A' \in \mathcal{A}'\} = f^{-1}(\mathcal{A}')$ . ◁

The following corollary immediately follows from Definition 2.26 and the definition of  $(\mathcal{A}, \mathcal{A}')$ -measurability (see Def. 2.5).

**Corollary 2.28 (A Condition Equivalent to Measurability)**

Let  $f: \Omega \rightarrow \Omega'$  be a mapping and let  $(\Omega', \mathcal{A}')$  be a measurable space. Then  $f$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable if and only if  $\sigma(f) \subset \mathcal{A}$ .

In the following lemma and the subsequent remark, we treat a  $\cap$ -stable generating system (see Def. 1.36). For a  $\sigma$ -algebra  $\mathcal{C}$  and a measurable function  $f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$  we use the notation  $\sigma(\mathcal{C}, f) := \sigma(\mathcal{C} \cup f^{-1}(\mathcal{A}'))$ .

**Lemma 2.29 ( $\cap$ -Stable Generating System)**

Let  $(\Omega, \mathcal{A})$  be a measurable space and let  $\mathcal{C} \subset \mathcal{A}$  be a  $\sigma$ -algebra. Furthermore, assume that  $\Omega'$  is finite or countable and let  $f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{P}(\Omega'))$  be a measurable mapping. Then the set

$$\mathcal{D} := \{C \cap f^{-1}(\{\omega'\}) : \omega' \in \Omega' \text{ and } C \in \mathcal{C}\}$$

is a  $\cap$ -stable generating system of  $\sigma(\mathcal{C}, f) := \sigma(\mathcal{C} \cup f^{-1}[\mathcal{P}(\Omega')])$ .

(Proof p. 71)

**Remark 2.30 (A Special Case)** Let us consider the special case in which  $\mathcal{C} = \{\Omega, \emptyset\}$ . In this case, Lemma 2.29 simplifies as follows: Let  $(\Omega, \mathcal{A})$  be a measurable space and let  $f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{P}(\Omega'))$  be a measurable mapping, where  $\Omega'$  is finite or countable. Then the set  $\{f^{-1}(\{\omega'\}) : \omega' \in \Omega'\} \cup \{\emptyset\}$  is a  $\cap$ -stable generating system of  $\sigma(f) := f^{-1}[\mathcal{P}(\Omega')]$ .  $\triangleleft$

**Example 2.31 ( $\sigma$ -Algebra Generated by an Indicator)** Let  $1_A: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  be the indicator of  $A \in \mathcal{A}$ . Then  $\sigma(1_A) = \{\Omega, \emptyset, A, A^c\}$ . The same  $\sigma$ -algebra is generated by  $1_A: (\Omega, \mathcal{A}) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  (see Remark 2.33 for the general proposition).  $\triangleleft$

**Example 2.32 (Flipping two Coins – continued)** In Example 2.2.2, we considered flipping two coins and the measurable mapping  $X = \text{number of flipping heads}$  with codomain  $\Omega' = \{0, 1, 2\}$ . In this example, all elements of  $X^{-1}[\mathcal{P}(\Omega')]$  have been listed in Table 2.1 (p. 45) as the inverse images  $X^{-1}(A')$  of the eight sets  $A' \in \mathcal{P}(\Omega')$ . Furthermore,  $X^{-1}[\mathcal{P}(\Omega')] = \mathcal{A}_2$ , where  $\mathcal{A}_2$  is the  $\sigma$ -algebra defined in Example 2.23.

Instead of choosing  $\Omega' = \{0, 1, 2\}$  as the codomain of  $X$ , we may also choose the set  $\mathbb{R}$  of real numbers, i. e.,  $X: \Omega \rightarrow \mathbb{R}$  is then considered to be a function into  $\mathbb{R}$ . In this case we use the *Borel  $\sigma$ -algebra*  $\mathcal{B}$  on  $\mathbb{R}$ . However, according to the following remark, the  $\sigma$ -algebra  $X^{-1}(\mathcal{B})$  generated by  $X$  and  $\mathcal{B}$  is the same as the  $\sigma$ -algebra  $X^{-1}[\mathcal{P}(\Omega')]$  generated by  $X$  and the power set of  $\Omega' = \{0, 1, 2\}$ .  $\triangleleft$

**Remark 2.33 ( $\sigma$ -Algebra Generated by a Function Into a Countable Set)** Let us consider a function  $f: \Omega \rightarrow \Omega' \subset \mathbb{R}$  and let  $\mathcal{B}$  denote the Borel  $\sigma$ -algebra on  $\mathbb{R}$ . If  $\Omega'$  is finite or countable, then  $f^{-1}[\mathcal{P}(\Omega')] = f^{-1}(\mathcal{B})$  (see Exercise 2-9).  $\triangleleft$

**Table 2.2.** Joe and Ann With Random Assignment and Measurable Mappings

Elements of $\Omega$				Measurable mappings		
Unit	Treatment	Success	$P(\{\omega\})$	Person variable $U$	Treatment variable $X$	Outcome variable $Y$
(Joe, no, -)			.09	Joe	0	0
(Joe, no, +)			.21	Joe	0	1
(Joe, yes, -)			.04	Joe	1	0
(Joe, yes, +)			.16	Joe	1	1
(Ann, no, -)			.24	Ann	0	0
(Ann, no, +)			.06	Ann	0	1
(Ann, yes, -)			.12	Ann	1	0
(Ann, yes, +)			.08	Ann	1	1

**Example 2.34 (Joe and Ann – continued)** Table 2.2 (p. 53) displays mappings on  $(\Omega, \mathcal{A})$ , all components of which have already been specified in Example 1.9. The *first* mapping displayed in Table 2.2 is the *person variable*  $U$  that assigns to each possible outcome  $\omega \in \Omega$  the value *Joe* if  $\omega \in \{\text{Joe}\} \times \Omega_X \times \Omega_Y$  and the value *Ann* if  $\omega \in \{\text{Ann}\} \times \Omega_X \times \Omega_Y$ . Hence,  $U: \Omega \rightarrow \Omega_U$  is a mapping with domain  $\Omega = \Omega_U \times \Omega_X \times \Omega_Y$  and codomain  $\Omega_U$ . It projects the first component  $u$  of  $\omega = (u, \omega_X, \omega_Y)$  onto the set  $\Omega_U$ . Therefore, it is also called the *first projection mapping*.

The *second* mapping in this table is the *treatment variable*  $X$ . It assigns to each possible outcome  $\omega \in \Omega$  the value 0 if  $\omega \in \Omega_U \times \{\text{no}\} \times \Omega_Y$  and the value 1 if  $\omega \in \Omega_U \times \{\text{yes}\} \times \Omega_Y$ . Hence,  $X: \Omega \rightarrow \Omega'$  is a function with domain  $\Omega$  and codomain  $\Omega' = \{0, 1\}$ .

The *third* mapping is the *outcome variable*  $Y$ . It assigns to each  $\omega \in \Omega$  the value 0 if  $\omega \in \Omega_U \times \Omega_X \times \{-\}$  and the value 1 if  $\omega \in \Omega_U \times \Omega_X \times \{+\}$ . Therefore,  $Y: \Omega \rightarrow \Omega'$  is a function with domain  $\Omega$  and codomain  $\Omega' = \{0, 1\}$ . Hence, all three mapping  $U$ ,  $X$ , and  $Y$  have the same domain  $\Omega$ .

Considering  $U: \Omega \rightarrow \Omega_U$  and the  $\sigma$ -algebra  $\mathcal{A}_U := \{\Omega_U, \emptyset, \{\text{Joe}\}, \{\text{Ann}\}\}$ , the  $\sigma$ -algebra  $U^{-1}(\mathcal{A}_U)$  consists of the following four inverse images: the event

$$U^{-1}(\{\text{Joe}\}) = \{(\text{Joe}, \text{no}, -), (\text{Joe}, \text{no}, +), (\text{Joe}, \text{yes}, -), (\text{Joe}, \text{yes}, +)\}$$

that *Joe is drawn*, the event

$$U^{-1}(\{\text{Ann}\}) = \{(\text{Ann}, \text{no}, -), (\text{Ann}, \text{no}, +), (\text{Ann}, \text{yes}, -), (\text{Ann}, \text{yes}, +)\}$$

that *Ann is drawn*, the sure event  $U^{-1}(\Omega_U) = \Omega$  that *Joe or Ann are drawn*, and the impossible event  $U^{-1}(\emptyset) = \emptyset$  that *neither Joe nor Ann are drawn*.  $\triangleleft$

### 2.3.3 Final $\sigma$ -Algebra

Consider the mapping  $f: \Omega \rightarrow \Omega'$ . As noted in Remark 2.25, for a  $\sigma$ -algebra  $\mathcal{A}'$  on  $\Omega'$ ,  $\sigma(f) = f^{-1}(\mathcal{A}')$  is the smallest  $\sigma$ -algebra on  $\Omega$  for which  $f$  is measurable. In contrast, now we consider a  $\sigma$ -algebra  $\mathcal{C}$  on  $\Omega$  and look for the largest  $\sigma$ -algebra  $\mathcal{C}'$  on  $\Omega'$  such that  $f$  is  $(\mathcal{C}, \mathcal{C}')$ -measurable. This  $\sigma$ -algebra is specified in the following lemma. It is called the *final  $\sigma$ -algebra*.

**Lemma 2.35 (Final  $\sigma$ -Algebra)**

Let  $f: \Omega \rightarrow \Omega'$  be a mapping and  $\mathcal{C}$  a  $\sigma$ -algebra on  $\Omega$ .

(i) Then

$$\mathcal{C}'_f := \{A' \subset \Omega' : f^{-1}(A') \in \mathcal{C}\} \quad (2.16)$$

is a  $\sigma$ -algebra on  $\Omega'$ .

(ii) Furthermore, if  $f: (\Omega, \mathcal{C}) \rightarrow (\Omega', \mathcal{A}')$  is a measurable mapping, then  $\mathcal{A}' \subset \mathcal{C}'_f$ .

(Proof p. 71)

Note that (ii) is a formal way of saying that  $\mathcal{C}'_f$  is the largest  $\sigma$ -algebra on  $\Omega'$  such that  $f$  is  $\mathcal{C}$ -measurable.

**Definition 2.36 (Final  $\sigma$ -Algebra)**

The  $\sigma$ -algebra  $\mathcal{C}'_f$  defined by Equation (2.16) is called the *final  $\sigma$ -algebra of  $\mathcal{C}$  under  $f$* .

### 2.3.4 Multivariate Mapping

Now consider the measurable space  $(\prod_{i=1}^n \Omega'_i, \otimes_{i=1}^n \mathcal{A}'_i)$  and note that the definitions of measurable mappings and of the  $\sigma$ -algebra generated by a mapping also apply to  $n$ -variate mappings  $f: \Omega \rightarrow \Omega'_1 \times \dots \times \Omega'_n$  and in particular to functions for which  $\Omega'_1 \times \dots \times \Omega'_n = \mathbb{R}^n$ .

**Lemma 2.37 ( $\sigma$ -Algebra Generated by a Multivariate Mapping)**

Let  $\Omega$  be a nonempty set,  $(\Omega'_i, \mathcal{A}'_i)$ ,  $i = 1, \dots, n$ ,  $n \in \mathbb{N}$ , be measurable spaces, and  $f = (f_1, \dots, f_n)$  be a multivariate mapping with  $f_i: \Omega \rightarrow \Omega'_i$ ,  $i = 1, \dots, n$ , i. e.,  $f: \Omega \rightarrow \prod_{i=1}^n \Omega'_i$ . Then

$$\sigma(f_1, \dots, f_n) := \sigma(f) = f^{-1}\left(\otimes_{i=1}^n \mathcal{A}'_i\right) = \sigma\left(\bigcup_{i=1}^n \sigma(f_i)\right). \quad (2.17)$$

(Proof p. 72)

According to the following theorem, a multivariate mapping is measurable if and only if all its components are measurable.

**Theorem 2.38 (Measurability of Multivariate Mappings)**

Under the assumptions of Lemma 2.37, the following two propositions are equivalent to each other:

- (a)  $f: (\Omega, \mathcal{A}) \rightarrow \left( \prod_{i=1}^n \Omega'_i, \otimes_{i=1}^n \mathcal{A}'_i \right)$  is a measurable mapping.  
 (b)  $\forall i = 1, \dots, n: f_i: (\Omega, \mathcal{A}) \rightarrow (\Omega'_i, \mathcal{A}'_i)$  is a measurable mapping.

(Proof p. 72)

**Remark 2.39 ( $\sigma$ -Algebra Generated by a Family of Mappings)** Let  $I$  be a (finite, countable, or uncountable) index set and let  $(f_i, i \in I)$  be a family of mappings  $f_i: (\Omega, \mathcal{A}) \rightarrow (\Omega'_i, \mathcal{A}'_i)$ . The  $\sigma$ -algebra generated by this family is defined as

$$\sigma(f_i, i \in I) := \sigma\left(\bigcup_{i \in I} \sigma(f_i)\right). \quad (2.18)$$

Then Equation (2.17) implies

$$\sigma(f) = \sigma(f_i, i \in I), \quad \text{where } I = \{i = 1, \dots, n\}. \quad (2.19)$$

◁

**Example 2.40 (Joe and Ann – continued)** In Example 2.34 we already considered the function  $X: \Omega \rightarrow \mathbb{R}$  indicating with its values 1 and 0 whether or not the drawn person is treated and the function  $Y: \Omega \rightarrow \mathbb{R}$  indicating with its values 1 and 0 whether or not the drawn person is successful. If we specify the  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$  such that  $X$  and  $Y$  are both  $(\mathcal{A}, \mathcal{B}_2)$ -measurable, then the bivariate function  $(X, Y): \Omega \rightarrow \mathbb{R}^2$  is  $(\mathcal{A}, \mathcal{B}_2)$ -measurable. And vice versa, if we specify the  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$  such that the bivariate function  $(X, Y): \Omega \rightarrow \mathbb{R}^2$  is  $(\mathcal{A}, \mathcal{B}_2)$ -measurable then  $X$  and  $Y$  are both  $(\mathcal{A}, \mathcal{B})$ -measurable. In this example  $X$ ,  $Y$ , and  $(X, Y)$  are measurable with respect to  $\mathcal{A}$  whenever the two inverse images  $X^{-1}(\{1\})$  and  $Y^{-1}(\{1\})$  are elements of  $\mathcal{A}$  (see Exercise 2-10). ◁

**Remark 2.41 (Lower Dimension Multivariate Mappings)** Lemma 2.37 and Remark 1.23 imply

$$\sigma(f_i, i \in J) \subset f^{-1}\left(\otimes_{i=1}^n \mathcal{A}'_i\right), \quad \forall J \subset \{1, \dots, n\}.$$

Furthermore, Theorem 2.38 implies: If

$$f = (f_1, \dots, f_n): (\Omega, \mathcal{A}) \rightarrow \left( \prod_{i=1}^n \Omega'_i, \otimes_{i=1}^n \mathcal{A}'_i \right)$$

is a measurable mapping and  $J = \{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ ,  $k \leq n$ , then

$$f_J := (f_{i_1}, \dots, f_{i_k}): (\Omega, \mathcal{A}) \rightarrow \left( \prod_{j=1}^k \Omega'_{i_j}, \otimes_{j=1}^k \mathcal{A}'_{i_j} \right)$$

is measurable as well. ◁

### 2.3.5 Projection Mapping

In Definition 1.31 we introduced the product  $\sigma$ -algebra  $\otimes_{i=1}^n \mathcal{A}_i$  for a finite number of measurable spaces  $(\Omega_i, \mathcal{A}_i)$ . Now we give an equivalent characterization. Let  $(\Omega_i, \mathcal{A}_i)$ ,  $i = 1, \dots, n$ , be measurable spaces. Then, for  $j = 1, \dots, n$ , the  $j$ th *projection mapping*  $\pi_j: \times_{i=1}^n \Omega_i \rightarrow \Omega_j$  is defined by

$$\pi_j(\omega_1, \dots, \omega_n) = \omega_j, \quad \forall (\omega_1, \dots, \omega_n) \in \times_{i=1}^n \Omega_i. \quad (2.20)$$

The inverse images are

$$\pi_j^{-1}(A_j) = \Omega_1 \times \dots \times \Omega_{j-1} \times A_j \times \Omega_{j+1} \times \dots \times \Omega_n, \quad \text{for } A_j \subset \Omega_j. \quad (2.21)$$

#### Lemma 2.42 (Product $\sigma$ -Algebra)

If  $(\Omega_i, \mathcal{A}_i)$ ,  $i = 1, \dots, n$ , are measurable spaces, then

$$\otimes_{i=1}^n \mathcal{A}_i = \sigma(\pi_1, \dots, \pi_n). \quad (2.22)$$

(Proof p. 73)

### 2.3.6 Measurability With Respect to a Mapping

In the next definition, we consider two mappings and the concept of a mapping being measurable with respect to another mapping.

#### Definition 2.43 (Measurability With Respect to a Mapping)

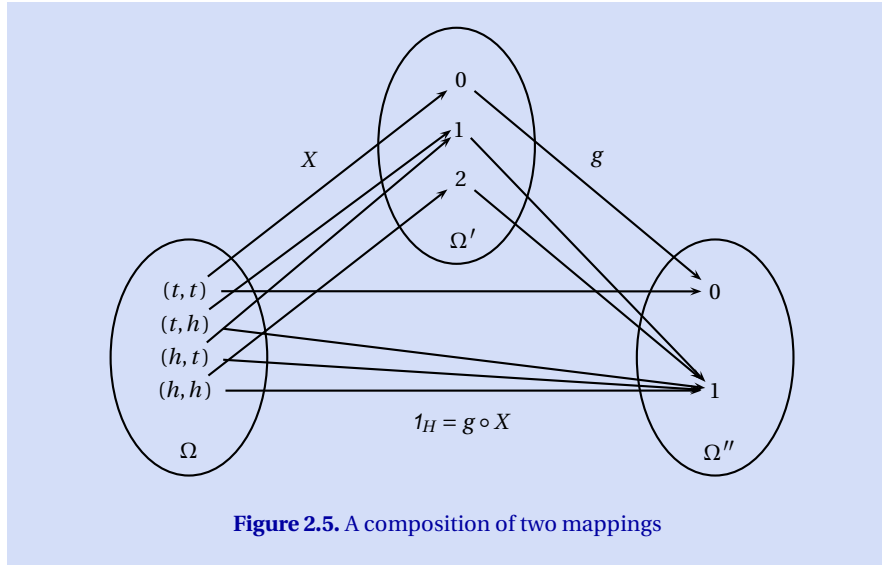
Let  $f: \Omega \rightarrow \Omega'$  and  $h: \Omega \rightarrow \Omega''$  be mappings, and let  $(\Omega', \mathcal{A}')$  and  $(\Omega'', \mathcal{A}'')$  be measurable spaces. Then  $h$  is called *measurable w.r.t.  $f$*  (or  *$f$ -measurable*) if

$$h^{-1}(\mathcal{A}'') \subset f^{-1}(\mathcal{A}'). \quad (2.23)$$

If  $\Omega'$  is finite or countable, then the following corollary provides a representation for all functions that are measurable with respect to  $f$ .

#### Corollary 2.44 (Measurability With Respect to a Discrete Function)

Let  $f: \Omega \rightarrow \Omega'$  be a mapping, let  $(\Omega', \mathcal{P}(\Omega'))$  be a measurable space, where  $\Omega'$  is finite or countable, and let  $h: \Omega \rightarrow \bar{\mathbb{R}}$  be a function. Then  $h$  is measurable w.r.t.  $f$  if and only if for all  $\omega' \in \Omega'$  there are  $\alpha_{\omega'} \in \bar{\mathbb{R}}$  such that



$$h = \sum_{\omega' \in \Omega'} \alpha_{\omega'} \cdot 1_{f^{-1}(\{\omega'\})}. \tag{2.24}$$

(Proof p. 73)

**Example 2.45 (Flipping Two Coins – continued)** Consider the mapping  $X =$  number of flipping heads with codomain  $\Omega' = \{0, 1, 2\}$ , let  $H := \{(h, t), (h, h), (t, h)\}$ , and let  $1_H: \Omega \rightarrow \Omega''$  denote the indicator of  $H$ , with  $\Omega'' = \{0, 1\}$ . Hence,  $1_H$  indicates with its values 1 and 0 whether or not at least one heads is flipped. If we consider the  $\sigma$ -algebra  $\mathcal{A}' = \mathcal{P}(\Omega')$  on  $\Omega'$  and the  $\sigma$ -algebra  $\mathcal{A}'' = \mathcal{P}(\Omega'')$  on  $\Omega''$ , then

$$X^{-1}(\mathcal{A}') = \{ \Omega, \emptyset, \{(h, h)\}, \{(h, t), (t, h)\}, \{(t, t)\}, \\ \{(h, h), (h, t), (t, h)\}, \{(h, h), (t, t)\}, \{(h, t), (t, h), (t, t)\} \}$$

and

$$1_H^{-1}(\mathcal{A}'') = \{ \Omega, \emptyset, \{(h, h), (h, t), (t, h)\}, \{(t, t)\} \}.$$

Obviously,  $1_H^{-1}(\mathcal{A}'') \subset X^{-1}(\mathcal{A}')$ . Therefore,  $1_H$  is measurable with respect to  $X$ , but not vice versa. That is,  $X$  represents a more detailed information about the outcome of the random experiment than  $1_H$ . Hence, if the value of  $X$  is known, then we can compute the value of  $1_H$ , but not vice versa. In our example, Figure 2.5 shows: if  $X(\omega) = 1$ , then  $1_H(\omega) = 1$ . However, if  $1_H(\omega) = 1$ , then  $X(\omega) = 1$  or  $X(\omega) = 2$ . (For a more general presentation of this property see Lemma 2.52.)  $\triangleleft$

## 2.4 Theorems on Measurable Mappings

In this section we consider compositions of mappings, which are defined as follows: Let  $\Omega$ ,  $\Omega'$ , and  $\Omega''$  be nonempty sets and let  $f: \Omega \rightarrow \Omega'$  and  $g: \Omega' \rightarrow \Omega''$  be mappings. Then the *composition of  $f$  and  $g$*  is the mapping  $g \circ f: \Omega \rightarrow \Omega''$  defined by:

$$g \circ f(\omega) := g[f(\omega)], \quad \forall \omega \in \Omega, \quad (2.25)$$

(see Fig. 2.5), where  $g \circ f(\omega)$  denotes the value of the mapping  $g \circ f$  for the argument  $\omega$ . Instead of  $g \circ f$ , we often use the notation  $g(f)$  and say that  $g(f)$  is a *function of  $f$* . Using this notation, Equation (2.25) can be written

$$g(f)(\omega) = g[f(\omega)], \quad \forall \omega \in \Omega. \quad (2.26)$$

**Lemma 2.46 (Compositions With a Finite or Countable Number of Values)**

Let  $f: \Omega \rightarrow \Omega'$  be a mapping, where  $\Omega'$  is finite or countable, and let  $g: \Omega' \rightarrow \mathbb{R}$  be a function. Furthermore, for  $\omega' \in \Omega'$ , define  $1_{f=\omega'} := 1_{f^{-1}(\{\omega'\})}$ . Then

$$g \circ f = g(f) = \sum_{\omega' \in \Omega'} g(\omega') \cdot 1_{f^{-1}(\{\omega'\})} = \sum_{\omega' \in \Omega'} g(\omega') \cdot 1_{f=\omega'}. \quad (2.27)$$

(Proof p. 73)

Hence, under the assumptions of Lemma 2.46, for all  $\omega \in \Omega$ ,

$$g \circ f(\omega) = g[f(\omega)] = \sum_{\omega' \in \Omega'} g(\omega') \cdot 1_{f^{-1}(\{\omega'\})}(\omega) = \sum_{\omega' \in \Omega'} g(\omega') \cdot 1_{f=\omega'}(\omega). \quad (2.28)$$

**Example 2.47 (Flipping two Coins – continued)** Let us consider  $X = \text{number of flipping heads}$  and the mapping  $g: \Omega' \rightarrow \Omega''$  defined by  $g(x) := 1_{\{1,2\}}(x)$ , for all  $x \in \Omega'$  (see Fig. 2.5). Then the composition  $g \circ X$  defines a new mapping  $g \circ X: \Omega \rightarrow \Omega''$ , where  $\Omega'' = \{0, 1\}$ . In this example, the composition  $g \circ X$  is identical to the indicator  $1_H$  of the event  $H = \{(h, h), (t, h), (h, t)\}$  that heads are flipped at least once.  $\triangleleft$

**Example 2.48 (Joe and Ann – continued)** In Example 2.34 we already considered the mapping  $U: \Omega \rightarrow \Omega_U = \{\text{Joe}, \text{Ann}\}$  showing which person is drawn and the mapping  $X: \Omega \rightarrow \Omega' = \{0, 1\}$  indicating whether or not the drawn person is treated. Now we can consider the bivariate random variable  $(U, X): \Omega \rightarrow \Omega_U \times \Omega'$  and we can write

$$X = g \circ (U, X) = g(U, X)$$

as the composition of  $(U, X)$  and a (projection) mapping  $g$ ,

$$g[(u, x)] = x, \quad \forall (u, x) \in \Omega_U \times \Omega'. \quad \triangleleft$$

### 2.4.1 Measurability of a Composition

The following theorem shows that measurability is preserved by the composition of mappings.

**Theorem 2.49 (Measurability of a Composition)**

If  $f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$  and  $g: (\Omega', \mathcal{A}') \rightarrow (\Omega'', \mathcal{A}'')$  are measurable mappings, then the composition  $g \circ f$  is  $(\mathcal{A}, \mathcal{A}'')$ -measurable.

(Proof p. 74)

**Remark 2.50 ( $\sigma$ -Algebra Generated by a Composition)** Note that

$$(g \circ f)^{-1}(\mathcal{A}'') = f^{-1}[g^{-1}(\mathcal{A}'')] \quad (2.29)$$

(see the proof of Theorem 2.49).  $\triangleleft$

**Example 2.51 (Flipping two Coins – continued)** Figure 2.5 illustrates Theorem 2.49. If (a)  $X$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable and (b)  $g$  is  $(\mathcal{A}', \mathcal{A}'')$ -measurable, then  $1_H = g \circ X$  is  $(\mathcal{A}, \mathcal{A}'')$ -measurable. Suppose  $\mathcal{A}' = \mathcal{P}(\Omega')$  and  $\mathcal{A}'' = \mathcal{P}(\Omega'')$ , where  $\Omega' = \{0, 1, 2\}$  and  $\Omega'' = \{0, 1\}$ . Then the premise '(a) and (b)' is satisfied if  $\mathcal{A}$  is such that  $X^{-1}(\mathcal{A}') \subset \mathcal{A}$ . If the premise '(a) and (b)' is *not* satisfied, then we cannot conclude that  $1_H$  is  $(\mathcal{A}, \mathcal{A}'')$ -measurable. Note that in this example  $1_H$  can be  $(\mathcal{A}, \mathcal{A}'')$ -measurable even if (a) and (b) do not hold. A sufficient requirement is that  $\{(t, t)\}$  and  $\{(t, h), (h, t), (h, h)\}$ , the inverse images of  $\{0\}$  and  $\{1\}$  under  $1_H$ , respectively, are elements of  $\mathcal{A}$  (see Cor. 2.21).  $\triangleleft$

If a mapping  $h$  is measurable with respect to a mapping  $f$ , then each element in the  $\sigma$ -algebra generated by  $h$  is an element in the  $\sigma$ -algebra generated by  $f$ . If  $h$  is measurable with respect to  $f$ , then in a sense, the information represented by  $h$  is already contained in  $f$  (cf. section 2.3.2). This is expressed in more formal terms in the following lemma, which is crucial, e. g., in the general definition of conditional expectations  $E(Y|X=x)$  (see ch. 10).

**Lemma 2.52 (Factorization Lemma of Measurable Functions)**

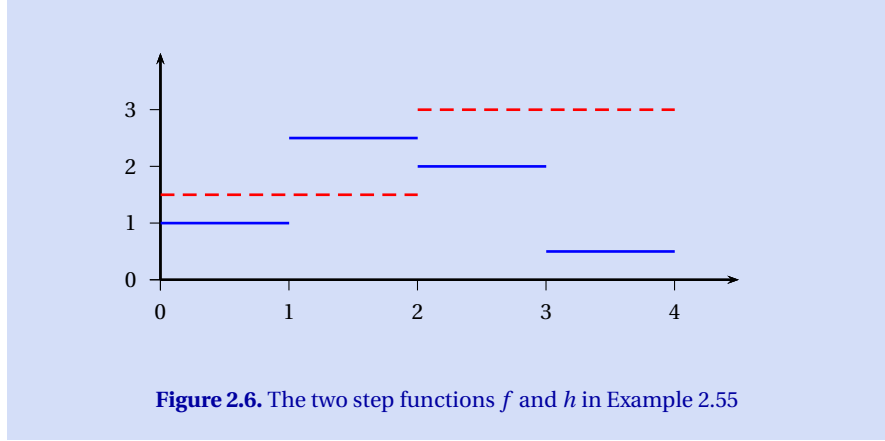
Let  $f: \Omega \rightarrow \Omega'$  be a mapping, let  $(\Omega', \mathcal{A}')$  be a measurable space, and let  $h: \Omega \rightarrow \overline{\mathbb{R}}$  be a function. Then  $h$  is measurable w.r.t.  $f$ , i. e.,  $h^{-1}(\overline{\mathcal{B}}) \subset f^{-1}(\mathcal{A}')$ , if and only if there is a measurable function  $g: (\Omega', \mathcal{A}') \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  such that

$$h = g \circ f \quad (2.30)$$

is the composition of  $f$  and  $g$ . We call  $g$  a factorization of  $h$  w.r.t.  $f$ .

For a proof see Klenke (2013, Corollary 1.97, pp. 38).

If, instead of  $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$  we consider a measurable space  $(\Omega'', \mathcal{P}(\Omega''))$ , where  $\Omega''$  is finite or countable, 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



**Figure 2.6.** The two step functions  $f$  and  $h$  in Example 2.55

the  $\sigma$ -algebra on  $\Omega''$  is the power set of  $\Omega''$  (see Example 2.9). Hence, Lemma 2.52 implies the following corollary:

**Corollary 2.53 (Factorization of a Mapping Into a Finite or Countable Set)**

Let  $f: \Omega \rightarrow \Omega'$  be a mapping,  $(\Omega', \mathcal{A}')$  a measurable space, and  $h: \Omega \rightarrow \Omega''$  a mapping, where  $\Omega''$  is finite or countable. Then  $h$  is measurable w.r.t.  $f$ , i. e.,  $h^{-1}[\mathcal{P}(\Omega'')] \subset f^{-1}(\mathcal{A}')$ , if and only if there is a measurable mapping  $g: (\Omega', \mathcal{A}') \rightarrow (\Omega'', \mathcal{P}(\Omega''))$  such that  $h = g \circ f$ .

**Example 2.54 (Flipping two Coins – continued)** If we specify  $\Omega' = \{0, 1, 2\}$ , the  $\sigma$ -algebra  $\mathcal{A}' = \mathcal{P}(\Omega')$ , the set  $\Omega'' = \{0, 1\}$ , the  $\sigma$ -algebra  $\mathcal{A}'' = \mathcal{P}(\Omega'')$ , and the function  $h = 1_H$ , then the example depicted in Figure 2.5 can be used to illustrate this corollary. The mapping  $g$  in this figure is such that  $1_H = g \circ X$ .  $\triangleleft$

**Example 2.55 (Two Step Functions)** Figure 2.6 presents an example in which  $\Omega = [0, 4]$ ,  $A_1 = [0, 1]$ ,  $A_2 = ]1, 2]$ ,  $A_3 = ]2, 3]$ , and  $A_4 = ]3, 4]$ . Note that the sets  $A_1, \dots, A_4$  are pairwise disjoint. The measurable function  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  is defined by

$$f = \sum_{i=1}^4 \alpha_i 1_{A_i},$$

where  $\alpha_1 = 1$ ,  $\alpha_2 = 2.5$ ,  $\alpha_3 = 2$ , and  $\alpha_4 = 0.5$ . Furthermore, the function  $h: \Omega \rightarrow \mathbb{R}$  is defined by

$$h = \sum_{j \in \{1,3\}} \beta_j 1_{A_j \cup A_{j+1}}$$

with  $\beta_1 = 1.5$  and  $\beta_3 = 3$ . Note that  $\sigma(h) = \sigma(\{A_j \cup A_{j+1}: j \in \{1,3\}\})$ , whereas  $\sigma(f) = \sigma(\{A_i: i = 1, \dots, 4\})$  (see Exercise 2-11). Therefore,  $h$  is measurable with respect to  $f$ , i. e.,  $\sigma(h) \subset \sigma(f)$ .

According to Lemma 2.52, there is a function  $g: \mathbb{R} \rightarrow \mathbb{R}$  such that  $h = g \circ f$ . In fact, if we define  $g$  by

$$g(x) = \sum_{j \in \{1,3\}} \beta_j 1_{\{\alpha_j, \alpha_{j+1}\}}(x), \quad \forall x \in \mathbb{R},$$

then  $h = g \circ f$ . The function  $g$  takes on the value 1.5 if  $x = \alpha_1 = 1$  or  $x = \alpha_2 = 2.5$  and the value 3 if  $x = \alpha_3 = 2$  or  $x = \alpha_4 = 0.5$ . For all other  $x \in \mathbb{R}$  the value of  $g$  is 0.  $\triangleleft$

**Example 2.56 (Square of a Real-Valued Function)** Suppose  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  is a real-valued measurable function and  $f^2(\omega) := f(\omega)^2$ , for all  $\omega \in \Omega$ .

- (i) If  $f$  is *nonnegative*, i. e., if  $f(\omega) \geq 0$ , for all  $\omega \in \Omega$ , then  $f$  and  $f^2$  are measurable with respect to each other, i. e.,  $\sigma(f) = \sigma(f^2)$ .
- (ii) If there are  $\omega_1, \omega_2 \in \Omega$  with  $f(\omega_1) < 0 < f(\omega_2)$  and  $f^2(\omega_1) = f^2(\omega_2)$ , then  $\sigma(f^2) \subset \sigma(f)$ , but  $\sigma(f) \neq \sigma(f^2)$

(see Exercise 2-13.) In a sense,  $\sigma(f) = \sigma(f^2)$  means that  $f$  and  $f^2$  contain the same information, whereas  $\sigma(f^2) \subset \sigma(f)$ ,  $\sigma(f) \neq \sigma(f^2)$  means that  $f^2$  contains less information than  $f$ . If, e. g.,  $f^2(\omega) = 4$ , then  $f(\omega) = 2$  or  $f(\omega) = -2$ .  $\triangleleft$

### 2.4.2 Theorems on Measurable Functions

In the first theorem we consider sums and differences as well as products and ratios of measurable functions. The *sum* of two functions  $f, h: \Omega \rightarrow \mathbb{R}^n$  is again a function  $(f + h): \Omega \rightarrow \mathbb{R}^n$  defined by

$$(f + h)(\omega) := \begin{pmatrix} f_1 + h_1 \\ \vdots \\ f_n + h_n \end{pmatrix}(\omega) := \begin{pmatrix} f_1(\omega) + h_1(\omega) \\ \vdots \\ f_n(\omega) + h_n(\omega) \end{pmatrix}, \quad \forall \omega \in \Omega.$$

The first parentheses in the term  $(f + h)(\omega)$  are used to make clear that  $f + h$  is a symbol of a new function on  $\Omega$ . Of course, the difference  $f - h$  is defined in the same way as  $f + h$  replacing  $+$  by  $-$ .

Similarly, the *product*  $f \cdot h$  of two functions  $f, h: \Omega \rightarrow \mathbb{R}$  is again a function  $(f \cdot h): \Omega \rightarrow \mathbb{R}$  defined by

$$(f \cdot h)(\omega) := f(\omega) \cdot h(\omega), \quad \forall \omega \in \Omega.$$

Correspondingly,  $f/h: \Omega \rightarrow \mathbb{R}$  is defined by

$$(f/h)(\omega) := f(\omega)/h(\omega), \quad \forall \omega \in \Omega,$$

provided that  $h(\omega) \neq 0$  for all  $\omega \in \Omega$ .

**Theorem 2.57 (Sums and Products of Measurable Functions)**

If  $f, h: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  are measurable functions, then  $f + h$  and  $f - h$  are  $(\mathcal{A}, \mathcal{B}_n)$ -measurable as well. Furthermore, if  $f, h: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  are measurable functions, then  $f \cdot h$  and  $f/h$  (with  $h(\omega) \neq 0$ , for all  $\omega \in \Omega$ ) are also  $(\mathcal{A}, \mathcal{B})$ -measurable.

For a proof see Klenke (2013, Theorem 1.91, p. 37).

**Remark 2.58 (Squared Function)** If  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  is a measurable function, then  $f^2 := f \cdot f$  is also  $(\mathcal{A}, \mathcal{B})$ -measurable. Obviously, this also applies to  $f^n$ ,  $n \in \mathbb{N}$ . Hence, if  $f$  is  $(\mathcal{A}, \mathcal{B})$ -measurable, then  $f^n$  is also  $(\mathcal{A}, \mathcal{B})$ -measurable.  $\triangleleft$

**Example 2.59 (Scale Transformations and Translations)** Remember that a constant real number can always be interpreted as a measurable function (see Example 2.10). Therefore, Theorem 2.57 implies that, for all  $\alpha \in \mathbb{R}$ , the functions  $f + \alpha$ ,  $f - \alpha$ , and  $\alpha \cdot f$  are  $(\mathbb{R}, \mathcal{B})$ -measurable if  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  is a measurable function.  $\triangleleft$

**Example 2.60 (Number of Flipping Heads)** Consider flipping a coin  $n$  times, let  $\Omega = \{h, t\}^n$ , and let  $1_{A_i}: \Omega \rightarrow \mathbb{R}$  denote the indicators of flipping heads at the  $i$ th flip of the coin. Then

$$X = \sum_{i=1}^n 1_{A_i}$$

is the number of flipping heads. If  $A_i \in \mathcal{A} = \mathcal{P}(\Omega)$ ,  $i = 1, \dots, n$ , then  $(\Omega, \mathcal{A})$  is a measurable space and  $X$  is measurable for any  $\sigma$ -algebra on  $\mathbb{R}$  (see Example 2.9). In the case  $\mathcal{A} = \mathcal{P}(\Omega)$  it is not necessary to apply Theorem 2.57.  $\triangleleft$

**Example 2.61 (Linear Combination of two Functions)** Let  $f, h: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  be measurable functions and  $\alpha, \beta \in \mathbb{R}$ . Then, according to Theorem 2.57, the function  $(\alpha \cdot f + \beta \cdot h): (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  defined by

$$(\alpha \cdot f + \beta \cdot h)(\omega) = \alpha \cdot f(\omega) + \beta \cdot h(\omega), \quad \forall \omega \in \Omega. \quad (2.31)$$

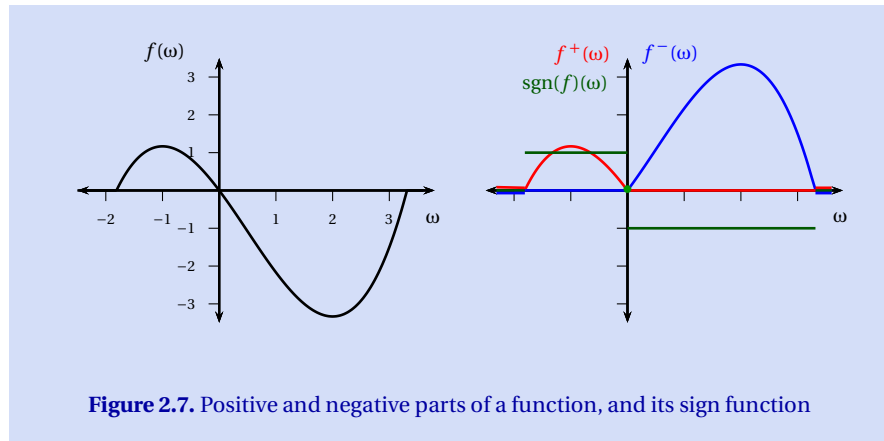
is  $(\mathcal{A}, \mathcal{B})$ -measurable.  $\triangleleft$

**Remark 2.62 (Positive and Negative Parts of a Function)** In Theorem 2.66 we consider the positive and the negative parts of a function  $f: \Omega \rightarrow \overline{\mathbb{R}}$ . The positive part  $f^+: \Omega \rightarrow \overline{\mathbb{R}}$  is defined by

$$f^+(\omega) := \max(f(\omega), 0), \quad \forall \omega \in \Omega,$$

and the negative part  $f^-: \Omega \rightarrow \overline{\mathbb{R}}$  by

$$f^-(\omega) := -\min(f(\omega), 0), \quad \forall \omega \in \Omega.$$



**Figure 2.7.** Positive and negative parts of a function, and its sign function

Hence, the value  $f^+(\omega)$  of the positive part of  $f$  is defined to be the *greater* one of the two numbers  $f(\omega)$  and 0 if they differ and  $f^+(\omega) = 0$  if  $f(\omega) = 0$ . In contrast, the value  $f^-(\omega)$  of the negative part of  $f$  is defined to be the *smaller* one of the two numbers  $f(\omega)$  and 0 *multiplied by*  $-1$  if they differ and  $f^-(\omega) = 0$  if  $f(\omega) = 0$ . Note that  $f^+$  and  $f^-$  are both nonnegative functions and that

$$f = f^+ - f^-.$$

◁

**Example 2.63 (Positive and Negative Parts of a Function)** The positive and negative parts of a function are illustrated by Figure 2.7 showing the graph of the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$f(x) = \begin{cases} \frac{x^3}{3} - \frac{x^2}{2} - 2x, & \text{if } -1.81 < x < 3.315, \\ 0, & \text{otherwise.} \end{cases}$$

The positive part  $f^+$  takes on the value 0 if  $x \leq 0$  (see the red line on the horizontal axis), whereas negative part  $f^-$  takes on the value 0 if  $x \geq 0$  (see the blue line on the horizontal axis).

◁

**Remark 2.64 (Absolute Value Function)** Furthermore, we consider the *absolute value function*  $|f|: \Omega \rightarrow \overline{\mathbb{R}}$  defined by

$$|f|(\omega) := |f(\omega)| := \begin{cases} f(\omega), & \text{if } f(\omega) \geq 0 \\ -f(\omega), & \text{if } f(\omega) < 0. \end{cases}$$

Note that  $|f| = f^+ + f^- = \max(f^+, f^-)$ . Hence, in Figure 2.7, the absolute value function is represented by the red and blue lines *above* (if  $-1.81 < x < 3.315$ ) or *on* (if  $x < -1.81$  or  $x > 3.315$ ) the horizontal axis.

◁

**Remark 2.65 (Sign Function)** In Theorem 2.66, we also refer to  $\text{sgn}(f): \Omega \rightarrow \mathbb{R}$ , called the *sign function*, which is defined by

$$\text{sgn}(f)(\omega) = \begin{cases} 1, & \text{if } f(\omega) > 0 \\ 0, & \text{if } f(\omega) = 0 \\ -1, & \text{if } f(\omega) < 0. \end{cases}$$

In Figure 2.7, the graph of this function is represented by the green lines above and below the horizontal axis, and by the green point.  $\triangleleft$

**Theorem 2.66 (Positive and Negative Parts of a Function)**

Let  $(\Omega, \mathcal{A})$  be a measurable space. If  $f: \Omega \rightarrow \overline{\mathbb{R}}$  is  $(\mathcal{A}, \overline{\mathcal{B}})$ -measurable, then the functions  $f^+$ ,  $f^-$ ,  $|f|$ , and  $\text{sgn}(f)$  are  $(\mathcal{A}, \overline{\mathcal{B}})$ -measurable as well.

For a proof see Klenke (2013, Corollary 1.89, p. 36). The positive part  $f^+$ , the negative part  $f^-$ , and the absolute value function  $|f|$  of a function  $f$  plays an important role in integration theory (see ch. 3).

Another implication of Theorem 2.57 on the measurability of some sets that are often used is formulated in the following remark.

**Remark 2.67 (Some Important Measurable Sets)** Let  $(\Omega, \mathcal{A})$  be a measurable space and let  $f, g: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be measurable functions. Then

- (a)  $\{\omega \in \Omega: f(\omega) \geq g(\omega)\} \in \mathcal{A}$
- (b)  $\{\omega \in \Omega: f(\omega) > g(\omega)\} \in \mathcal{A}$
- (c)  $\{\omega \in \Omega: f(\omega) = g(\omega)\} \in \mathcal{A}$

(see Exercise 2-12).  $\triangleleft$

## 2.5 Equivalence of Two Mappings With Respect to a Measure

Now we study some properties of mappings  $f: \Omega \rightarrow \Omega'$  involving a measure space  $(\Omega, \mathcal{A}, \mu)$ . In this case, we use the notation

$$f: (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$$

to express that  $f: \Omega \rightarrow \Omega'$  is a mapping and that  $\mu$  is a measure on the measurable space  $(\Omega, \mathcal{A})$ . If there is also a  $\sigma$ -algebra  $\mathcal{A}'$  on  $\Omega'$ , then we use the notation

$$f: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$$

to express that the mapping  $f: \Omega \rightarrow \Omega'$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable and that  $\mu$  is a measure on the measurable space  $(\Omega, \mathcal{A})$ .

Remember, two mappings  $f$  and  $g$  are *identical*, i. e.,  $f = g$ , if and only if

$$\{\omega \in \Omega: f(\omega) \neq g(\omega)\} = \emptyset.$$

A less restrictive concept is their *equivalence with respect to the measure  $\mu$* .

**Definition 2.68 (Equivalence of Two Mappings With Respect to a Measure)**

Let  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$  be mappings. Then  $f$  and  $g$  are called  $\mu$ -equivalent, denoted by

$$f \stackrel{\mu}{=} g,$$

if there is an  $A \in \mathcal{A}$  with  $\mu(A) = 0$  and, for all  $\omega \in \Omega \setminus A$ ,  $f(\omega) = g(\omega)$ .

Because  $\mu(\emptyset) = 0$ ,  $f = g$  implies  $f \stackrel{\mu}{=} g$ .

**Remark 2.69 (A Note on Notation)** If  $f, g$  are denoted by  $f: (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'_f$  and  $g: (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'_g$ , then we can choose  $\Omega' = \Omega'_f \cup \Omega'_g$  and denote  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$  (see Rem. 2.4).  $\triangleleft$

**Remark 2.70 (An Alternative Notation)** If  $f \stackrel{\mu}{=} g$  we also say that  $f = g$ ,  $\mu$ -almost everywhere ( $\mu$ -a.e.). Furthermore, we also write

$$f(\omega) = g(\omega), \quad \text{for } \mu\text{-almost all } \omega \in \Omega, \quad (2.32)$$

and use  $f(\omega) \stackrel{\mu\text{-a.e.}}{=} g(\omega)$  as a shortcut.  $\triangleleft$

**Remark 2.71 (Singleton With a Positive Value of a Measure)** If  $f \stackrel{\mu}{=} g$  or, equivalently,  $f(\omega) \stackrel{\mu\text{-a.e.}}{=} g(\omega)$ , and  $\{\omega^*\} \in \mathcal{A}$  with  $\mu(\{\omega^*\}) > 0$ , then

$$f(\omega^*) = g(\omega^*).$$

$\triangleleft$

**Remark 2.72 ( $\mu$ -Equivalence, Restricted Functions, and Compositions)** Let  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$  be mappings.

(i) If  $\Omega' = \mathbb{R}$ , then

$$f \stackrel{\mu}{=} g \Rightarrow 1_A \cdot f \stackrel{\mu}{=} 1_A \cdot g, \quad \forall A \in \mathcal{A}. \quad (2.33)$$

(ii) If  $h: \Omega' \rightarrow \Omega''$  is also a mapping, then

$$f \stackrel{\mu}{=} g \Rightarrow h \circ f \stackrel{\mu}{=} h \circ g \quad (2.34)$$

(see Exercise 2-14).  $\triangleleft$

**Remark 2.73 (Equivalence Relation)** If  $\mathcal{M}$  is a set of mappings  $(\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$ , then  $\stackrel{\mu}{=}$  is an *equivalence relation* on  $\mathcal{M}$  (see Exercise 2-15). In other words, if  $f, g, h \in \mathcal{M}$ , then the following propositions hold for  $\stackrel{\mu}{=}$ :

- (i)  $f \stackrel{\mu}{=} f$  (reflexivity).
- (ii)  $g \stackrel{\mu}{=} f$  if and only if  $f \stackrel{\mu}{=} g$  (symmetry).
- (iii) If  $f \stackrel{\mu}{=} g$  and  $g \stackrel{\mu}{=} h$ , then  $f \stackrel{\mu}{=} h$  (transitivity).

&lt;

**Definition 2.74 (Equivalence Class With Respect to a Measure)**

Let  $\mathcal{M}$  be a set of mappings  $(\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$  and let  $f \in \mathcal{M}$ . Then

$$C(f) := \{g \in \mathcal{M} : g \stackrel{\mu}{=} f\}$$

is called the  $\mu$ -equivalence class of  $f$  in  $\mathcal{M}$  and  $f$  a representative of the class  $C(f)$ .

**Remark 2.75 (A Partition of the Set  $\mathcal{M}$ )** If  $\mathcal{M}$  is a set of mappings  $(\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$ , then the set  $\{C(f) : f \in \mathcal{M}\}$  is a partition of  $\mathcal{M}$ , i. e.,

- (a)  $\forall f \in \mathcal{M} : C(f) \neq \emptyset$ .
- (b)  $\forall f, g \in \mathcal{M} : C(f) = C(g)$  or  $C(f) \cap C(g) = \emptyset$ .
- (c)  $\bigcup_{f \in \mathcal{M}} C(f) = \mathcal{M}$ .

(see Exercise 2-16).

&lt;

**Remark 2.76 (Other Properties of  $\mu$ -Equivalence)**

- (i) Let  $f, g : (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$ . If  $\mu(\Omega) > 0$ , then

$$\forall \alpha \in \Omega', \forall \beta \in \Omega' : f \stackrel{\mu}{=} \alpha \wedge g \stackrel{\mu}{=} \beta \wedge f \stackrel{\mu}{=} g \Rightarrow \alpha = \beta. \quad (2.35)$$

- (ii) If  $f, g, f^*, g^* : (\Omega, \mathcal{A}, \mu) \rightarrow \mathbb{R}$ , then

$$\begin{aligned} f \stackrel{\mu}{=} f^* \wedge g \stackrel{\mu}{=} g^* &\Rightarrow f + g \stackrel{\mu}{=} f^* + g^*, \\ f - g \stackrel{\mu}{=} f^* - g^*, & \\ f \cdot g \stackrel{\mu}{=} f^* \cdot g^*. & \end{aligned} \quad (2.36)$$

Furthermore, suppose  $\mu(\{\omega \in \Omega : g(\omega) = 0\}) = 0$ , and define  $\frac{f}{g} : \Omega \rightarrow \mathbb{R}$  by

$$\frac{f}{g}(\omega) := \begin{cases} \frac{f(\omega)}{g(\omega)}, & \text{if } g(\omega) \neq 0, \\ 0, & \text{otherwise,} \end{cases} \quad \forall \omega \in \Omega,$$

and let  $\frac{f^*}{g^*}$  be defined analogously. Then

$$f \stackrel{\mu}{=} f^* \wedge g \stackrel{\mu}{=} g^* \Rightarrow \frac{f}{g} \stackrel{\mu}{=} \frac{f^*}{g^*}. \quad (2.37)$$

(iii) If  $f_i, f_i^* : (\Omega, \mathcal{A}, \mu) \rightarrow \mathbb{R}$  and  $\alpha_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ , then

$$(\forall i = 1, \dots, n: f_i \stackrel{\mu}{=} f_i^*) \Rightarrow \sum_{i=1}^n \alpha_i f_i \stackrel{\mu}{=} \sum_{i=1}^n \alpha_i f_i^*. \quad (2.38)$$

(iv) If  $f_1, f_2, \dots, f_1^*, f_2^*, \dots : (\Omega, \mathcal{A}, \mu) \rightarrow \mathbb{R}$  and  $\alpha_1, \alpha_2, \dots \in \mathbb{R}$ , then

$$(\forall i = 1, 2, \dots: f_i \stackrel{\mu}{=} f_i^*) \Rightarrow \sum_{i=1}^{\infty} \alpha_i f_i \stackrel{\mu}{=} \sum_{i=1}^{\infty} \alpha_i f_i^*, \quad (2.39)$$

provided that the limits denoted by the infinite sums (see Box 0.1) exist.

For proofs see Exercise 2-17.  $\triangleleft$

**Remark 2.77 (Order Relations for Real-Valued Functions)** For two mappings  $f, g : \Omega \rightarrow \mathbb{R}$  we write  $f < g$ , if and only if

$$\{\omega \in \Omega: f(\omega) \geq g(\omega)\} = \emptyset.$$

The notation  $f > g$ ,  $f \leq g$ , and  $f \geq g$  is used correspondingly.  $\triangleleft$

**Remark 2.78 (Order Relations With Respect to a Measure  $\mu$ )** For functions  $f, g, h : (\Omega, \mathcal{A}, \mu) \rightarrow \mathbb{R}$ , we also use the notation

$$f \stackrel{\mu}{<} g,$$

if there is an  $A \in \mathcal{A}$  with  $f(\omega) < g(\omega)$  for all  $\omega \in \Omega \setminus A$  and  $\mu(A) = 0$ . The notation  $f \stackrel{\mu}{>} g$ ,  $f \stackrel{\mu}{\leq} g$ , and  $f \stackrel{\mu}{\geq} g$  is used correspondingly. Furthermore,

$$f \stackrel{\mu}{<} g \quad \text{and} \quad g \stackrel{\mu}{=} h \quad \Rightarrow \quad f \stackrel{\mu}{<} h. \quad (2.40)$$

The analog propositions hold for  $\stackrel{\mu}{>}$ ,  $\stackrel{\mu}{\leq}$ , and  $\stackrel{\mu}{\geq}$  (see Exercise 2-18).  $\triangleleft$

## 2.6 Image Measure

In the definition of a measurable mapping  $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$  we required: for all  $A' \in \mathcal{A}'$ :  $f^{-1}(A') \in \mathcal{A}$ . Because a measure  $\mu$  assigns a value to *all* elements  $A \in \mathcal{A}$ , the measure  $\mu$  also assigns a value to each  $f^{-1}(A') := \{\omega \in \Omega: f(\omega) \in A'\}$ . This is the reason for choosing the term *measurable mapping*: If  $\mu$  is a measure on  $\mathcal{A}$  and  $f$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable, then there is a value  $\mu[f^{-1}(A')]$  for all inverse images  $f^{-1}(A')$ ,  $A' \in \mathcal{A}'$ .

According to the following theorem, a measurable mapping  $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$  induces a measure on the codomain space  $(\Omega', \mathcal{A}')$ .

**Theorem 2.79 (Image Measure)**

Let  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$  be a measurable mapping. Then the function  $\mu_f: \mathcal{A}' \rightarrow \overline{\mathbb{R}}$  defined by

$$\mu_f(A') := \mu[f^{-1}(A')], \quad \forall A' \in \mathcal{A}', \quad (2.41)$$

is a measure on the measurable space  $(\Omega', \mathcal{A}')$ .

(Proof p. 74)

**Definition 2.80 (Image Measure)**

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$  is a measurable mapping, then  $\mu_f: \mathcal{A}' \rightarrow \mathbb{R}$  defined by Equation (2.41) is called the image measure of  $\mu$  under  $f$ .

**Example 2.81 (Rectangles – continued)** Now we consider a measure  $\mu: \mathcal{A} \rightarrow \mathbb{R}$  restricted to  $\mathcal{A}$ , which is specified by

$$\mu(A) = (7-2) \cdot (5-2) = 15$$

and

$$\mu(\Omega) = (10-0) \cdot (6-0) = 60.$$

This specification determines the areas of all four sets in  $\mathcal{A}$ , because  $\mu(A^c) = \mu(\Omega) - \mu(A) = 60 - 15 = 45$  and  $\mu(\emptyset) = 0$ . Hence, the measure space  $(\Omega, \mathcal{A}, \mu)$  is completely determined. Note that  $\mu$  is the restriction of the Lebesgue measure  $\lambda_2$  to the  $\sigma$ -algebra  $\mathcal{A}$ , i. e.,  $\mu(A) = \lambda_2(A)$ , for all  $A \in \mathcal{A}$ .

In Example 2.2.1 we considered the mapping  $f: \Omega \rightarrow \Omega' = \Omega$  defined by  $f(x) = \frac{3}{4}x$ . Furthermore, we considered the rectangle  $B' = ]4.5, 7.5] \times [0, 4.5]$  and the  $\sigma$ -algebra  $\mathcal{A}' = \{\Omega', \emptyset, B', (B')^c\}$ . If we specify  $\mathcal{A}$  such that  $f^{-1}(B') \in \mathcal{A}$ , then  $f$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable. In this case, all inverse images  $f^{-1}(A')$  of sets  $A' \in \mathcal{A}'$  are elements of the  $\sigma$ -algebra  $\mathcal{A}$ . Therefore, the areas  $\lambda_2[f^{-1}(A')]$  of these inverse images are defined by the measure  $\lambda_2$  on  $\mathcal{A}$  that assigns the area to *all* elements of  $\mathcal{A}$ . If we specify  $\mathcal{A} = \{\Omega, \emptyset, f^{-1}(B'), f^{-1}[(B')^c]\}$ , then

$$\lambda_{2_f}(B') = \lambda_2[f^{-1}(B')] = \lambda_2([6, 10] \times [0, 6]) = (10-6) \cdot (6-0) = 24,$$

$$\lambda_{2_f}((B')^c) = \lambda_2(f^{-1}[(B')^c]) = \lambda_2([0, 6] \times [0, 6]) = (6-0) \cdot (6-0) = 36,$$

$\lambda_{2_f}(\Omega') = 60$ , and  $\lambda_{2_f}(\emptyset) = 0$ . Then the function  $\lambda_{2_f}: \mathcal{A}' \rightarrow \mathbb{R}$  defined by

$$\lambda_{2_f}(B') = \lambda_2[f^{-1}(B')], \quad \forall B' \in \mathcal{A}',$$

is again a measure, the *image measure* of  $\lambda_2$  under  $f$ . Therefore,  $(\Omega', \mathcal{A}', \lambda_{2_f})$  is a measure space.

Note that the image measure  $\lambda_{2_f}$  on the  $\sigma$ -algebra  $\mathcal{A}'$  differs from the area measure on  $\mathcal{A}'$ . In fact, the area of  $B'$  is  $(7.5-4.5) \cdot 4.5 = 13.5$  and the area of  $(B')^c$  is  $60 - 13.5 = 46.5$ .  $\triangleleft$

**Remark 2.82 (Cumulation of the Values  $\mu(\{\omega\})$ )** If  $\{\omega\} \in \mathcal{A}$ , for all  $\omega \in \Omega$ , then

$$\mu_f(\{\omega'\}) = \sum_{\omega: f(\omega)=\omega'} \mu(\{\omega\}), \quad (2.42)$$

provided that the sum is over a finite or countable number of summands. The measure  $\mu$  assigns to the singletons and other elements  $A \in \mathcal{A}$  a nonnegative number  $\mu(A)$ , and  $f$  maps each element  $\omega \in \Omega$  to an element  $\omega'$  in  $\Omega'$ . Thereby it translates the values  $\mu(A)$  of the measure  $\mu$  to their images  $f(A)$ . In particular, this applies to the singletons  $\{\omega\}$ . This is illustrated in the following example.  $\triangleleft$

**Example 2.83 (Flipping Two Coins – continued)** In this example,

$$P(\{\omega\}) = \frac{1}{4}, \quad \forall \omega \in \Omega, \quad (2.43)$$

uniquely defines a measure  $P: \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  and the measure space  $(\Omega, \mathcal{P}(\Omega), P)$ . The reason is that the singletons  $\{\omega\}$  are pairwise disjoint and Rule (x) of Box 1.1 implies

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}), \quad \forall A \in \mathcal{A}.$$

For instance, the set  $A = \text{flipping one and only one head}$  is the union  $A = \{(h, t)\} \cup \{(t, h)\} = \{(h, t), (t, h)\}$ . Hence,

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = P(\{(h, t)\}) + P(\{(t, h)\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Now consider Figure 2.4 and realize that each arrow translates the value  $\mu(\{\omega\}) = \frac{1}{4}$  from left to right. According to Equation (2.42), this yields

$$P_X(\{0\}) = P[X^{-1}(\{0\})] = P(\{(t, t)\}) = \frac{1}{4},$$

$$P_X(\{1\}) = P[X^{-1}(\{1\})] = P(\{(t, h), (h, t)\}) = \frac{2}{4},$$

and

$$P_X(\{2\}) = P[X^{-1}(\{2\})] = P(\{(h, h)\}) = \frac{1}{4}.$$

$\triangleleft$

**Example 2.84 (Image Measure Under a Step Function)** If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$  is measurable such that  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  with pairwise different  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ ,  $\alpha_i \neq 0$ , and pairwise disjoint  $A_i \in \mathcal{A}$ ,  $i = 1, \dots, n$ , and if we define  $A_{n+1} := \Omega \setminus (\bigcup_{i=1}^n A_i)$ , and  $\alpha_{n+1} := 0$ , then the image measure is

$$\mu_f = \sum_{i=1}^{n+1} \mu(A_i) \cdot \delta_{\alpha_i} \quad (2.44)$$

(see Exercise 2-19). Equation (2.44) generalizes Equation (2.42): For all  $\omega \in A_i$

$$f(\omega) = \alpha_i \cdot 1_{A_i}(\omega) = \alpha_i.$$

Hence,  $f$  translates the value  $\mu(A_i)$  to  $\alpha_i \in \mathbb{R}$  and  $\mu_f$  assigns the value  $\mu(A_i)$  to the singleton  $\{\alpha_i\}$ ,  $i = 1, \dots, n+1$ .  $\triangleleft$

Our next theorem deals with the image measures of  $\mu$ -equivalent measurable mappings. As a random variable is a particular measurable mapping and the distribution of a random variable a particular image measure (see section 5.1), this theorem has important implications on all concepts that in some sense describe properties of distributions of random variables such as expectations, variances, covariances, etc.

**Theorem 2.85 ( $\mu$ -Equivalence Implies Equality of Image Measures)**

If  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$  are measurable mappings, then

$$f \stackrel{\mu}{=} g \quad \Rightarrow \quad \mu_f = \mu_g. \quad (2.45)$$

(Proof p. 74)

In the following theorem we present a necessary and sufficient condition for  $\mu$ -equivalence of two compositions  $g \circ f$  and  $g^* \circ f$ .

**Theorem 2.86 ( $\mu$ -Equivalence of Compositions)**

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$  and  $g, g^*: (\Omega', \mathcal{A}') \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  are measurable mappings, then

$$g \stackrel{\mu_f}{=} g^* \quad \Leftrightarrow \quad g \circ f \stackrel{\mu}{=} g^* \circ f. \quad (2.46)$$

(Proof p. 75)

## 2.7 Proofs

**Proof of Lemma 2.19**

(a) If  $f = \sum_i \alpha_i 1_{A_i}$ , then for all  $B \in \bar{\mathcal{B}}$ ,

$$f^{-1}(B) = \bigcup_{i: \alpha_i \in B} A_i \in \mathcal{A},$$

because  $\mathcal{A}$  is closed with respect to finite and countable unions.

(b) Assume that there are no  $\alpha_1, \alpha_2, \dots \in \bar{\mathbb{R}}$  such that  $f = \sum_i \alpha_i 1_{A_i}$ . Then there are an  $i$  and elements  $\omega_1, \omega_2 \in A_i$  with  $f(\omega_1) \neq f(\omega_2)$ . Applying Equation (2.5) yields

$$f^{-1}(\{f(\omega_1)\}) \cap f^{-1}(\{f(\omega_2)\}) = \emptyset.$$

Furthermore, because  $\omega_j \in f^{-1}(\{f(\omega_j)\})$ ,  $j = 1, 2$ ,

$$f^{-1}(\{f(\omega_1)\}) \cap A_i \neq \emptyset, \quad \text{and} \quad f^{-1}(\{f(\omega_2)\}) \cap A_i \neq \emptyset.$$

Therefore, we conclude:  $f^{-1}(\{f(\omega_1)\}) \notin \mathcal{A}$  and  $f^{-1}(\{f(\omega_2)\}) \notin \mathcal{A}$ . Because  $\{f(\omega_1)\}, \{f(\omega_2)\} \in \bar{\mathcal{B}}$ , it follows that  $f$  is not  $(\mathcal{A}, \bar{\mathcal{B}})$ -measurable.

**Proof of Lemma 2.29**

(i)  $\cap$ -stability of  $\mathcal{D}$ . If  $C_1, C_2 \in \mathcal{C}$ , and  $\omega'_1, \omega'_2 \in \Omega'$ , then

$$\begin{aligned} & [C_1 \cap f^{-1}(\{\omega'_1\})] \cap [C_2 \cap f^{-1}(\{\omega'_2\})] \\ &= (C_1 \cap C_2) \cap [f^{-1}(\{\omega'_1\}) \cap f^{-1}(\{\omega'_2\})] \quad [\cap \text{ is associative and commutative}] \\ &= \begin{cases} (C_1 \cap C_2) \cap f^{-1}(\{\omega'_1\}), & \text{if } \omega'_1 = \omega'_2 \\ \emptyset, & \text{if } \omega'_1 \neq \omega'_2 \end{cases} \quad [(2.5)] \end{aligned}$$

is an element of  $\mathcal{D}$ , because  $C_1 \cap C_2 \in \mathcal{C}$  and  $\emptyset \in \mathcal{C}$ , which follows from the definition of a  $\sigma$ -algebra.

(ii) Denote  $\mathcal{A}' = \mathcal{D}(\Omega')$  and define  $\sigma(\mathcal{D}) = \sigma(\mathcal{C}, f) := \sigma[\mathcal{C} \cup f^{-1}(\mathcal{A}')]$ .

(a)  $\sigma(\mathcal{D}) \subset \sigma[\mathcal{C} \cup f^{-1}(\mathcal{A}')]$ . Obviously,  $\{f^{-1}(\{\omega'\}) : \omega' \in \Omega'\} \subset f^{-1}(\mathcal{A}')$ . Therefore,

$$\begin{aligned} \mathcal{D} &= \{C \cap f^{-1}(\{\omega'\}) : \omega' \in \Omega', C \in \mathcal{C}\} \\ &\subset \{C \cap f^{-1}(A') : A' \in \mathcal{A}', C \in \mathcal{C}\} \\ &\subset \sigma[\mathcal{C} \cup f^{-1}(\mathcal{A}')]. \quad [\text{Rem. 1.2}] \end{aligned}$$

Hence, according to Remark 1.23,  $\sigma(\mathcal{D}) \subset \sigma[\mathcal{C} \cup f^{-1}(\mathcal{A}')]$ .

(b)  $\sigma[\mathcal{C} \cup f^{-1}(\mathcal{A}')] \subset \sigma(\mathcal{D})$ . Because  $\Omega \in \mathcal{C}$  and  $\Omega \in f^{-1}(\mathcal{A}')$ , all  $C \in \mathcal{C}$  and all  $f^{-1}(A')$  are elements of  $\sigma(\mathcal{D})$  (see Def. 1.1,  $\Omega'$  is finite or countable). Therefore,  $\mathcal{C} \cup f^{-1}(\mathcal{A}') \subset \sigma(\mathcal{D})$ . Proposition (1.11) then implies  $\sigma[\mathcal{C} \cup f^{-1}(\mathcal{A}')] \subset \sigma(\mathcal{D})$ .

**Proof of Lemma 2.35**

(i) We have to show that  $\mathcal{C}'_f$  satisfies conditions (a) to (c) of Definition 1.1.

(a)

$$\Omega = f^{-1}(\Omega') \in \mathcal{C} \Rightarrow \Omega' \in \mathcal{C}'_f. \quad [(2.16)]$$

(b)

$$\begin{aligned} A' \in \mathcal{C}'_f &\Rightarrow f^{-1}(A') \in \mathcal{C} \quad [(2.16)] \\ &\Rightarrow f^{-1}(A')^c = f^{-1}[(A')^c] \in \mathcal{C} \quad [\text{Def. 1.1 (b), (2.4)}] \\ &\Rightarrow (A')^c \in \mathcal{C}'_f. \quad [(2.16)] \end{aligned}$$

(c)

$$\begin{aligned} A'_1, A'_2, \dots \in \mathcal{C}'_f &\Rightarrow f^{-1}(A'_1), f^{-1}(A'_2), \dots \in \mathcal{C} \quad [(2.16)] \\ &\Rightarrow \bigcup_{i=1}^{\infty} f^{-1}(A'_i) = f^{-1}\left(\bigcup_{i=1}^{\infty} A'_i\right) \in \mathcal{C} \quad [\text{Def. 1.1 (c), (2.6)}] \\ &\Rightarrow \bigcup_{i=1}^{\infty} A'_i \in \mathcal{C}'_f. \quad [(2.16)] \end{aligned}$$

(ii) For all  $A' \in \mathcal{A}'$ ,

$$\begin{aligned} A' \in \mathcal{A}' &\Rightarrow f^{-1}(A') \in \mathcal{C} \quad [(\mathcal{C}, \mathcal{A}')\text{-measurability of } f] \\ &\Rightarrow A' \in \mathcal{C}'_f. \quad [(2.16)] \end{aligned}$$

Hence,  $(\mathcal{C}, \mathcal{A}')$ -measurability of  $f$  implies  $\mathcal{A}' \subset \mathcal{C}'_f$ .

**Proof of Lemma 2.37**

First, note that, for  $A'_i \in \mathcal{A}'_i$ ,  $i = 1, \dots, n$ ,

$$\begin{aligned}
 f^{-1}(A'_1 \times \dots \times A'_n) &= \{\omega \in \Omega: f(\omega) \in A'_1 \times \dots \times A'_n\} \\
 &= \{\omega \in \Omega: (f_1(\omega), \dots, f_n(\omega)) \in A'_1 \times \dots \times A'_n\} \\
 &= \{\omega \in \Omega: f_1(\omega) \in A'_1, \dots, f_n(\omega) \in A'_n\} \\
 &= \bigcap_{i=1}^n \{\omega \in \Omega: f_i(\omega) \in A'_i\} \\
 &= \bigcap_{i=1}^n f_i^{-1}(A'_i).
 \end{aligned} \tag{2.47}$$

Hence,

$$\begin{aligned}
 \sigma(f) &= \{f^{-1}(A'): A' \in \bigotimes_{i=1}^n \mathcal{A}'_i\} && \text{[Def. 2.26]} \\
 &= \sigma(\{f^{-1}(A'_1 \times \dots \times A'_n): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}) && \text{[Th. 2.20, Defs. 1.13, 1.31]} \\
 &= \sigma(\{f_1^{-1}(A'_1) \cap \dots \cap f_n^{-1}(A'_n): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}) && \text{[(2.47)]} \\
 &\supset \sigma\left(\bigcup_{i=1}^n \{f_i^{-1}(A'_i) \cap \bigcap_{j=1, j \neq i}^n f_j^{-1}(\Omega'_j): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}\right) && \text{[Rem. 1.23]} \\
 &= \sigma\left(\bigcup_{i=1}^n \{f_i^{-1}(A'_i): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}\right) && \text{[}f_j^{-1}(\Omega'_j) = \Omega\text{]} \\
 &= \sigma\left(\bigcup_{i=1}^n \sigma(f_i)\right). && \text{[Def. 2.26]}
 \end{aligned}$$

Furthermore,

$$\begin{aligned}
 &\{f_1^{-1}(A'_1) \cap \dots \cap f_n^{-1}(A'_n): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\} \\
 &\subset \sigma\left(\bigcup_{i=1}^n \{f_i^{-1}(A'_i): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}\right). && \text{[Rem. 1.2, finite intersections]}
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \sigma\left(\bigcup_{i=1}^n \sigma(f_i)\right) &= \sigma\left(\bigcup_{i=1}^n \{f_i^{-1}(A'_i): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}\right) && \text{[Def. 2.26]} \\
 &\supset \sigma(\{f_1^{-1}(A'_1) \cap \dots \cap f_n^{-1}(A'_n): A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}) && \text{[Rem. 1.23]} \\
 &= \sigma(f).
 \end{aligned}$$

Hence,  $\sigma(f) = \sigma(\bigcup_{i=1}^n \sigma(f_i))$ .

**Proof of Theorem 2.38**

(b)  $\Rightarrow$  (a): For all  $i = 1, \dots, n$ : Let  $A'_i \in \mathcal{A}'_i$ . If  $f_i$  is measurable, then  $f_i^{-1}(A'_i) \in \mathcal{A}$ . Hence,

$$\begin{aligned}
 f^{-1}(A'_1 \times \dots \times A'_n) &= \{\omega \in \Omega: f(\omega) \in A'_1 \times \dots \times A'_n\} \\
 &= \bigcap_{i=1}^n f_i^{-1}(A'_i) \in \mathcal{A}.
 \end{aligned}$$

Because  $\{A'_1 \times \dots \times A'_n : A'_i \in \mathcal{A}'_i, i = 1, \dots, n\}$  is a generating system of  $\bigotimes_{i=1}^n \mathcal{A}'_i$ , Theorem 2.20 implies that  $f$  is measurable.

(a)  $\Rightarrow$  (b): If  $f$  is measurable, then for all  $i = 1, \dots, n$ ,

$$\begin{aligned} f_i^{-1}(\mathcal{A}'_i) &= \{f_i^{-1}(A'_i) : A'_i \in \mathcal{A}'_i\} \\ &= \{f_i^{-1}(A'_i) \cap \bigcap_{j=1, j \neq i}^n f_j^{-1}(\Omega'_j) : A'_i \in \mathcal{A}'_i\} \\ &= \{f^{-1}(\Omega'_1 \times \dots \times \Omega'_{i-1} \times A'_i \times \Omega'_{i+1} \times \dots \times \Omega'_n) : A'_i \in \mathcal{A}'_i\} \quad [(2.47)] \\ &\subset f^{-1}\left(\bigotimes_{i=1}^n \mathcal{A}'_i\right) \subset \mathcal{A}. \end{aligned}$$

### **Proof of Lemma 2.42**

Consider

$$\pi = (\pi_1, \dots, \pi_n) : \left( \prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i \right) \rightarrow \left( \prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathcal{A}_i \right).$$

Analogously to the proof of Lemma 2.37 and using Definition 1.31,

$$\begin{aligned} \bigotimes_{i=1}^n \mathcal{A}_i &= \sigma(\{A_1 \times \dots \times A_n : A_i \in \mathcal{A}_i, i = 1, \dots, n\}) \\ &= \sigma(\{\pi^{-1}(A_1 \times \dots \times A_n) : A_i \in \mathcal{A}_i, i = 1, \dots, n\}) \quad [(2.20), (2.21)] \\ &= \sigma(\pi) \quad [\text{Th. 2.20}, (2.12)] \\ &= \sigma(\pi_i, i = 1, \dots, n). \quad [\text{Lem. 2.37}, (2.18)] \end{aligned}$$

### **Proof of Corollary 2.44**

If  $\Omega'$  is finite or countable and we consider the measurable space  $(\Omega', \mathcal{P}(\Omega'))$ , then  $\sigma(f) = \sigma[\{f^{-1}(\{\omega'\}) : \omega' \in \Omega'\}]$  [see Lemma 2.29 with  $\mathcal{C} = \{\Omega, \emptyset\}$ ]. Because  $\{f^{-1}(\{\omega'\}) : \omega' \in \Omega'\}$  is a finite or countable partition of  $\Omega$ , this corollary is an immediate implication of Lemma 2.19.

### **Proof of Lemma 2.46**

For all  $\omega \in \Omega$  and all  $\omega' \in \Omega'$ ,

$$g[f(\omega)] \cdot 1_{f=\omega'}(\omega) = g(\omega') \cdot 1_{f=\omega'}(\omega) = \begin{cases} 0, & \text{if } f(\omega) \neq \omega' \\ g(\omega'), & \text{if } f(\omega) = \omega'. \end{cases} \quad (2.48)$$

This equation is equivalent to

$$g(f) \cdot 1_{f=\omega'} = g(\omega') \cdot 1_{f=\omega'}. \quad (2.49)$$

Because the set  $\{f^{-1}(\{\omega'\}) : \omega' \in \Omega'\}$  is a finite or countable partition of  $\Omega$  we can conclude:

$$1_\Omega = \sum_{\omega' \in \Omega'} 1_{f=\omega'}.$$

Therefore,

$$g(f) = g(f) \cdot 1_{\Omega} = \sum_{\omega' \in \Omega'} g(f) \cdot 1_{f=\omega'} = \sum_{\omega' \in \Omega'} g(\omega') \cdot 1_{f=\omega'},$$

and this implies Equation (2.27).

### **Proof of Theorem 2.49**

If  $f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ ,  $g: (\Omega', \mathcal{A}') \rightarrow (\Omega'', \mathcal{A}'')$  are measurable mappings, then, according to Corollary 2.28,  $f^{-1}(\mathcal{A}') \subset \mathcal{A}$  and  $g^{-1}(\mathcal{A}'') \subset \mathcal{A}'$ . Hence, for all  $A'' \in \mathcal{A}''$ ,

$$\begin{aligned} (g \circ f)^{-1}(A'') &= \{\omega \in \Omega: g[f(\omega)] \in A''\} \\ &= \{\omega \in \Omega: f(\omega) \in g^{-1}(A'')\} \\ &= f^{-1}[g^{-1}(A'')]. \end{aligned}$$

Furthermore,

$$\begin{aligned} f^{-1}[g^{-1}(\mathcal{A}'')] &\subset f^{-1}(\mathcal{A}') && [(2.15)] \\ &\subset \mathcal{A}. && [(2.15)] \end{aligned}$$

### **Proof of Theorem 2.79**

We show that  $\mu_f$  has the properties (a) to (c) required in Definition 1.43. For each property of  $\mu_f$  we use the corresponding property of  $\mu$ .

- (a)  $\mu_f(\emptyset) = \mu[f^{-1}(\emptyset)] = \mu(\emptyset) = 0$ .
- (b) For all  $A' \in \mathcal{A}'$ :  $\mu_f(A') = \mu[f^{-1}(A')] \geq 0$ .
- (c) If  $A'_1, A'_2, \dots \in \mathcal{A}'$  are pairwise disjoint, then, according to Equation (2.5), for  $i \neq j$ ,

$$f^{-1}(A'_i) \cap f^{-1}(A'_j) = f^{-1}(A'_i \cap A'_j) = f^{-1}(\emptyset),$$

i. e., the inverse images  $f^{-1}(A'_1), f^{-1}(A'_2), \dots$  are pairwise disjoint as well. Therefore,

$$\mu_f\left(\bigcup_{i=1}^{\infty} A'_i\right) = \mu\left(f^{-1}\left(\bigcup_{i=1}^{\infty} A'_i\right)\right) \quad [(2.41)]$$

$$= \mu\left(\bigcup_{i=1}^{\infty} f^{-1}(A'_i)\right) \quad [(2.6)]$$

$$= \sum_{i=1}^{\infty} \mu(f^{-1}(A'_i)) \quad [\text{Def. 1.43 (c)}]$$

$$= \sum_{i=1}^{\infty} \mu_f(A'_i). \quad [(2.41)]$$

### **Proof of Theorem 2.85**

If  $f \stackrel{\mu}{=} g$ , then there is a set  $A$  satisfying

$$f(\omega) = g(\omega), \quad \forall \omega \in \Omega \setminus A \text{ and } \mu(A) = 0.$$

Monotonicity of  $\mu$  implies  $\mu(\{\omega \in A: f(\omega) \in A'\}) = 0 = \mu(\{\omega \in A: g(\omega) \in A'\})$  for all  $A' \in \mathcal{A}'$ . Hence, using additivity of  $\mu$ ,

$$\begin{aligned}
\mu_f(A') &= \mu[f^{-1}(A')] \\
&= \mu(\{\omega \in \Omega \setminus A: f(\omega) \in A'\}) + \mu(\{\omega \in A: f(\omega) \in A'\}) \\
&= \mu(\{\omega \in \Omega \setminus A: g(\omega) \in A'\}) + \mu(\{\omega \in A: g(\omega) \in A'\}) \\
&= \mu[g^{-1}(A')] = \mu_g(A').
\end{aligned}$$

### ***Proof of Theorem 2.86***

For measurable functions  $g, g^*: (\Omega', \mathcal{A}') \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  define  $A' := \{\omega' \in \Omega': g(\omega') \neq g^*(\omega')\}$ . Note that  $A' \in \mathcal{A}'$  [see Rem. 2.67 (c)]. Then

$$\begin{aligned}
f^{-1}(A') &= \{\omega \in \Omega: f(\omega) \in A'\} \\
&= \{\omega \in \Omega: g[f(\omega)] \neq g^*[f(\omega)]\} \\
&= \{\omega \in \Omega: (g \circ f)(\omega) \neq (g^* \circ f)(\omega)\}.
\end{aligned}$$

Hence,  $g \stackrel{\mu_f}{=} g^* \Leftrightarrow \mu_f(A') = 0 \Leftrightarrow \mu[f^{-1}(A')] = 0 \Leftrightarrow g \circ f \stackrel{\mu}{=} g^* \circ f$ .

## **2.8 Exercises**

▷ **Exercise 2-1** Prove Equations (2.4) to (2.6).

▷ **Exercise 2-2** Consider Example 2.2.1 and compute the inverse images of the sets  $\{(4.5, 0)\}$ ,  $\{(7.5, 0)\}$ ,  $\{(7.5, 4.5)\}$ , and  $\{(4.5, 4.5)\}$  under the function

$$f(x_1, x_2) = \frac{3}{4} \cdot (x_1, x_2) = \left(\frac{3}{4} \cdot x_1, \frac{3}{4} \cdot x_2\right).$$

▷ **Exercise 2-3** Consider Example 2.2.1 and specify the inverse images of the rectangles  $[8, 10] \times [0, 2]$  and  $[3, 7.5] \times [0, 3]$  under the function  $f: \Omega \rightarrow \Omega'$  defined by  $f(x_1, x_2) = \frac{3}{4} \cdot (x_1, x_2)$ .

▷ **Exercise 2-4** Consider Example 2.2.1 and use Equation (2.4) to determine the inverse image  $f^{-1}[(C')^c]$ .

▷ **Exercise 2-5** Prove the proposition of Example 2.14.

▷ **Exercise 2-6** Prove the proposition of Example 2.15.

▷ **Exercise 2-7** Prove the proposition of Example 2.17.

▷ **Exercise 2-8** Prove the proposition of Remark 2.18.

▷ **Exercise 2-9** Prove the proposition formulated in Remark 2.33.

▷ **Exercise 2-10** Consider Example 2.40 and show that  $X$ ,  $Y$ , and  $(X, Y)$  are measurable with respect to  $\mathcal{A}$  whenever the two inverse images  $X^{-1}(\{1\})$  and  $Y^{-1}(\{1\})$  are elements of  $\mathcal{A}$ .

▷ **Exercise 2-11** In Example 2.55 we considered  $\Omega = [0, 4]$ ,  $A_1 = [0, 1]$ ,  $A_2 = ]1, 2]$ ,  $A_3 = ]2, 3]$ , and  $A_4 = ]3, 4]$ . There, we also defined the functions  $f$  and  $h$ . Show that  $\sigma(h) = \sigma(\{A_j \cup A_{j+1}: j \in J\})$  and  $\sigma(f) = \sigma(\{A_i: i = 1, \dots, 4\})$ . Furthermore, show  $\sigma(h) \subset \sigma(f)$ .

- ▷ **Exercise 2-12** Prove the propositions of Remark 2-12.
- ▷ **Exercise 2-13** Prove the proposition of Example 2.56.
- ▷ **Exercise 2-14** Prove the propositions of Remark 2.72.
- ▷ **Exercise 2-15** Consider Remark 2.73 and show: If  $\mathcal{M}$  is a set of mappings  $(\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$ , then  $\equiv_{\mu}$  is an equivalence relation on  $\mathcal{M}$ .
- ▷ **Exercise 2-16** Show that  $\{C(f): f \in \mathcal{M}\}$  is a partition of  $\mathcal{M}$  (see Remark 2.75).
- ▷ **Exercise 2-17** Prove the propositions of Remark 2.76.
- ▷ **Exercise 2-18** Prove proposition (2.40).
- ▷ **Exercise 2-19** Prove the proposition of Example 2.84.

## Solutions

- ▷ **Solution 2-1** Equation (2.4):

$$f^{-1}[(A')^c] = \{\omega \in \Omega: f(\omega) \in (A')^c\} = \{\omega \in \Omega: f(\omega) \notin A'\} = [f^{-1}(A')]^c.$$

Equation (2.5):

$$\begin{aligned} f^{-1}\left(\bigcap_{i \in I} A'_i\right) &= \{\omega \in \Omega: f(\omega) \in \bigcap_{i \in I} A'_i\} = \{\omega \in \Omega: f(\omega) \in A'_i, \forall i \in I\} \\ &= \bigcap_{i \in I} \{\omega \in \Omega: f(\omega) \in A'_i\} = \bigcap_{i \in I} f^{-1}(A'_i). \end{aligned}$$

Equation (2.6):

$$\begin{aligned} f^{-1}\left(\bigcup_{i \in I} A'_i\right) &= \{\omega \in \Omega: f(\omega) \in \bigcup_{i \in I} A'_i\} = \{\omega \in \Omega: \exists i \in I: f(\omega) \in A'_i\} \\ &= \bigcup_{i \in I} \{\omega \in \Omega: f(\omega) \in A'_i\} = \bigcup_{i \in I} f^{-1}(A'_i). \end{aligned}$$

- ▷ **Solution 2-2** The inverse images are the sets  $f^{-1}\{(4.5, 0)\} = \{(6, 0)\}$ ,  $f^{-1}\{(7.5, 0)\} = \{(10, 0)\}$ ,  $f^{-1}\{(7.5, 4.5)\} = \{(10, 6)\}$ , and  $f^{-1}\{(4.5, 4.5)\} = \{(6, 6)\}$ .
- ▷ **Solution 2-3**  $f^{-1}([8, 10] \times [0, 2]) = \emptyset$  and  $f^{-1}([3, 7.5] \times [0, 3]) = [4, 10] \times [0, 4]$ .
- ▷ **Solution 2-4** According to Equation (2.4), the inverse image of  $(C')^c$  under  $f$  is

$$\begin{aligned} f^{-1}[(C')^c] &= [f^{-1}(C')]^c \\ &= \Omega \setminus ([6, 10] \times [0, 6]) \\ &= ([0, 10] \times [0, 6]) \setminus ([6, 10] \times [0, 6]) \\ &= [0, 6] \times [0, 6]. \end{aligned}$$

▷ **Solution 2-5** If  $f: \Omega \rightarrow \Omega'$  is constant, then, according to Example 2.10, it is  $(\mathcal{A}, \mathcal{A}')$ -measurable for  $\mathcal{A} = \{\Omega, \emptyset\}$ . Now, assume that  $f$  is not constant, i. e.,  $\exists \omega_1, \omega_2 \in \Omega: f(\omega_1) \neq f(\omega_2)$ . According to our assumptions,

$$\{f(\omega_1)\}, \{f(\omega_2)\} \in \mathcal{A}'.$$

Furthermore,  $\omega_i \in f^{-1}\{f(\omega_i)\}$ , for  $i = 1, 2$ , i. e., the inverse images are nonempty sets. Now,  $f(\omega_1) \neq f(\omega_2)$  implies

$$\{f(\omega_1)\} \cap \{f(\omega_2)\} = \emptyset,$$

and, using Equation (2.5),

$$f^{-1}(\{f(\omega_1)\}) \cap f^{-1}(\{f(\omega_2)\}) = f^{-1}(\{f(\omega_1)\} \cap \{f(\omega_2)\}) = f^{-1}(\emptyset) = \emptyset.$$

Hence, the inverse images are nonempty disjoint sets, and therefore none of them is in  $\mathcal{A} = \{\Omega, \emptyset\}$ . This implies that  $f$  is not  $(\mathcal{A}, \mathcal{A}')$ -measurable if it is not constant.

▷ **Solution 2-6** We consider  $\{A, A^c, \Omega, \emptyset\}$ . If  $f = \alpha_1 1_A + \alpha_2 1_{A^c}$ , then for all  $A' \in \mathcal{B}$ ,

$$f^{-1}(A') = \begin{cases} \emptyset, & \text{if } \alpha_1 \notin A', \alpha_2 \notin A' \\ A, & \text{if } \alpha_1 \in A', \alpha_2 \notin A' \\ A^c, & \text{if } \alpha_1 \notin A', \alpha_2 \in A' \\ \Omega, & \text{if } \alpha_1 \in A', \alpha_2 \in A'. \end{cases}$$

Hence,  $f$  is  $(\mathcal{A}, \mathcal{B})$ -measurable. (Note that this also holds if  $A = \emptyset$  or  $A = \Omega$ , and also if  $\alpha_1 = \alpha_2$ .)

Now assume that  $f$  is an  $(\mathcal{A}, \mathcal{B})$ -measurable function.

(a) If  $f$  takes one only one single value, say  $\alpha$ , then

$$f = \alpha 1_\Omega = \alpha_1 1_A + \alpha_2 1_{A^c}, \quad \text{with } \alpha_1 = \alpha_2 = \alpha.$$

(b) If  $f$  takes on exactly two different values  $\beta_1 \neq \beta_2$ , then  $f^{-1}(\{\beta_1, \beta_2\}) = f^{-1}(\{\beta_1\}) \cup f^{-1}(\{\beta_2\}) = \Omega$ , and according to Equation (2.5),  $f^{-1}(\{\beta_1\}) \cap f^{-1}(\{\beta_2\}) = \emptyset$ , and  $f^{-1}(\{\beta_i\}) \neq \emptyset$ , for  $i = 1, 2$ . Hence,  $f$  is  $(\mathcal{A}, \mathcal{B})$ -measurable if and only if

$$f^{-1}(\{\beta_1\}) = A \quad \text{or} \quad f^{-1}(\{\beta_1\}) = A^c \quad \text{and} \quad A, A^c \neq \emptyset.$$

This implies

$$f = \beta_1 1_A + \beta_2 1_{A^c} \quad \text{or} \quad f = \beta_2 1_A + \beta_1 1_{A^c},$$

respectively.

(c) If  $f$  takes on *three or more* pairwise different values, then, using the same kind of argument as in (a), we can conclude that there are at least three pairwise disjoint and nonempty inverse images under  $f$ , say  $A_1, A_2, A_3 \subset \Omega$ . Hence, in this case  $f$  is not  $(\mathcal{A}, \mathcal{B})$ -measurable.

▷ **Solution 2-7** If  $A_1, \dots, A_n \in \mathcal{A}$  are pairwise disjoint and we define  $A_{n+1} := \Omega \setminus (\bigcup_{i=1}^n A_i)$ , and  $\alpha_{n+1} := 0$ , then

$$f = \sum_{i=1}^n \alpha_i 1_{A_i} = \sum_{i=1}^{n+1} \alpha_i 1_{A_i}.$$

Because  $A_1, \dots, A_{n+1}$  are pairwise disjoint and  $\bigcup_{i=1}^{n+1} A_i = \Omega$ , there is, for all  $\omega \in \Omega$ , exactly one  $i \in \{1, \dots, n+1\}$  such that  $\omega \in A_i$ , and therefore  $f(\omega) = \alpha_i$ . Hence, the codomain of  $f$  is  $\{\alpha_1, \dots, \alpha_{n+1}\}$ . Vice versa, for all  $\alpha_i, i = 1, \dots, n+1$ , we obtain the inverse image

$$f^{-1}(\{\alpha_i\}) = \{\omega \in \Omega: f(\omega) = \alpha_i\} = \bigcup_{j: \alpha_j = \alpha_i} A_j \quad (2.50)$$

(If the  $\alpha_1, \dots, \alpha_{n+1}$  are pairwise different, then  $f^{-1}(\{\alpha_i\}) = A_i$ .) Now, for all  $A' \subset \mathbb{R}$ ,

$$\begin{aligned} f^{-1}(A') &= \{\omega \in \Omega: f(\omega) \in A'\} && \text{[Def. 2.1]} \\ &= \{\omega \in \Omega: f(\omega) \in \bigcup_{i: \alpha_i \in A'} \{\alpha_i\}\} && \text{[codomain of } f \text{ is } \{\alpha_1, \dots, \alpha_{n+1}\}] \\ &= f^{-1}\left(\bigcup_{i: \alpha_i \in A'} \{\alpha_i\}\right) && \text{[Def. 2.1]} \\ &= \bigcup_{i: \alpha_i \in A'} f^{-1}(\{\alpha_i\}) && \text{[(2.6)]} \\ &= \bigcup_{i: \alpha_i \in A'} A_i. && \text{[(2.50)]} \end{aligned}$$

▷ **Solution 2-8** If  $A_1, \dots, A_n \in \mathcal{A}$  are pairwise disjoint and if we define  $A_{n+1} := \Omega \setminus (\bigcup_{i=1}^n A_i)$ , and  $\alpha_{n+1} := 0$ , then

$$\forall A' \in \mathcal{B}: f^{-1}(A') = \bigcup_{\substack{i=1, \dots, n+1, \\ \alpha_i \in A'}} A_i \in \mathcal{A},$$

[see Eq. (2.10)].

If  $A_1, \dots, A_n \in \mathcal{A}$  are *not* pairwise disjoint, define the  $2^n$  sets

$$B_j := A_1^{c_1(j)} \cap \dots \cap A_n^{c_n(j)} \in \mathcal{A}, \quad j = 1, \dots, 2^n,$$

with  $(c_1(j), \dots, c_n(j)) \in \{0, 1\}^n$  and

$$A_i^0 := A_i, \quad A_i^1 := A_i^c.$$

Note that some of the sets  $B_j$  can be empty. Then

$$f = \sum_{i=1}^n \alpha_i 1_{A_i} = \sum_{j=1}^{2^n} \beta_j 1_{B_j},$$

with  $\beta_j = \sum_{i: c_i(j)=0} \alpha_i$ . Because  $B_1, \dots, B_{2^n}$  are pairwise disjoint and  $\bigcup_{j=1}^{2^n} B_j = \Omega$ , the function  $f$  is  $(\mathcal{A}, \mathcal{B})$ -measurable (see the first part of this solution).

▷ **Solution 2-9** If all values of  $f$  are elements of  $\Omega'$ , then

$$f^{-1}(B) = f^{-1}(\Omega' \cap B), \quad \forall B \in \mathcal{B}.$$

Therefore,  $f^{-1}(\mathcal{B}) = f^{-1}(\mathcal{B}|_{\Omega'})$ , where  $\mathcal{B}|_{\Omega'}$  denotes the trace of  $\mathcal{B}$  in  $\Omega'$  (see Example 1.10). Note that  $\mathcal{B}|_{\Omega'} = \mathcal{P}(\Omega')$  (see Exercise 1-13). Hence,  $f^{-1}(\mathcal{B}) = f^{-1}(\mathcal{B}|_{\Omega'}) = f^{-1}[\mathcal{P}(\Omega')]$ .

▷ **Solution 2-10** First of all note that  $X, Y$  are measurable with respect to  $\mathcal{A}$  if and only if and  $(X, Y)$  is measurable with respect to  $\mathcal{A}$  (see Theorem 2.38). Hence, it suffices to show that  $X$  is measurable with respect to  $\mathcal{A}$  if  $X^{-1}(\{1\}) \in \mathcal{A}$ . Because  $X: \Omega \rightarrow \mathbb{R}$  is an indicator,  $X^{-1}(\{0\}) = X^{-1}(\{1\}^c) = (X^{-1}(\{1\}))^c$  [see Eq. (2.5)]. Hence, if  $X^{-1}(\{1\}) \in \mathcal{A}$ , then  $(X^{-1}(\{1\}))^c = X^{-1}(\{0\}) \in \mathcal{A}$ . Furthermore, for all  $B \in \mathcal{B}$ ,

$$X^{-1}(B) = \begin{cases} \emptyset, & \text{if } 0 \notin B, 1 \notin B, \\ X^{-1}(\{1\}), & \text{if } 0 \notin B, 1 \in B \\ \Omega \setminus X^{-1}(\{1\}), & \text{if } 0 \in B, 1 \notin B \\ \Omega, & \text{if } \{0, 1\} \subset B. \end{cases}$$

(The proof for  $Y$  is analog.)

▷ **Solution 2-11** Because the codomain of  $f$  is  $\mathbb{R}$ ,

$$\begin{aligned} \sigma(f) &= \sigma\left(f^{-1}(\mathcal{P}(\{1, 2.5, 2, 0.5\}))\right) && \text{[Rem. 2.33]} \\ &= \sigma\left(\{f^{-1}(\{1\}), f^{-1}(\{2.5\}), f^{-1}(\{2\}), f^{-1}(\{0.5\})\}\right) && \text{[(2.12)]} \\ &= \sigma(\{A_1, A_2, A_3, A_4\}). \end{aligned}$$

Analogously,

$$\begin{aligned} \sigma(h) &= \sigma\left(h^{-1}(\mathcal{P}(\{1.5, 3\}))\right) && \text{[Rem. 2.33]} \\ &= \sigma\left(\{h^{-1}(\{1.5\}), h^{-1}(\{3\})\}\right) && \text{[(2.12)]} \\ &= \sigma(\{A_1 \cup A_2, A_3 \cup A_4\}). \end{aligned}$$

Because

$$\{A_1 \cup A_2, A_3 \cup A_4\} \subset \sigma(\{A_1, A_2, A_3, A_4\}) \quad \text{[Rem. 1.2]}$$

monotonicity of the generated  $\sigma$ -algebras (see Remark 1.23) implies  $\sigma(h) \subset \sigma(f)$ .

▷ **Solution 2-12** Denote  $A_\infty := \{\omega \in \Omega: f(\omega) = \infty\}$ ,  $A_{-\infty} := \{\omega \in \Omega: f(\omega) = -\infty\}$ ,  $B_\infty := \{\omega \in \Omega: g(\omega) = \infty\}$ , and  $B_{-\infty} := \{\omega \in \Omega: g(\omega) = -\infty\}$ . Because  $\{\infty, -\infty\} \in \overline{\mathcal{B}}$ , all four sets defined above are elements of  $\mathcal{A}$ . Furthermore,

$$A := \{\omega \in \Omega: -\infty < f(\omega) < \infty\} = f^{-1}(\mathbb{R}) \in \mathcal{A}$$

and

$$B := \{\omega \in \Omega: -\infty < g(\omega) < \infty\} = g^{-1}(\mathbb{R}) \in \mathcal{A}.$$

(a) Now

$$\begin{aligned} \{\omega \in \Omega: f(\omega) \geq g(\omega)\} &= A_\infty \cup \{\omega \in A \cap B: f(\omega) \geq g(\omega)\} \cup B_{-\infty} \\ &= A_\infty \cup \{\omega \in A \cap B: f(\omega) - g(\omega) \geq 0\} \cup B_{-\infty} \\ &= A_\infty \cup [1_{A \cap B} \cdot (f - g)]^{-1}([0, \infty[) \cup B_{-\infty} && \text{[Def. 2.1]} \\ &\in \mathcal{A}. && \text{[Th. 2.57]} \end{aligned}$$

(b) Analogously,

$$\begin{aligned} \{\omega \in \Omega: f(\omega) > g(\omega)\} &= (A_\infty \cap (B \cup B_{-\infty})) \cup \{\omega \in A \cap B: f(\omega) > g(\omega)\} \\ &= (A_\infty \cap (B \cup B_{-\infty})) \cup [1_{A \cap B} \cdot (f - g)]^{-1}([0, \infty[) && \text{[Def. 2.1]} \\ &\in \mathcal{A}. && \text{[Th. 2.57]} \end{aligned}$$

(c) Finally,

$$\begin{aligned} \{\omega \in \Omega: f(\omega) = g(\omega)\} &= \{\omega \in \Omega: f(\omega) \geq g(\omega)\} \setminus \{\omega \in \Omega: f(\omega) > g(\omega)\} \\ &\in \mathcal{A}. && \text{[Rem. 1.2]} \end{aligned}$$

▷ **Solution 2-13** For any real-valued measurable function  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ , Lemma 2.52 yields  $\sigma(f^2) \subset \sigma(f)$ , because  $f^2 = g(f)$  for  $g: (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$  with the measurable function  $g(x) = x^2$ , for all  $x \in \mathbb{R}$ . [Note that  $g$  is a continuous function that is  $\mathcal{B}$ -measurable (see Klenke, 2013, Th. 1.88, p. 36)].

(i) We assume that  $f$  is nonnegative and measurable. Then  $f^2(\omega) = x$  if and only if  $f(\omega) = \sqrt{x}$ , for all  $x \geq 0$ . Hence, for all  $A \in \mathcal{A}$ ,

$$\begin{aligned} A \in \sigma(f) &\Rightarrow \exists B_1 \in \mathcal{B}: A = f^{-1}(B_1) \\ &\Rightarrow \exists B_2 \in \mathcal{B}: A = (f^2)^{-1}(B_2) && [\text{choose } B_2 := g^{-1}(B_1)] \\ &\Rightarrow A \in \sigma(f^2). \end{aligned}$$

This implies  $\sigma(f) \subset \sigma(f^2)$ .

(ii) Assume that there are  $\omega_1, \omega_2 \in \Omega$  with  $f(\omega_1) < 0 < f(\omega_2)$  and  $f^2(\omega_1) = f^2(\omega_2)$ . Then  $A := f^{-1}(-\infty, 0)$  implies  $A \in \sigma(f^2)$ , and  $\omega_1 \in A$  and  $\omega_2 \notin A$ . Furthermore, for all  $B \in \mathcal{B}: \{\omega_1, \omega_2\} \subset (f^2)^{-1}(B)$  if  $f^2(\omega_1) \in B$  and  $\{\omega_1, \omega_2\} \cap (f^2)^{-1}(B) = \emptyset$  if  $f^2(\omega_1) \notin B$ . Hence,  $A \notin (f^2)^{-1}(\mathcal{B})$ .

▷ **Solution 2-14** (i) If  $f \stackrel{\mu}{=} g$ , then there is a set  $B \in \mathcal{A}$  with  $\mu(B) = 0$  and  $f(\omega) = g(\omega)$  for all  $\omega \in \Omega \setminus B$ . Hence,  $1_A(\omega) \cdot f(\omega) = 1_A(\omega) \cdot g(\omega)$  for all  $\omega \in \Omega \setminus B$ . According to Definition 2.68,  $1_A f \stackrel{\mu}{=} 1_A g$ .

(ii) Note that  $A_h := \{\omega \in \Omega: h[f(\omega)] \neq h[g(\omega)]\} \subset \{\omega \in \Omega: f(\omega) \neq g(\omega)\} =: A$ . Therefore,  $\mu(A) = 0$  implies  $\mu(A_h) = 0$  [see Box 1.1 (v)].

▷ **Solution 2-15** Reflexivity.  $\mu(\{\omega \in \Omega: f(\omega) \neq f(\omega)\}) = \mu(\emptyset) = 0$ . Hence,  $f \stackrel{\mu}{=} f$ .

Symmetry. Assume that  $f, g \in \mathcal{M}$  and  $f \stackrel{\mu}{=} g$ . Then

$$\mu(\{\omega \in \Omega: g(\omega) \neq f(\omega)\}) = \mu(\{\omega \in \Omega: f(\omega) \neq g(\omega)\}) = 0.$$

Hence,  $g \stackrel{\mu}{=} f$ .

Transitivity. Assume that  $f, g, h \in \mathcal{M}$ ,  $f \stackrel{\mu}{=} g$ , and  $g \stackrel{\mu}{=} h$ . Then transitivity of  $=$  and subadditivity of  $\mu$  yield

$$\begin{aligned} \mu(\{\omega \in \Omega: f(\omega) \neq h(\omega)\}) &\leq \mu(\{\omega \in \Omega: f(\omega) \neq g(\omega)\} \cup \{\omega \in \Omega: g(\omega) \neq h(\omega)\}) \\ &\leq \mu(\{\omega \in \Omega: f(\omega) \neq g(\omega)\}) + \mu(\{\omega \in \Omega: g(\omega) \neq h(\omega)\}) \\ &= 0 + 0 \quad [f \stackrel{\mu}{=} g, g \stackrel{\mu}{=} h] \\ &= 0. \end{aligned}$$

Therefore,  $f \stackrel{\mu}{=} h$ .

▷ **Solution 2-16** (a)  $\forall f \in \mathcal{M}: f \stackrel{\mu}{=} f$  (reflexivity). This implies:  $\forall f \in \mathcal{M}: f \in C(f)$  and therefore  $\forall f \in \mathcal{M}: C(f) \neq \emptyset$ .

(b) Let  $f, g \in \mathcal{M}$ . We consider two cases,  $f \stackrel{\mu}{=} g$  and  $\neg(f \stackrel{\mu}{=} g)$ .  
 $f \stackrel{\mu}{=} g$ . Transitivity implies:  $\forall h \in \mathcal{M}: f \stackrel{\mu}{=} h$  if and only if  $g \stackrel{\mu}{=} h$ . Hence,  $\forall h \in \mathcal{M}: h \in C(f)$  if and only if  $h \in C(g)$ . This means that  $C(f) = C(g)$ .

$\neg(f \stackrel{\mu}{=} g)$ . We show  $C(f) \cap C(g) = \emptyset$  by contraposition. Assume:  $\exists h \in \mathcal{M}: h \in C(f) \cap C(g)$ . Then  $f \stackrel{\mu}{=} h$ ,  $h \stackrel{\mu}{=} g$ , and transitivity implies:  $f \stackrel{\mu}{=} g$ , which is a contradiction to  $\neg(f \stackrel{\mu}{=} g)$ .

(c) In part (a) we showed that  $\forall f \in \mathcal{M}: f \in C(f)$ . Therefore,  $\forall f \in \mathcal{M}: f \in \bigcup_{f^* \in \mathcal{M}} C(f^*)$ .

▷ **Solution 2-17** (i) Suppose that  $\mu \neq 0$ ,  $\alpha, \beta \in \Omega'$  and  $f \stackrel{\mu}{=} \alpha \wedge g \stackrel{\mu}{=} \beta \wedge f \stackrel{\mu}{=} g$ . If  $\alpha \neq \beta$ , then subadditivity and monotonicity of  $\mu$  yield

$$\begin{aligned}
0 &< \mu(\Omega) \\
&= \mu(\{\omega \in \Omega: f(\omega) = \alpha \wedge g(\omega) = \beta\}) + \mu(\{\omega \in \Omega: f(\omega) \neq \alpha \vee g(\omega) \neq \beta\}) \quad [\text{Box 1.1 (iv)}] \\
&= \mu(\{\omega \in \Omega: f(\omega) = \alpha \wedge g(\omega) = \beta\}) \quad [f \stackrel{\mu}{=} \alpha, g \stackrel{\mu}{=} \beta] \\
&\leq \mu(\{\omega \in \Omega: f(\omega) \neq g(\omega)\}) \quad [\alpha \neq \beta] \\
&= 0, \quad [f \stackrel{\mu}{=} g]
\end{aligned}$$

which proves (i) by contraposition.

(ii) If  $f \stackrel{\mu}{=} f^*$  and  $g \stackrel{\mu}{=} g^*$  and  $A_0 := \{\omega \in \Omega: f(\omega) \neq f^*(\omega) \vee g(\omega) \neq g^*(\omega)\}$ , then

$$\begin{aligned}
\mu(A_0) &= \mu(\{\omega \in \Omega: f(\omega) \neq f^*(\omega)\} \cup \{\omega \in \Omega: g(\omega) \neq g^*(\omega)\}) \\
&\leq \mu(\{\omega \in \Omega: f(\omega) \neq f^*(\omega)\}) + \mu(\{\omega \in \Omega: g(\omega) \neq g^*(\omega)\}) \quad [\text{Box 1.1 (vii), (v)}] \\
&= 0. \quad [f \stackrel{\mu}{=} f^*, g \stackrel{\mu}{=} g^*]
\end{aligned}$$

Note that  $\{\omega \in \Omega: f(\omega) + g(\omega) \neq f^*(\omega) + g^*(\omega)\} \subset A_0$ , and that this also holds for the corresponding sets for the difference, product, and ratio. This implies Equations (2.36) and (2.37).

(iii), (iv) Denote  $I := \{i = 1, \dots, n\}$  for (iii) and  $I := \mathbf{N}$  for (iv), respectively. Furthermore, define

$$A_0 := \bigcup_{i \in I} \{\omega \in \Omega: f_i(\omega) \neq f_i^*(\omega)\} = \{\omega \in \Omega: \exists i \in I: f_i(\omega) \neq f_i^*(\omega)\}.$$

Then

$$\begin{aligned}
\mu(A_0) &\leq \sum_{i \in I} \mu(\{\omega \in \Omega: f_i(\omega) \neq f_i^*(\omega)\}) \quad [\text{Box 1.1 (xi)}] \\
&= 0, \quad \text{if } f_i \stackrel{\mu}{=} f_i^*, \forall i \in I.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mu\left(\left\{\omega \in \Omega: \sum_{i \in I} \alpha_i f_i(\omega) \neq \sum_{i \in I} \alpha_i f_i^*(\omega)\right\}\right) &\leq \mu(A_0) \quad [\text{Box 1.1 (v)}] \\
&= 0, \quad \text{if } f_i \stackrel{\mu}{=} f_i^*, \forall i \in I.
\end{aligned}$$

▷ **Solution 2-18**

$$\forall \omega \in \Omega: (f(\omega) < g(\omega) \wedge g(\omega) = h(\omega)) \Rightarrow f(\omega) < h(\omega),$$

which, by contraposition, is equivalent to

$$\forall \omega \in \Omega: f(\omega) \geq h(\omega) \Rightarrow (f(\omega) \geq g(\omega) \vee g(\omega) \neq h(\omega)).$$

Therefore,

$$\{\omega \in \Omega: f(\omega) \geq h(\omega)\} \subset \{\omega \in \Omega: f(\omega) \geq g(\omega)\} \cup \{\omega \in \Omega: g(\omega) \neq h(\omega)\}.$$

Now  $(f \stackrel{\mu}{<} g \wedge g \stackrel{\mu}{=} h)$  implies

$$\mu(\{\omega \in \Omega: f(\omega) \geq h(\omega)\})$$

$$\begin{aligned}
&\leq \mu(\{\omega \in \Omega: f(\omega) \geq g(\omega)\}) + \mu(\{\omega \in \Omega: g(\omega) \neq h(\omega)\}) && \text{[Box 1.1 (vii)]} \\
&= 0 + 0 = 0. && [f \stackrel{\mu}{\leq} g, g \stackrel{\mu}{=} h]
\end{aligned}$$

Because a measure is nonnegative, this implies  $\mu(\{\omega \in \Omega: f(\omega) \geq h(\omega)\}) = 0$ , which is equivalent to  $f \stackrel{\mu}{<} h$ .

▷ **Solution 2-19** For all  $A' \in \mathcal{A}'$ ,

$$\mu_f(A') = \mu[f^{-1}(A')] \quad [(2.41)]$$

$$= \mu \left( \bigcup_{\substack{i=1, \dots, n+1, \\ \alpha_i \in A'}} A_i \right) \quad \text{[Def. 2.16]}$$

$$= \sum_{\substack{i=1, \dots, n+1, \\ \alpha_i \in A'}} \mu(A_i) \quad \text{[Def. 1.43, (c)]}$$

$$= \sum_{i=1}^{n+1} \mu(A_i) \cdot \delta_{\alpha_i}(A'). \quad [(1.39)]$$

## Chapter 3

# Integral

In the preceding chapters we introduced the most important concepts of measure theory related to the concepts of a measure and a measurable mapping. In this chapter we introduce the *integral* of measurable functions. This concept is fundamental also for probability theory, because the expectation of a numerical random variable with respect to a probability measure is the *integral* of a measurable function with respect to a probability measure. In chapter 6 we shall see that this also applies to variances, covariances, and correlations. We start defining the integral of a measurable function with respect to a measure  $\mu$ . Then we study the most important rules of computation and other properties of integrals, introduce the concept of a *measure with density*, and treat the relationship between the Riemann integral and the integral with respect to the Lebesgue measure. The next section is on *absolute continuity* and the *Radon-Nikodym Theorem*. Both issues are crucial for conditional expectations (see ch. 10). A section on the integral with respect to a product measure concludes this chapter.

### 3.1 Definition

At first we define the integral for *nonnegative step functions*, then we extend the integral to *nonnegative measurable functions*, and finally we introduce the integral for *measurable functions* that may take on negative or nonnegative values.

#### 3.1.1 Integral of a Nonnegative Step Function

In this subsection we introduce the integral of a *nonnegative step function*, also called *nonnegative simple function* or *elementary function*.

#### Nonnegative Step Function

**Definition 3.1 (Nonnegative Step Function and Normal Representation)**

Let  $(\Omega, \mathcal{A})$  be a measurable space. Then  $f: \Omega \rightarrow \mathbb{R}$  is called a *nonnegative step function*, if there is a finite sequence  $A_1, \dots, A_n \in \mathcal{A}$  and a finite sequence  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, n$ , such that

$$f = \sum_{i=1}^n \alpha_i 1_{A_i}. \quad (3.1)$$

If  $A_1, \dots, A_n \in \mathcal{A}$  are pairwise disjoint, then  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  is called a normal representation of  $f$ .

**Remark 3.2 (Step Functions Take on Finitely Many Values)** A nonnegative step function  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  is a measurable function  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  taking on only a finite number of *nonnegative* values. These values are not necessarily  $\alpha_1, \dots, \alpha_n$ . However, note:

- (i) If  $\mathcal{E} = \{A_1, \dots, A_n\}$  is a partition of  $\Omega$ , then  $\alpha_1, \dots, \alpha_n$  are the values of  $f$ .
- (ii) If  $A_1, \dots, A_n$  are pairwise disjoint but  $\mathcal{E}$  is not a partition of  $\Omega$ , i. e.,  $A_{n+1} := \Omega \setminus \bigcup_{i=1}^n A_i \neq \emptyset$ , then

$$f = \sum_{i=1}^n \alpha_i 1_{A_i} + 0 \cdot 1_{A_{n+1}}.$$

This implies:  $f(\omega) = 0$ , for all  $\omega \in A_{n+1}$ .

- (iii) If  $A_1, \dots, A_n$  are pairwise disjoint and additionally  $\alpha_1, \dots, \alpha_n$  are pairwise different and not 0, then  $A_i = f^{-1}(\{\alpha_i\})$ ,  $i = 1, \dots, n$ .
- (iv) If  $A_1, \dots, A_n$  are pairwise disjoint, then, for all  $\alpha_i \neq 0$ ,

$$f^{-1}(\{\alpha_i\}) = \bigcup_{j: \alpha_j = \alpha_i} A_j.$$

Hence in this case, the inverse image of the set  $\{\alpha_i\}$  under  $f$  is the union of all sets  $A_j$ ,  $j \in \{1, \dots, n\}$ , for which  $\alpha_j = \alpha_i$ .

◁

**Remark 3.3 (Different Representations of a Nonnegative Step Function)** Note that nonnegative step functions can have different representations and also different normal representations (see Example 3.7). ◁

**Example 3.4 (Indicator Function)** Let  $(\Omega, \mathcal{A})$  be a measurable space and  $A \in \mathcal{A}$ . The indicator function  $1_A$ , which has already been introduced in Example 2.12, can also be written  $1 \cdot 1_A + 0 \cdot 1_{A^c}$ . Hence, because  $A \in \mathcal{A}$  and 1 is a real number,  $1_A$  is a nonnegative step function. Note that  $1 \cdot 1_A$  is also a normal representation of a nonnegative step function. ◁

**Example 3.5 (Two Nonnegative Step Functions)** In Example 2.55 we already presented two nonnegative step functions  $f$  and  $h$  and an illustrating Figure (see Fig. 2.6). The representations of both functions are normal. ◁

**Example 3.6 (Tossing a Dice)** Consider the set  $\Omega = \{\omega_1, \dots, \omega_6\}$  of possible outcomes of tossing a dice. Furthermore, let  $\mathcal{A} = \mathcal{P}(\Omega)$  be the power set of  $\Omega$ , and define  $X: \Omega \rightarrow \mathbb{R}$  by

$$X(\omega_i) = i, \quad \forall \omega_i \in \Omega.$$

Hence,  $X(\omega_i)$  is the *number of dots*. Considering the elements  $\{\omega_1\}, \dots, \{\omega_6\}$  of  $\mathcal{A}$ , and

$$X = \sum_{i=1}^6 i \cdot \mathbf{1}_{\{\omega_i\}}$$

shows that  $X$  has a normal representation of a nonnegative step function. (For a related example see Exercise 3-1.)  $\triangleleft$

**Example 3.7 (Several Representations of Nonnegative Step Functions)** Consider the measurable space  $(\mathbb{R}, \mathcal{B})$  and the nonnegative function  $f: \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$f(x) = \begin{cases} 2, & \text{if } x \in [0, 1[ \\ 5, & \text{if } x \in [1, 2] \\ 4, & \text{if } x \in ]2, 3] \\ 1, & \text{if } x \in ]3, 4] \\ 0, & \text{otherwise.} \end{cases}$$

This function can also be represented by

$$\begin{aligned} f &= 2 \cdot \mathbf{1}_{[0,1[} + 5 \cdot \mathbf{1}_{[1,2]} + 4 \cdot \mathbf{1}_{]2,3]} + 1 \cdot \mathbf{1}_{]3,4]} \\ &= 2 \cdot \mathbf{1}_{[0,.5]} + 2 \cdot \mathbf{1}_{].5,1]} + 5 \cdot \mathbf{1}_{[1,2]} + 4 \cdot \mathbf{1}_{]2,3]} + 1 \cdot \mathbf{1}_{]3,4]} \\ &= 2 \cdot \mathbf{1}_{[0,2]} + 3 \cdot \mathbf{1}_{[1,3]} + 1 \cdot \mathbf{1}_{[2,4]} \\ &= 1 \cdot \mathbf{1}_{[0,4]} + 1 \cdot \mathbf{1}_{[0,3]} + 2 \cdot \mathbf{1}_{[1,3]} + 1 \cdot \mathbf{1}_{[1,2]}. \end{aligned} \quad (3.2)$$

The first two representations are normal, the latter two are nonnormal representations of  $f$ .  $\triangleleft$

**Remark 3.8 (Existence of a Normal Representation)** For every nonnegative step function there exists a normal representation (see Exercise 3-2).

If  $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$  is a normal representation of a nonnegative step function, then there may be another sequence  $C_1, \dots, C_m$  of pairwise disjoint elements of  $\mathcal{A}$  and another sequence  $\gamma_1, \dots, \gamma_m$  of nonnegative real numbers such that

$$f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} = \sum_{i=1}^m \gamma_i \mathbf{1}_{C_i}.$$

Both sum terms are normal representations. The first two representations of  $f$  in Equation (3.2) provide an example.  $\triangleleft$

### Integral of a Nonnegative Step Function

The following uniqueness property holds for two normal representations of a nonnegative step function:

**Lemma 3.9 (A Uniqueness Property)**

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space. If  $f: \Omega \rightarrow \mathbb{R}$  is a nonnegative step function and  $f = \sum_{i=1}^n \alpha_i 1_{A_i} = \sum_{i=1}^m \gamma_i 1_{C_i}$  are two normal representations, then

$$\sum_{i=1}^n \alpha_i \mu(A_i) = \sum_{i=1}^m \gamma_i \mu(C_i). \quad (3.3)$$

For a proof see Klenke (2013, Lemma 4.1, p. 85). Note, by convention  $0 \cdot \infty = 0$ .

According to this lemma, the number  $\sum_{i=1}^n \alpha_i \mu(A_i)$  assigned to a nonnegative step function  $f$  does not depend on the specific normal representation of  $f$  (for an illustration see Exercise 3-3). This property allows us to define the *integral* of a nonnegative step function with respect to a measure  $\mu$  as follows:

**Definition 3.10 (Integral of a Nonnegative Step Function)**

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space and let  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  be a normal representation of a nonnegative step function  $f: \Omega \rightarrow \mathbb{R}$ . Then the number

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i) \quad (3.4)$$

is called the *integral of  $f$  (over  $\Omega$ ) with respect to  $\mu$* .

**Remark 3.11 (Integral of a Constant)** Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space. If  $f = \alpha$ ,  $\alpha \in \mathbb{R}$ , then Equation (3.4) immediately implies

$$\int \alpha d\mu = \alpha \cdot \mu(\Omega). \quad (3.5)$$

◁

**Remark 3.12 (Integral Over a Subset of  $\Omega$ )** Let  $A \in \mathcal{A}$ . If  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  is a normal representation of a nonnegative step function, then the product  $1_A \cdot f$  is a nonnegative step function as well and can be written

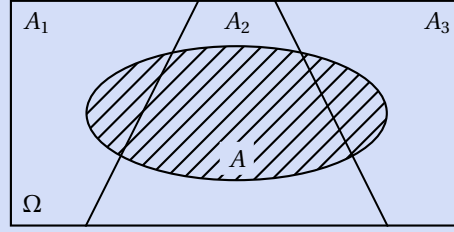
$$1_A \cdot f = \sum_{i=1}^n \alpha_i 1_{A \cap A_i}, \quad (3.6)$$

which is a normal representation of  $1_A \cdot f$  (see Fig. 3.1 and Exercise 3-4). Hence, we may also consider the integral  $\int 1_A \cdot f d\mu$  and define the *integral of  $f$  over a subset  $A$  of  $\Omega$*  by

$$\int_A f d\mu = \int 1_A \cdot f d\mu. \quad (3.7)$$

◁

The following corollary is a special case of Equation (3.7) if  $f = \alpha$ ,  $\alpha \in \mathbb{R}$ . (For proof see Exercise 3-5).



**Figure 3.1.** A partition and a subset of  $\Omega$

**Corollary 3.13 (Constants)**

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space and  $\alpha \in \mathbb{R}$ . If  $A \in \mathcal{A}$ , then

$$\int_A \alpha d\mu = \alpha \mu(A). \quad (3.8)$$

**Examples**

**Example 3.14 (Indicator Function)** Consider a measure space  $(\Omega, \mathcal{A}, \mu)$  and the indicator  $1_A$  of  $A \in \mathcal{A}$ . Then  $1_A = 1 \cdot 1_A$  is a normal representation of  $1_A$ . Therefore,

$$\int_A d\mu = \int 1_A d\mu = 1 \cdot \mu(A) = \mu(A). \quad (3.9)$$

◁

**Example 3.15 (Nonnegative Step Function and Dirac Measure)** Let  $(\Omega, \mathcal{A})$  be a measurable space, and for  $\omega \in \Omega$  let  $\delta_\omega$  denote the *Dirac measure* at  $\omega$  (see Example 1.52). Furthermore, consider a normal representation  $f = \sum_{i=1}^m \alpha_i 1_{A_i}$  of a nonnegative step function. Its integral with respect to the Dirac measure is

$$\int f d\delta_\omega = \sum_{i=1}^m \alpha_i \delta_\omega(A_i) = \sum_{i=1}^m \alpha_i 1_{A_i}(\omega) = f(\omega). \quad (3.10)$$

According to this equation, the integral of a nonnegative step function  $f$  with respect to the Dirac measure at  $\omega$  is the value of  $f$  for the argument  $\omega$ . Furthermore, if  $f = 1_A$  is the indicator of  $A \in \mathcal{A}$ , then

$$\int 1_A d\delta_\omega = 1_A(\omega). \quad (3.11)$$

Hence, in this special case, the integral is the value of the indicator  $1_A$  for the argument  $\omega$ . ◁

**Example 3.16 (Nonnegative Step Function and Counting Measure)** Suppose  $\Omega = \{1, \dots, n\}$ ,  $n \in \mathbb{N}$ . For the measurable space  $(\Omega, \mathcal{P}(\Omega))$ , the counting measure  $\mu_{\#}$  on the power set  $\mathcal{P}(\Omega)$  is defined by

$$\mu_{\#}(A) = \sum_{\omega=1}^n 1_A(\omega), \quad \forall A \subset \Omega, \quad (3.12)$$

(see Example 1.54). Hence,  $\mu_{\#}(A)$  is simply the number of elements, i. e., the cardinality of  $A$ . Now consider a nonnegative step function with normal representation  $f = \sum_{i=1}^m \alpha_i 1_{A_i}$ . According to Equations (3.4) and (3.12), its integral with respect to the counting measure is

$$\begin{aligned} \int f d\mu_{\#} &= \sum_{i=1}^m \alpha_i \mu_{\#}(A_i) = \sum_{i=1}^m \alpha_i \sum_{\omega=1}^n 1_{A_i}(\omega) = \sum_{\omega=1}^n \sum_{i=1}^m \alpha_i 1_{A_i}(\omega) \\ &= \sum_{\omega=1}^n f(\omega). \end{aligned} \quad (3.13)$$

Hence, the integral of a nonnegative step function  $f$  with respect to the counting measure  $\mu_{\#}$  is the sum over all values of  $f$  (see also Exercise 3-6). Using Equations (1.41) and (3.10), this integral can also be written

$$\int f d\mu_{\#} = \int f d\left(\sum_{\omega=1}^n \delta_{\omega}\right) = \sum_{\omega=1}^n \int f d\delta_{\omega}. \quad (3.14)$$

&lt;

### 3.1.2 Integral of a Nonnegative Measurable Function

In this section we extend the concept of an integral to nonnegative measurable functions. Before introducing the definition we consider a theorem according to which every nonnegative measurable function can be represented as a limit of an increasing sequence of nonnegative step functions. We begin with an example.

**Example 3.17 (Increasing Sequence of Nonnegative Step Functions)** Consider the measurable space  $(\mathbb{R}, \mathcal{B})$  and the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$f(x) = \begin{cases} 1 - x^2, & \forall x \in [0, 1], \\ 0, & \text{otherwise.} \end{cases} \quad (3.15)$$

Now we construct three functions  $f_i: \mathbb{R} \rightarrow \mathbb{R}$ ,  $i = 1, 2, 3$ , with  $f_1 \leq f_2 \leq f_3 \leq f$  that approximate  $f$  (see Fig. 3.2). Let us start with  $f_1$  defined by

$$f_1(x) = \begin{cases} \alpha_1 = .50, & \text{if } x \in A_1 = [0, (1 - .50)^{1/2}], \\ \alpha_2 = 0, & \text{if } x \in A_1^c, \end{cases}$$

where  $[0, (1 - .50)^{1/2}]$  denotes the closed interval between 0 and  $(1 - .50)^{1/2} \approx .707$ . Because  $A_1$  is an element of  $\mathcal{B}$  and .50 is a nonnegative real number,  $f_1 = \alpha_1 1_{A_1}$  is a nonnegative step function. Next consider  $f_2$  defined by

$$f_2(x) = \begin{cases} \beta_1 = .75, & \text{if } x \in B_1 = [0, (1 - .75)^{1/2}], \\ \beta_2 = .50, & \text{if } x \in B_2 = ](1 - .75)^{1/2}, (1 - .50)^{1/2}], \\ \beta_3 = .25, & \text{if } x \in B_3 = ](1 - .50)^{1/2}, (1 - .25)^{1/2}], \\ \beta_4 = 0, & \text{if } x \in (B_1 \cup B_2 \cup B_3)^c. \end{cases}$$

Because  $B_1, B_2, B_3$  are elements of  $\mathcal{B}$  and  $.75, .50, .25$  are nonnegative real numbers,  $f_2 = \sum_{i=1}^3 \beta_i 1_{B_i}$  is a nonnegative step function. Finally, let  $f_3$  be defined by

$$f_3(x) = \begin{cases} \gamma_1 = .875, & \text{if } x \in C_1 = [0, (1 - .875)^{1/2}], \\ \gamma_2 = .750, & \text{if } x \in C_2 = ](1 - .875)^{1/2}, (1 - .750)^{1/2}], \\ \gamma_3 = .625, & \text{if } x \in C_3 = ](1 - .750)^{1/2}, (1 - .625)^{1/2}], \\ \gamma_4 = .500, & \text{if } x \in C_4 = ](1 - .625)^{1/2}, (1 - .500)^{1/2}], \\ \gamma_5 = .375, & \text{if } x \in C_5 = ](1 - .500)^{1/2}, (1 - .375)^{1/2}], \\ \gamma_6 = .250, & \text{if } x \in C_6 = ](1 - .375)^{1/2}, (1 - .250)^{1/2}], \\ \gamma_7 = .125, & \text{if } x \in C_7 = ](1 - .250)^{1/2}, (1 - .125)^{1/2}], \\ \gamma_8 = 0, & \text{if } x \in (C_1 \cup \dots \cup C_7)^c. \end{cases}$$

Again,  $C_1, \dots, C_7$  is a sequence of elements of  $\mathcal{B}$  and  $.875, .750, .625, .500, .375, .250, .125$  is a sequence of nonnegative real numbers. Therefore,  $f_3 = \sum_{i=1}^7 \gamma_i 1_{C_i}$  is a nonnegative step function. The integral of the functions  $f_1$  and  $f_2$  are computed in Exercise 3-7.  $\triangleleft$

### Convergence of an Increasing Sequence of Nonnegative Step Functions

**Example 3.18 (Convergence)** Figure 3.2 shows that  $f_1(\omega) \leq f_2(\omega) \leq f_3(\omega) \leq f(\omega)$  for all  $\omega \in \Omega$ . Hence,  $f_1, f_2, f_3$  is a *finite increasing sequence* of nonnegative step functions. The interval  $[0, 1]$  on the vertical axis is partitioned and these partitions are refined step by step. In our example, we started with the partition  $\{[0, .50[, [.50, 1]\}$ . Then we partitioned

$$[0, .50[ \text{ to } \{[0, .25[, [.25, .50]\} \quad \text{and} \quad [.50, 1] \text{ to } \{[.50, .75[, [.75, 1]\}, \text{ etc.}$$

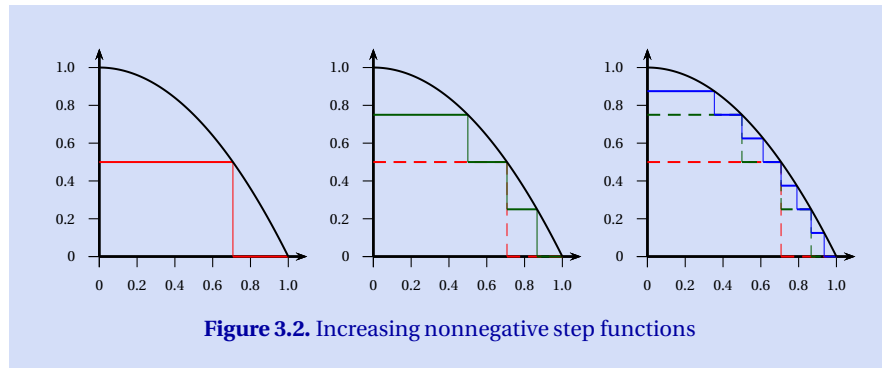
Following this idea, we can define functions  $f_4, f_5, \dots$  such that  $f_1, f_2, \dots$  is an infinite sequence of nonnegative step functions with  $f_1(\omega) \leq f_2(\omega) \leq \dots \leq f(\omega)$ , for all  $\omega \in \Omega$ , and  $\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega)$ , for all  $\omega \in \Omega$ . According to the following theorem, this holds for *all* nonnegative measurable functions  $f: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ .  $\triangleleft$

#### Theorem 3.19 (Approximation of Nonnegative Functions)

Let  $(\Omega, \mathcal{A})$  be a measurable space and  $f: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  a nonnegative measurable function. Then:

(i) There is a sequence  $f_1, f_2, \dots$  of nonnegative step functions such that

$$f_1(\omega) \leq f_2(\omega) \leq \dots, \quad \forall \omega \in \Omega \quad (3.16)$$



and

$$\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega), \quad \forall \omega \in \Omega. \quad (3.17)$$

(ii) There is a sequence of sets  $A_1, A_2, \dots \in \mathcal{A}$  and a sequence of nonnegative real numbers  $\alpha_1, \alpha_2, \dots$  such that

$$f = \sum_{i=1}^{\infty} \alpha_i 1_{A_i}. \quad (3.18)$$

For a proof see Klenke (2013, Theorem 1.96, p. 38).

**Remark 3.20 (Infinite Sums)** Equation (3.18) can be visualized by Figure 3.2. The function  $f_3$  on the right-hand side of this figure is already close to  $f$ . Partitioning the intervals *on the vertical axis* again and again leads to better approximations of  $f$ . Note that the horizontal axis does not have to be a subset of  $\mathbb{R}$ ; instead, it can be any nonempty set  $\Omega$ .

Remember, the right-hand side of Equation (3.18) is just a symbol for the corresponding limit, i. e.,

$$\sum_{i=1}^{\infty} \alpha_i 1_{A_i} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \alpha_i 1_{A_i}. \quad (3.19)$$

Note that, for  $\alpha_i \geq 0$ , this limit always exists. ◁

Before turning to the definition of the integral of a nonnegative measurable function let us use the properties (3.16) and (3.17) to define the concepts *increasing sequence of nonnegative step functions* and *pointwise convergence*.

**Definition 3.21 (Increasing Sequence of Nonnegative Step Functions)**

Let  $(\Omega, \mathcal{A})$  be a measurable space and  $f: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  a nonnegative measurable function. A sequence  $f_1, f_2, \dots$  of nonnegative step functions satisfying

(3.16) is called *increasing*. If it also satisfies (3.17), then we say that  $f_1, f_2, \dots$  converges pointwise to  $f$  and denote it by  $f_n \uparrow f$ .

### Uniqueness of the Limits of an Integral

In Theorem 3.19 we have seen that every nonnegative measurable function  $f$  can be represented by the limit  $\lim_{n \rightarrow \infty} f_n$  of an increasing sequence  $f_1, f_2, \dots$  of nonnegative step functions, i. e.,

$$f = \lim_{n \rightarrow \infty} f_n. \quad (3.20)$$

The definition of the integral of nonnegative step functions implies that the *integrals* of the functions  $f_n$  are increasing as well, i. e.,

$$f_n \leq f_{n+1} \Rightarrow \int f_n d\mu \leq \int f_{n+1} d\mu, \quad \forall n \in \mathbb{N},$$

[see Bauer, 2001, proposition (10.7), p. 55]. Hence, the sequence of the integrals either converges to a (finite) real number or diverges to  $+\infty$ .

In Figure 3.2 we presented the first three nonnegative step functions  $f_1, f_2$ , and  $f_3$  of such an increasing sequence  $f_1, f_2, \dots$  that approximates the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  defined by Equation (3.15). Figure 3.3 visualizes the convergence of the integrals  $\int f_n d\lambda$  with respect to the Lebesgue measure  $\lambda$  on  $(\mathbb{R}, \mathcal{B})$  (see the shaded areas in Figure 3.3).

It should be noted, however, that there is not only one single increasing sequence of nonnegative step functions whose limit is  $f$ . This is illustrated in the following example.

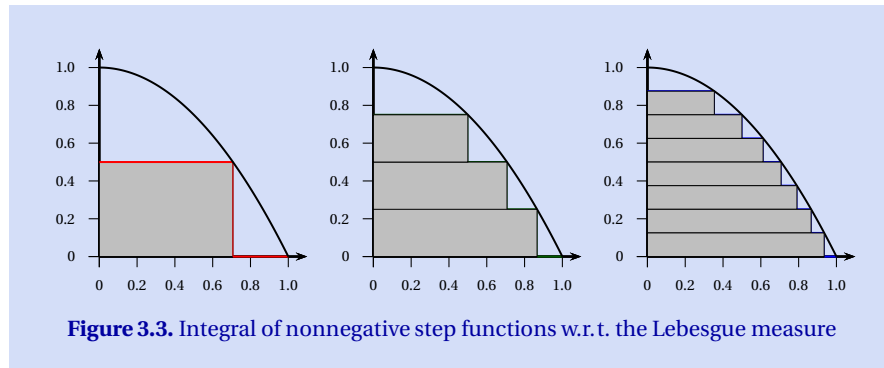
**Example 3.22 (Uniqueness)** For an example, construct a sequence  $g_1, g_2, \dots$  analogously to the sequence  $f_1, f_2, \dots$  in Example 3.17, using other partitions of the interval  $[0, 1]$  on the vertical axis, e. g.,

$$\{[0, .40[, [.40, 1]\} \quad \text{and} \quad \{[0, .20[, [.20, .40[, [.40, .80[, [.80, 1]\}, \quad \text{etc.}$$

Then  $g_1, g_2, \dots$  is a second increasing sequence that also approximates  $f$ . Figure 3.3 suggests that the specific choice of an increasing sequence of nonnegative step functions is irrelevant for the limit of their integrals. And in fact, according to the following theorem this does not only apply to our example and to the Lebesgue measure  $\lambda$  on  $\mathcal{B}$ , but to *any* nonnegative measurable function and *any* measure  $\mu$ . ◁

### Theorem 3.23 (Uniqueness of the Limits of Integrals)

If  $f_1, f_2, \dots$  and  $g_1, g_2, \dots$  are two increasing sequences of nonnegative step functions  $f_n, g_n: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ , then  $\lim_{n \rightarrow \infty} f_n = \lim_{n \rightarrow \infty} g_n$  implies



$$\lim_{n \rightarrow \infty} \int f_n d\mu = \lim_{n \rightarrow \infty} \int g_n d\mu. \quad (3.21)$$

For a proof see Bauer (2001, Corollary 11.2, p. 58).

According to this theorem, if we consider two increasing sequences of nonnegative step functions with identical limits, then we know that the limits of their integrals are identical.

### Definition of the Integral of a Nonnegative Measurable Function

Based on the result of Theorem 3.23, we define the integral of any nonnegative measurable function.

#### Definition 3.24 (Integral of a Nonnegative Measurable Function)

Assume that  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is a nonnegative measurable function and let  $f = \lim_{n \rightarrow \infty} f_n$  be a representation of  $f$  as the limit of an increasing sequence  $f_1, f_2, \dots$  of nonnegative step functions. Then

$$\int f d\mu := \lim_{n \rightarrow \infty} \int f_n d\mu \quad (3.22)$$

is called the *integral of  $f$  (over  $\Omega$ ) with respect to  $\mu$* .

Note that the integral of a nonnegative measurable function is either a nonnegative real number or  $+\infty$ .

**Example 3.25 (Integral With Respect to a Dirac Measure)** Suppose the assumptions of Definition 3.24 hold. Then, for  $\omega \in \Omega$ ,

$$\begin{aligned} \int f d\delta_\omega &= \lim_{n \rightarrow \infty} \int f_n d\delta_\omega && [(3.22)] \\ &= \lim_{n \rightarrow \infty} f_n(\omega) && [(3.10)] \quad (3.23) \\ &= f(\omega). && [(3.20)] \end{aligned}$$

Hence, the integral of a nonnegative measurable function  $f$  with respect to the Dirac measure at  $\omega$  is the value of  $f$  for  $\omega$ .  $\triangleleft$

We conclude this section by the following lemma on monotonicity of the integrals of nonnegative measurable functions.

**Lemma 3.26 (Monotonicity)**

If  $f, g: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  are nonnegative and measurable, then

$$f \leq g \Rightarrow \int f d\mu \leq \int g d\mu. \quad (3.24)$$

For a proof see Bauer (2001, p. 59, Eq. (11.8)).

**Remark 3.27 (Bounds of the Integral of a Bounded Function)** Let  $f: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be nonnegative and measurable and  $\alpha \in \mathbb{R}$ . Then, for  $g = \alpha$ , Equations (3.24) and (3.5) imply

$$f \leq \alpha \Rightarrow \int f d\mu \leq \alpha \cdot \mu(\Omega), \quad (3.25)$$

and

$$f \geq \alpha \Rightarrow \int f d\mu \geq \alpha \cdot \mu(\Omega). \quad (3.26)$$

$\triangleleft$

### 3.1.3 Integral of a Measurable Function

Now we define the integral of a measurable function  $f: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  using the positive part  $f^+$  and the negative part  $f^-$  of  $f$  that have been introduced in section 2.4.2. According to Theorem 2.66,  $f^+$  and  $f^-$  are both nonnegative measurable functions. Reading the following definition, remember the conventions:  $\infty + \infty = \infty$ ,  $-\infty - \infty = -\infty$ ,  $x + \infty = \infty$ , for all  $x \in \mathbb{R}$ ,  $x - \infty = -\infty$ , for all  $x \in \mathbb{R}$ . Also note that  $\infty - \infty$  is *not defined*, which has to be observed whenever integrals are not necessarily finite.

**Definition 3.28 (Integral of a Measurable Function)**

Let  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a measurable function. If  $\int f^+ d\mu$  or  $\int f^- d\mu$

are finite, then  $f$  is called *quasi-integrable with respect to  $\mu$* , or simply *quasi- $\mu$ -integrable*, and

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu \quad (3.27)$$

is called the *integral of  $f$  (over  $\Omega$ ) with respect to  $\mu$* . If  $\int f^+ \, d\mu$  and  $\int f^- \, d\mu$  are both finite, then  $f$  is called *integrable with respect to  $\mu$* , or simply  *$\mu$ -integrable*.

**Remark 3.29 (Integrability and Quasi-Integrability)** Of course, every integrable measurable function is quasi-integrable and each nonnegative function is also quasi-integrable. Furthermore, assuming that a function  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is integrable or quasi-integrable includes the assumption that  $f$  is measurable. Finally, if  $f$  is  $\mu$ -integrable, then

$$-\infty < \int f \, d\mu < +\infty,$$

i. e., the integral is *finite*, taking a value in  $\mathbb{R}$ . If  $f$  is quasi- $\mu$ -integrable, then the integral may also be *infinite*, i. e., it may also take on the values  $+\infty$  or  $-\infty$ .  $\triangleleft$

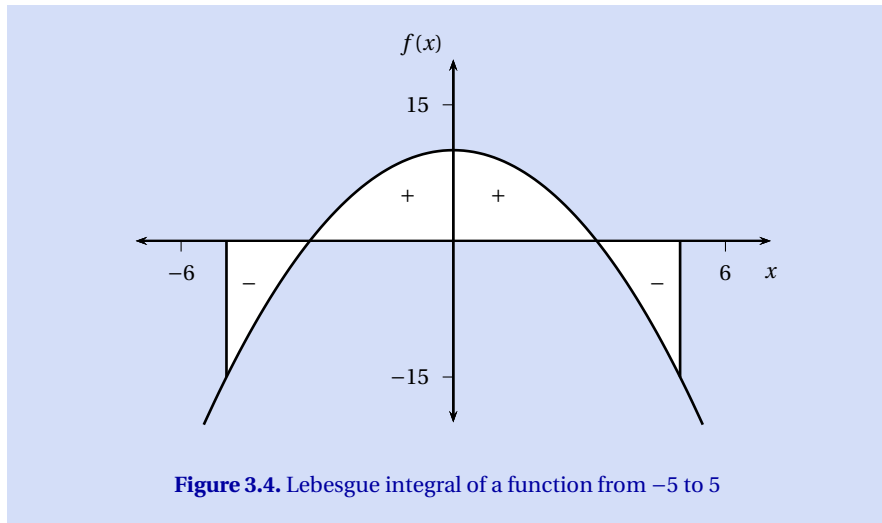
**Remark 3.30 (A Standard Method for Proofs)** The integral of a quasi-integrable function has been defined in three steps, for *nonnegative measurable step functions*, for *nonnegative measurable functions*, and for *quasi-integrable functions*. Oftentimes, these steps are also followed in proofs of propositions involving integrals. That is, in a first step it is shown that the proposition holds for nonnegative measurable step functions. In a second step, using Equation (3.22), it is proven for nonnegative measurable functions, and finally, Equation (3.27) is applied to complete the proof for all quasi-integrable functions. An example is the proof of Theorem 3.36. Oftentimes, we only detail the first step, in particular, if the remaining two steps are straightforward.  $\triangleleft$

**Example 3.31 (Integral With Respect to the Lebesgue Measure  $\lambda$ )** Figure 3.4 displays the integral of a function  $f: (\mathbb{R}, \mathcal{B}, \lambda) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  with respect to the Lebesgue measure  $\lambda$ . Because  $f^+$  and  $f^-$  are both nonnegative (see Rem. 2.62), the integrals  $\int f^+ \, d\mu$  and  $\int f^- \, d\mu$  are positive and identical to the white areas marked + and – in Figure 3.4. According to Equation (3.27), the integral of  $f$  is the difference between the area  $\int f^+ \, d\mu$  and the area  $\int f^- \, d\mu$ .  $\triangleleft$

**Remark 3.32 (An Alternative Notation)** An alternative notation for the integral of  $f$  is

$$\int f \, d\mu = \int f(\omega) \, \mu(d\omega) = \int_{\Omega} f(\omega) \, \mu(d\omega), \quad (3.28)$$

which explicitly uses the values  $f(\omega)$  of  $f$ . This notation conveys the idea that the values  $f(\omega)$  of  $f$  are weighted by the measure of  $d\omega$ . If  $\Omega = \mathbb{R}$ , then  $d\omega$  symbolizes



the length of an infinitesimal interval between two elements in  $\mathbb{R}$ . If  $\Omega$  is finite or countable, then  $\mu(d\omega)$  symbolizes the value of  $\mu$  for the singleton  $\{\omega\}$  and the integral can be written as a sum (see Example 3.16).  $\triangleleft$

**Lemma 3.33 (Integrability Carries Over to Restrictions of Functions)**

- (i) If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is quasi- $\mu$ -integrable and  $A \in \mathcal{A}$ , then  $1_A f$  is quasi- $\mu$ -integrable.
- (ii) If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is  $\mu$ -integrable and  $A \in \mathcal{A}$ , then  $1_A f$  is  $\mu$ -integrable.

*(Proof p. 111)*

**Remark 3.34 (Integral of  $1_A f$ )** Lemma 3.33 (ii) means: If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is  $\mu$ -integrable and  $A \in \mathcal{A}$ , then

$$\int f d\mu \text{ is finite} \quad \Rightarrow \quad \int 1_A f d\mu \text{ is finite.} \quad (3.29)$$

$\triangleleft$

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is quasi- $\mu$ -integrable and  $A \in \mathcal{A}$ , then Lemma 3.33 implies that the integral  $\int 1_A f d\mu$  is well-defined. Hence, we can now introduce the integral of  $f$  over a subset  $A$  of  $\Omega$  as follows:

**Definition 3.35 (Integral Over a Subset  $A$  of  $\Omega$ )**

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is quasi- $\mu$ -integrable and  $A \in \mathcal{A}$ , then

$$\int_A f \, d\mu := \int 1_A f \, d\mu \quad (3.30)$$

is called the integral of  $f$  over  $A$  with respect to  $\mu$ .

Because  $1_\Omega f = f$ , a special case of Equation (3.30) is

$$\int_\Omega f \, d\mu = \int 1_\Omega f \, d\mu = \int f \, d\mu. \quad (3.31)$$

**3.2 Properties**

In this section we consider some important properties and rules of computation for the integral of a measurable function  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ .

**Theorem 3.36 (Linearity)**

Consider the functions  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  and  $g: (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$ .

(i) If  $f$  is quasi- $\mu$ -integrable and  $\alpha \in \mathbb{R}$ , then  $\alpha f$  is quasi- $\mu$ -integrable and

$$\int \alpha f \, d\mu = \alpha \int f \, d\mu. \quad (3.32)$$

(ii) If  $f$  is quasi- $\mu$ -integrable and  $g$  is  $\mu$ -integrable, then  $f + g$  is quasi- $\mu$ -integrable, and

$$\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu. \quad (3.33)$$

(Proof p. 111)

Combining propositions (i) and (ii) of Theorem 3.36 immediately yields the following corollary.

**Corollary 3.37 (Linearity)**

Consider the functions  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ ,  $g: (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$  and let  $\alpha, \beta \in \mathbb{R}$ . If  $f$  is quasi- $\mu$ -integrable and  $g$  is  $\mu$ -integrable, then  $\alpha f + \beta g$  is quasi- $\mu$ -integrable, and

$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu. \quad (3.34)$$

Linearity can also be used to prove the following corollary on the equivalence of integrability of a measurable function  $f$  and finiteness of the integral of the absolute value function  $|f|$ .

**Corollary 3.38 (Integrability and Absolute Value Function)**

The function  $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is  $\mu$ -integrable if and only if

$$\int |f| d\mu < \infty.$$

(Proof p. 115)

**Example 3.39 (Integral Over the Union of Two Sets)** If  $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is  $\mu$ -integrable and  $A, B \in \mathcal{A}$ , then  $1_{A \cup B} f$  is  $\mu$ -integrable and

$$\int_{A \cup B} f d\mu = \int 1_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu - \int_{A \cap B} f d\mu. \quad (3.35)$$

If  $A \cap B = \emptyset$  and  $f$  is quasi- $\mu$ -integrable, then

$$\int_{A \cup B} f d\mu = \int 1_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu. \quad (3.36)$$

(see Exercise 3-8).

◁

**Lemma 3.40 (Measures That are Identical on a Sub- $\sigma$ -Algebra)**

Assume that  $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is nonnegative or  $\mu$ -integrable. Furthermore, let  $\mathcal{C} \subset \mathcal{A}$  be a  $\sigma$ -algebra, let  $f$  be  $\mathcal{C}$ -measurable, and suppose that  $\nu(A) = \mu(A)$ , for all  $A \in \mathcal{C}$ . Then  $(\Omega, \mathcal{C}, \nu)$  is a measure space and

$$\int f d\nu = \int f d\mu. \quad (3.37)$$

(Proof p. 115)

Hence, the integral  $\int f d\mu$  only depends on the values of  $\mu$  on the  $\sigma$ -algebra  $\sigma(f)$ , the  $\sigma$ -algebra generated by  $f$ .

**Lemma 3.41 (Integrable Functions are  $\mu$ -a.e. Real-Valued)**

Let  $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be measurable. If  $f$  is  $\mu$ -integrable, then  $f$  is real-valued  $\mu$ -almost everywhere. If  $f$  is quasi- $\mu$ -integrable, then

$$\mu(\{\omega \in \Omega : f(\omega) = \infty\}) > 0 \Rightarrow \int f d\mu = \infty, \quad (3.38)$$

$$\mu(\{\omega \in \Omega : f(\omega) = -\infty\}) > 0 \Rightarrow \int f d\mu = -\infty. \quad (3.39)$$

(Proof p. 115)

**Remark 3.42 (Integrable Functions are Assumed to be Real-Valued)** Contraposition of (3.38) and (3.39) yields: If  $\int f d\mu$  is finite, then  $f(\omega) \in \mathbb{R}$  (i. e.,  $-\infty < f(\omega) < \infty$ ), for  $\mu$ -almost all  $\omega \in \Omega$  (see Def. 2.68 and Remark 2.70). In this case, there is a *real-valued* measurable function  $f^*: (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$  with  $f^* \stackrel{\mu}{=} f$ . (For example, define  $f^* := 1_A \cdot f + 1_{A^c} \cdot 0 = 1_A \cdot f$  for  $A := \{\omega \in \Omega: f(\omega) \in \mathbb{R}\}$ ). Therefore, without substantial loss of generality, for simplicity, we often assume that a function is real-valued if it has a finite integral.  $\triangleleft$

### 3.2.1 Integral of $\mu$ -Equivalent Functions

The concept of equivalence of two measurable functions with respect to a measure has already been introduced in section 2.5. Now we treat the relationship of this concept to the integrals of two numerical functions.

**Theorem 3.43 (A Condition Equivalent to  $f \stackrel{\mu}{=} 0$ )**

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is a nonnegative measurable function, then

$$\int f d\mu = 0 \Leftrightarrow f \stackrel{\mu}{=} 0. \quad (3.40)$$

For a proof see Bauer (2001, Theorem 13.2, p. 71).

**Lemma 3.44 (Integral of a Positive Function)**

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is quasi- $\mu$ -integrable and there is an  $A \in \mathcal{A}$  such that  $\mu(A) > 0$  and  $f(\omega) > 0$ , for all  $\omega \in A$ , then

$$\int 1_A \cdot f d\mu > 0. \quad (3.41)$$

(Proof p. 116)

If  $(\Omega, \mathcal{A}, \mu)$  is a measure space, then a set  $A \in \mathcal{A}$  with  $\mu(A) = 0$  is called a *null set* with respect to  $\mu$ . In the following lemma we consider the integral over such a null set (see Exercise 3-9).

**Lemma 3.45 (Integral Over a Null Set)**

Let  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be measurable. If  $A \in \mathcal{A}$ , with  $\mu(A) = 0$ , then  $1_A \cdot f$  is  $\mu$ -integrable and

$$\int_A f d\mu = \int 1_A \cdot f d\mu = 0. \quad (3.42)$$

(Proof p. 116)

**Remark 3.46 (Integration Over Null Sets Can be Neglected)** The conjunction of Equations (3.36) and (3.42) implies: If  $f$  is quasi- $\mu$ -integrable and  $A \in \mathcal{A}$  with  $\mu(A) = 0$ , then

$$\int f d\mu = \int_{\Omega} f d\mu = \int_{\Omega \setminus A} f d\mu + \int_A f d\mu = \int_{\Omega \setminus A} f d\mu. \quad (3.43)$$

&lt;

**Lemma 3.47 (Integrals of  $\mu$ -Equivalent Functions)**

Let  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be quasi- $\mu$ -integrable. Then

$$f \stackrel{\mu}{=} g \Rightarrow \int f d\mu = \int g d\mu. \quad (3.44)$$

(Proof p. 116)

The following theorem presents a condition that is necessary and sufficient for  $\mu$ -equivalence of  $f$  and  $g$ .

**Theorem 3.48 (Identity of Integrals of  $\mu$ -Equivalent Functions)**

If  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  are  $\mu$ -integrable, then

$$f \stackrel{\mu}{=} g \Leftrightarrow \int_A f d\mu = \int_A g d\mu, \quad \forall A \in \mathcal{A}. \quad (3.45)$$

(Proof p. 117)

In section 3.4 we shall see that, if  $f$  and  $g$  are  $\mu$ -integrable and nonnegative, then it is sufficient to consider the integrals over all sets  $A$  in a  $\cap$ -stable generating system of  $\mathcal{A}$  in order to show  $\mu$ -equivalence of  $f$  and  $g$  (see Th. 3.68).

**Remark 3.49 (A Counter-Example)** Note that Equation (3.45) does not hold if  $f, g$  are nonnegative but not  $\mu$ -integrable measurable functions. This is exemplified as follows: Consider  $f, g: (\mathbb{R}, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ , where  $\mathcal{A} = \{\mathbb{R}, \emptyset\}$ ,  $\mu(\mathbb{R}) = \infty$ ,  $f = 1$ , and  $g = 2$ . Then  $\int_{\mathbb{R}} f d\mu = \int_{\mathbb{R}} g d\mu = \infty$  and  $\int_{\emptyset} f d\mu = \int_{\emptyset} g d\mu = 0$ . Hence,  $\int_A f d\mu = \int_A g d\mu$ , for all  $A \in \mathcal{A}$ , but  $f$  and  $g$  are *not equivalent* with respect to  $\mu$ .

&lt;

**Remark 3.50 (Some Special Cases)** Theorem 3.48 implies: If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is a measurable function, then  $f \stackrel{\mu}{=} \alpha$ ,  $\alpha \in \mathbb{R}$ , is equivalent to

$$\int_A f d\mu = \int_A \alpha d\mu = \int 1_A \alpha d\mu = \alpha \cdot \int 1_A d\mu = \alpha \cdot \mu(A), \quad \forall A \in \mathcal{A}. \quad (3.46)$$

Furthermore,  $f \stackrel{\mu}{=} 0$  is equivalent to

$$\int_A f d\mu = 0, \quad \forall A \in \mathcal{A}, \quad (3.47)$$

using the convention  $0 \cdot \infty = 0$ , if necessary.

An immediate implication of Equation (3.46) for  $A = \Omega$  is

$$f \stackrel{\mu}{=} \alpha, \quad \alpha \in \mathbb{R} \quad \Rightarrow \quad \int f d\mu = \alpha \mu(\Omega). \quad (3.48)$$

For  $\alpha = 0$ , this yields

$$f \stackrel{\mu}{=} 0 \quad \Rightarrow \quad \int f d\mu = 0. \quad (3.49)$$

◁

**Remark 3.51 (Almost Everywhere)** The notion of  $\mu$ -equivalence of  $f$  and  $g$  is an example of a property that holds for all  $\omega \in \Omega \setminus A$  with  $\mu(A) = 0$ . We also say that such a property holds  $\mu$ -almost everywhere ( $\mu$ -a.e.). Another example is the property

$$f(\omega) \leq g(\omega), \quad \forall \omega \in \Omega \setminus A \text{ and } \mu(A) = 0,$$

which is denoted by  $f \stackrel{\mu}{\leq} g$ .

◁

The proposition of Lemma 3.47 analogously holds for the relations  $\stackrel{\mu}{\leq}$  and  $\stackrel{\mu}{<}$ . The following theorem generalizes Lemma 3.26.

### Theorem 3.52 (Monotonicity)

Let  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be measurable functions.

(i) If  $f$  and  $g$  are quasi- $\mu$ -integrable, then

$$f \stackrel{\mu}{\leq} g \quad \Rightarrow \quad \int f d\mu \leq \int g d\mu. \quad (\text{monotonicity}) \quad (3.50)$$

(ii) If  $\mu(\Omega) > 0$  and  $f, g$  are  $\mu$ -integrable, then

$$f \stackrel{\mu}{<} g \quad \Rightarrow \quad \int f d\mu < \int g d\mu. \quad (\text{strict monotonicity}) \quad (3.51)$$

(Proof p. 117)

### 3.2.2 Integral With Respect to a Weighted Sum of Measures

In Example 1.62 we already noted that a weighted sum of measures with nonnegative weights is again a measure. As a special case, if  $\mu$  is a measure on  $(\Omega, \mathcal{A})$  and  $\alpha$  is a nonnegative number, then  $\alpha \cdot \mu$  is a measure on  $(\Omega, \mathcal{A})$  as well. Furthermore, if  $f$  is  $\mu$ -integrable, then

$$\int f d(\alpha\mu) = \int \alpha f d\mu = \alpha \int f d\mu \quad (3.52)$$

(see Exercise 3-10). This is generalized in the following theorem.

**Theorem 3.53 (Integral With Respect to a Weighted Sum of Measures)**

If  $f : (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is measurable and nonnegative,  $\mu_1, \mu_2, \dots$  are measures on  $(\Omega, \mathcal{A})$ , and  $\alpha_1, \alpha_2, \dots \in \mathbb{R}$  are nonnegative, then

$$\int f d\left(\sum_{i=1}^{\infty} \alpha_i \mu_i\right) = \sum_{i=1}^{\infty} \alpha_i \int f d\mu_i. \quad (3.53)$$

For a proof, see Equation (3.52) and Bauer (2001, Example 3 on page 61).

If we consider a *finite* weighted sum of measures, the assumption that  $f$  is nonnegative can be replaced by integrability of  $f$ . In the following theorem we consider a weighted sum of two measures. In Remark 3.55 we extend the result to a finite weighted sum of measures.

**Theorem 3.54 (Integral With Respect to a Weighted Sum of two Measures)**

Let  $\mu_1, \mu_2$  be measures on  $(\Omega, \mathcal{A})$ . If  $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is integrable with respect to  $\mu_1$  and  $\mu_2$ , and  $0 \leq \alpha_1, \alpha_2 \in \mathbb{R}$ , then  $f$  is integrable with respect to  $\alpha_1 \mu_1 + \alpha_2 \mu_2$ , and

$$\int f d(\alpha_1 \mu_1 + \alpha_2 \mu_2) = \alpha_1 \int f d\mu_1 + \alpha_2 \int f d\mu_2. \quad (3.54)$$

For a proof, see Equation (3.52) and Bauer (2001, Example 5 on page 67).

**Remark 3.55 (Integral With Respect to a Finite Weighted Sum of Measures)**

By induction, Theorem 3.54 yields, for nonnegative  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ ,

$$\int f d\left(\sum_{i=1}^n \alpha_i \mu_i\right) = \sum_{i=1}^n \alpha_i \int f d\mu_i, \quad (3.55)$$

provided that  $f$  is integrable with respect to all measures  $\mu_1, \dots, \mu_n$ . ◁

**Example 3.56 (Integral With Respect to the Weighted Sum of Dirac Measures)**

Let  $(\Omega, \mathcal{A})$  be a measurable space and, for  $i \in \mathbb{N}$ , let  $\omega_i \in \Omega$ ,  $\alpha_i \in \mathbb{R}$ ,  $\alpha_i \geq 0$ , and  $\delta_{\omega_i}$  denote the Dirac measure at  $\omega_i$ . Then

$$\mu = \sum_{i=1}^{\infty} \alpha_i \delta_{\omega_i} \quad (3.56)$$

defined by  $\mu(A) = \sum_{i=1}^{\infty} \alpha_i \delta_{\omega_i}(A)$ , for all  $A \in \mathcal{A}$ , is a measure on  $(\Omega, \mathcal{A})$  (see Example 1.62). For any nonnegative measurable function  $f : \Omega \rightarrow \mathbb{R}$  we obtain

$$\int f d\mu = \int f d \sum_{i=1}^{\infty} \alpha_i \delta_{\omega_i} \quad [(3.56)]$$

$$= \sum_{i=1}^{\infty} \alpha_i \int f d\delta_{\omega_i} \quad [(3.53)] \quad (3.57)$$

$$= \sum_{i=1}^{\infty} \alpha_i f(\omega_i). \quad [(3.23)]$$

For  $\mu(A) = \sum_{i=1}^n \alpha_i \delta_{\omega_i}(A)$ , for all  $A \in \mathcal{A}$ , Equation (3.57) with  $\alpha_i = 0$  for  $i > n$ , yields

$$\int f d\mu = \sum_{i=1}^n \alpha_i f(\omega_i). \quad (3.58)$$

Hence, the integral of a nonnegative measurable function  $f$  with respect to a finite or countable weighted sum of Dirac measures with nonnegative weights is a weighted sum of values of  $f$ .  $\triangleleft$

### 3.2.3 Integral With Respect to an Image Measure

The next theorem is relevant whenever we consider the integral of a composition  $g \circ f$  of a mapping  $f$  with a numerical function  $g$  [see Eq. (2.25)] or the integral with respect to the image measure  $\mu_f$  of  $\mu$  under  $f$  [see Def. 2.80].

#### Theorem 3.57 (Transformation Theorem)

Let  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$  and  $g: (\Omega', \mathcal{A}') \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}}')$  be measurable.

(i) If  $g$  is nonnegative or integrable with respect to  $\mu_f$ , then

$$\int g d\mu_f = \int g \circ f d\mu. \quad (3.59)$$

(ii)  $g$  is integrable with respect to  $\mu_f$  if and only if  $g \circ f$  is  $\mu$ -integrable.

For a proof, see Bauer (2001, Corollary 19.2.1, p. 110).

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  is a numerical measurable function and we replace  $g$  by the identity function  $id: (\bar{\mathbb{R}}, \bar{\mathcal{B}}) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ , then Theorem 3.57 implies the following corollary.

#### Corollary 3.58 (An Implication of the Transformation Theorem)

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  is nonnegative or  $\mu$ -integrable, then

$$\int id d\mu_f = \int_{\bar{\mathbb{R}}} id d\mu_f = \int f d\mu = \int_{\Omega} f d\mu. \quad (3.60)$$

Using the alternative notation of an integral introduced in Remark 3.32, Equation (3.59) can also be written

$$\int_{\overline{\mathbb{R}}} g(x) \mu_f(dx) = \int_{\Omega} g[f(\omega)] \mu(d\omega). \quad (3.61)$$

Correspondingly, Equation (3.60) can also be written

$$\int_{\overline{\mathbb{R}}} x \mu_f(dx) = \int_{\Omega} f(\omega) \mu(d\omega). \quad (3.62)$$

In Definition 3.10 we considered the case in which  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  is a nonnegative step function and defined its integral by  $\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$ , presuming that  $A_1, \dots, A_n$  are pairwise disjoint. Now we consider a measurable function  $f$  with a finite number of values, which can be 0, positive, or negative.

**Corollary 3.59 (Integral of a Function With a Finite Number of Values)**

If  $(\Omega, \mathcal{A}, \mu)$  is a measure space and  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  with pairwise different  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ ,  $\alpha_i \neq 0$ , and pairwise disjoint  $A_1, \dots, A_n \in \mathcal{A}$ , then  $f$  is  $\mu$ -integrable if and only if  $\mu(A_i) < \infty$  for all  $i = 1, \dots, n$ . If  $f$  is  $\mu$ -integrable, then

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i) = \sum_{i=1}^n \alpha_i \mu_f(\{\alpha_i\}). \quad (3.63)$$

(Proof p. 118)

### 3.2.4 Convergence Theorems

The next two theorems deal with convergence of integrals. In the first one, we assume that  $f_1, f_2, \dots$  is an increasing sequence of measurable functions that converge to  $f$ .

**Theorem 3.60 (Monotone Convergence; B. Levi)**

Let the functions  $f_n: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be measurable, for all  $n \in \mathbb{N}$ .

- (i) If the sequence  $f_1, f_2, \dots$  is increasing with  $\lim_{n \rightarrow \infty} f_n = f$  and the functions  $f_n$  are nonnegative for all  $n \in \mathbb{N}$  or  $\mu$ -integrable for all  $n \in \mathbb{N}$ , then

$$\int f d\mu = \lim_{i \rightarrow \infty} \int f_n d\mu. \quad (3.64)$$

- (ii) If the functions  $f_i$  are nonnegative for all  $i \in \mathbb{N}$ , then

$$\int \left( \sum_{i=1}^{\infty} f_i \right) d\mu = \sum_{i=1}^{\infty} \int f_i d\mu. \quad (3.65)$$

The integrals on both sides are finite or  $+\infty$ .

For a proof of (i), assuming nonnegativity, see Bauer (2001, Theorem 11.4, p. 59). For a proof of (i), assuming integrability, see Klenke (2013, Theorem 4.20, p. 93). For a proof of (ii), see Bauer (2001, Corollary 11.5, p. 60).

Note that, by definition of an ‘infinite sum’ (see Box 0.1), Equation (3.65) is equivalent to

$$\int \left( \lim_{n \rightarrow \infty} \sum_{i=1}^n f_i \right) d\mu = \lim_{n \rightarrow \infty} \sum_{i=1}^n \int f_i d\mu. \quad (3.66)$$

In the next theorem, we replace the assumption that  $f_1, f_2, \dots$  is increasing by the assumption that there is a  $\mu$ -integrable function  $g$  dominating the absolute value functions of all  $f_n$ .

**Theorem 3.61 (Dominated Convergence; Lebesgue Convergence Theorem)**

If  $g, f_n: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ ,  $n \in \mathbb{N}$ , are  $\mu$ -integrable and there is a measurable function  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  with  $\lim_{n \rightarrow \infty} f_n = f$ , and  $|f_n| \leq g$  for all  $n \in \mathbb{N}$ , then

$$\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu, \quad (3.67)$$

and this integral is finite.

For a proof, see Bauer (2001, Theorem 15.6, p. 83).

### 3.3 Lebesgue and Riemann Integral

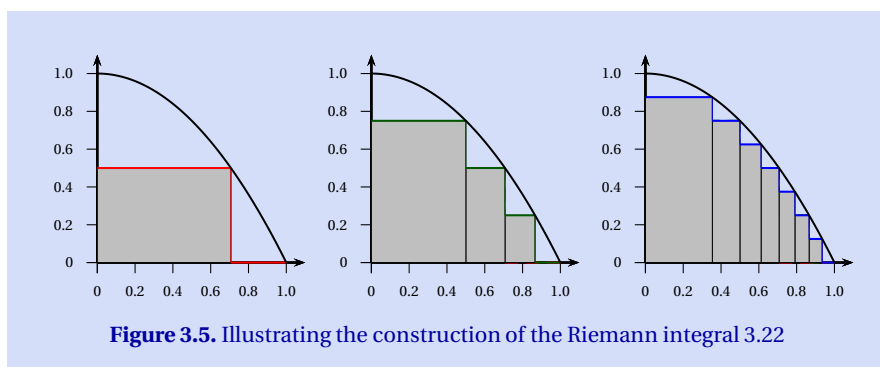
The Lebesgue measures  $\lambda_n$  on  $(\mathbb{R}^n, \mathcal{B}_n)$ ,  $n = 1, 2, 3$ , represent *length*, *area*, and *volume*, respectively. As the examples illustrated by Figure 3.4 shows, the integral of the  $(\mathcal{B}, \mathcal{B})$ -measurable function  $f: \mathbb{R} \rightarrow \mathbb{R}$  with respect to the Lebesgue measure  $\lambda = \lambda_1$ , i. e., the Lebesgue integral, yields the difference between the *areas* marked by + and the areas marked by –, respectively.

It is useful to know conditions under which the Lebesgue integral and the Riemann integrals are identical, because a lot of tools are available for Riemann integration (see, e. g., Ellis & Gulick, 2006). The following theorem is proved in Klenke (2013, Theorem 4.23, p. 96), who also provides a brief definition of the Riemann integral and Riemann integrability.

**Theorem 3.62 (Lebesgue Integral and Riemann Integral)**

Let  $\lambda$  denote the Lebesgue measure on  $(\mathbb{R}, \mathcal{B})$  and let  $[a, b]$ ,  $a, b \in \mathbb{R}$ ,  $a < b$ , be a closed interval. If  $f: [a, b] \rightarrow \mathbb{R}$  is Riemann integrable on  $[a, b]$ , then  $f$  is  $\lambda$ -integrable, and

$$\int_a^b f(x) dx = \int_{[a,b]} f d\lambda = \int_{[a,b]} f(x) \lambda(dx) = \int 1_{[a,b]} \cdot f d\lambda, \quad (3.68)$$



where  $\int_a^b f(x) dx$  denotes the Riemann integral from  $a$  to  $b$ .

Note that

$$\int_a^b f(x) dx = F(x) \Big|_a^b := F(b) - F(a), \quad (3.69)$$

where  $F$  is an antiderivative of  $f$ .

**Remark 3.63 (Lebesgue vs. Riemann Integral)** If we want to define the integral of a measurable function  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ , where the set  $\Omega$  is not necessarily a subset of  $\mathbb{R}$ , then this means that the traditional Riemann integral cannot be used. The *Riemann integral* is constructed by partitioning the *domain* of  $f$ , the set  $\mathbb{R}$  of real numbers into small intervals and adding the area of the rectangles on these intervals in order to approximate the area under the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  (see Fig. 3.5). If  $\Omega \not\subset \mathbb{R}$ , then this idea does not work any more. Instead, the *Lebesgue integral* is constructed by partitioning the *codomain* of  $f$ , which is the set  $\mathbb{R}$  of real numbers into small intervals (see Fig. 3.3). This is also possible if  $\Omega \not\subset \mathbb{R}$ , and in this aspect, the Lebesgue integral is more general than the Riemann integral.  $\triangleleft$

Note, however, that even if the domain of  $f$  is a subset of the set of real numbers, there are functions for which the Lebesgue integral exists and the Riemann integral does not exist (see, e. g., Klenke, 2013, Example 4.24, p. 97). Also note that there are functions that are Riemann integrable on a half-open or unbounded interval but not Lebesgue integrable (see, e. g., Klenke, 2013, Remark 4.25, p. 97).

**Example 3.64 (Using the Riemann Integral)** As a simple application of Theorem 3.62 consider the function  $f$  defined by  $f(x) = 10 - x^2$  on a closed interval  $[a, b]$ . Because  $f$  is a continuous function, it is Riemann integrable. Hence, we can apply Equation (3.68). For  $a = -5$  and  $b = 5$ , this equation yields

$$\int_{[-5,5]} f d\lambda = \int_{-5}^5 f(x) dx = 16.\bar{6}$$

(see Exercise 3-11). This integral is the difference between the areas marked by + and the areas marked by – in Figure 3.4.  $\triangleleft$

### 3.4 Density

A density  $f$  can be interpreted as a weighting function of the values of the original measure  $\mu$ . If we consider a measure  $\mu$  on a measurable space  $(\Omega, \mathcal{A})$  with  $\mu(\{x\}) > 0$  for all  $x \in \Omega$ , then this means that the values  $\mu(\{x\})$  of the singletons  $\{x\}$  are multiplied by a nonnegative number  $f(x)$ . If we consider the Lebesgue measure  $\lambda$  on  $\mathbb{R}$ , then, intuitively speaking, any infinitesimal interval  $dx$  gets a weight  $f(x)$ . Using such a density, a new measure  $\nu$  on  $\mathcal{A}$  is introduced, where  $\nu(A)$  is the integral of  $f$  over  $A$  with respect to  $\mu$ . The most important examples are densities with respect to the Lebesgue measure (see Example 3.69).

#### Theorem 3.65 (Measure With Density)

Let  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  be a nonnegative measurable function. Then the function  $\nu: \mathcal{A} \rightarrow \bar{\mathbb{R}}$  defined by

$$\nu(A) = \int_A f d\mu, \quad \forall A \in \mathcal{A}, \quad (3.70)$$

is a measure, called the measure with density  $f$  with respect to  $\mu$ . It is denoted by  $f \circ \mu$ , i. e.,  $f \circ \mu := \nu$ .

For a proof, see Bauer (2001, Theorem 17.1, p. 96).

The notation  $f \circ \mu$  has been adopted from Elstrodt (2007, p. 127). Using this notation, Equation (3.70) can also be written

$$f \circ \mu(A) = \int_A f d\mu, \quad \forall A \in \mathcal{A}. \quad (3.71)$$

#### Definition 3.66 (Density)

Let  $\nu$  be a measure on  $(\Omega, \mathcal{A})$ . If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  is a nonnegative measurable function satisfying Equation (3.70), then it is called a density of  $\nu$  with respect to  $\mu$ .

The following theorem generalizes Equation (3.52).

**Theorem 3.67 (Integral With Respect to a Measure With Density)**

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space and  $f: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  a nonnegative measurable function. Furthermore, let  $f \circ \mu: \mathcal{A} \rightarrow \overline{\mathbb{R}}$  be the measure with density  $f$  with respect to  $\mu$  and let  $g: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  be measurable.

(i) If  $g$  is nonnegative, then

$$\int g \, d f \circ \mu = \int g \cdot f \, d \mu. \quad (3.72)$$

(ii)  $g$  is integrable with respect to  $f \circ \mu$  if and only if  $g \cdot f$  is  $\mu$ -integrable.

(iii) If  $g$  is integrable with respect to  $f \circ \mu$ , then Equation (3.72) holds.

For a proof, see Bauer (2001, Theorem 17.3, p. 96, 97).

In the following theorem we summarize some necessary and sufficient conditions for  $\mu$ -equivalence of measurable functions on a measure space.

**Theorem 3.68 (Necessary and Sufficient Condition of  $\mu$ -Equivalence)**

Let  $f, g: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be measurable functions, let  $\mathcal{E} \subset \mathcal{A}$ , and consider:

(a)  $f \stackrel{\mu}{=} g$ .

(b)  $\int_A f \, d \mu = \int_A g \, d \mu, \quad \forall A \in \mathcal{A}.$

(c)  $f \circ \mu = g \circ \mu.$

(d)  $\int_A f \, d \mu = \int_A g \, d \mu, \quad \forall A \in \mathcal{E}.$

Then:

(i) (a)  $\Rightarrow$  (b), if  $f$  and  $g$  are quasi- $\mu$ -integrable.

(ii) (a)  $\Leftrightarrow$  (b), if  $f, g$  are  $\mu$ -integrable.

(iii) (a)  $\Leftrightarrow$  (c), if  $f, g$  are  $\mu$ -integrable and nonnegative.

(iv) (a)  $\Leftrightarrow$  (b)  $\Leftrightarrow$  (c)  $\Leftrightarrow$  (d), if  $f, g$  are  $\mu$ -integrable, nonnegative, and  $\mathcal{E} \subset \mathcal{A}$  is  $\cap$ -stable with  $\sigma(\mathcal{E}) = \mathcal{A}$ .

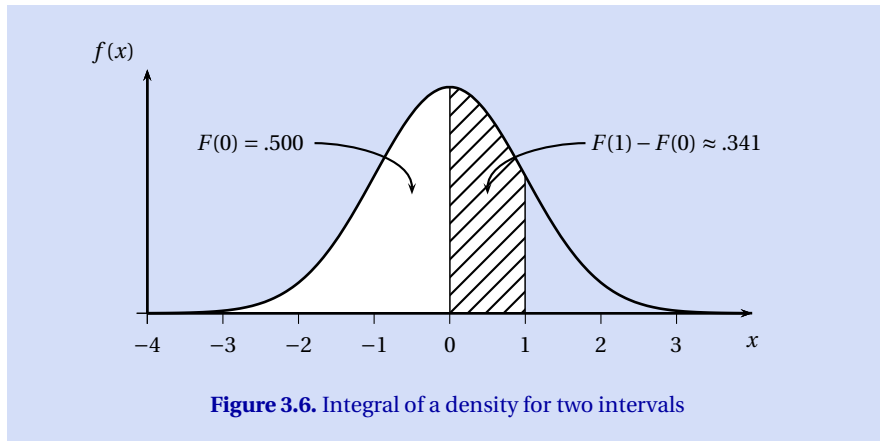
(Proof p. 119)

**Example 3.69 (A Density of the Normal Distribution)** As a special case of Equation (3.70) consider

$$\nu(A) = \int_A f \, d \lambda, \quad \forall A \in \mathcal{A}, \quad (3.73)$$

with

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-x^2}{2}\right), \quad x \in \mathbb{R}. \quad (3.74)$$



In this case, the measure  $\nu = f \circ \lambda$  is a probability measure and it is called the *standard normal distribution*. For an interval  $[a, b]$ , Theorem 3.67 yields

$$\nu([a, b]) = \int 1_{[a, b]} d\nu = \int 1_{[a, b]} d f \circ \lambda = \int 1_{[a, b]} f d\lambda = \int_a^b f(x) dx, \quad (3.75)$$

because  $f$  is Riemann  $\mu$ -integrable (see Th. 3.62). According to this equation, the value  $\nu([a, b])$  of the interval  $[a, b]$  can be represented as the area between the density and the  $x$ -axis above  $[a, b]$ . Figure 3.6 illustrates this fact for the interval  $[0, 1]$ . In this figure,

$$F(\alpha) = \int_{-\infty}^{\alpha} f(x) dx, \quad \alpha \in \mathbb{R}, \quad (3.76)$$

denotes the corresponding distribution function (see Def. 5.81), which is a special antiderivative of  $f$  [see Eq. (3.69)].  $\triangleleft$

### 3.5 Absolute Continuity and the Radon-Nikodym Theorem

Let  $\mu$  and  $\nu$  be measures on a measurable space  $(\Omega, \mathcal{A})$ . A necessary and sufficient condition for the existence of a density of  $\nu$  with respect to  $\mu$  is formulated in the Radon-Nikodym Theorem (see Th. 3.72), which is used not only for densities but also introducing conditional expectations (see ch. 10). The following definition prepares this theorem.

#### Definition 3.70 (Absolute Continuity)

Let  $\mu$  and  $\nu$  be measures on a measurable space  $(\Omega, \mathcal{A})$ .

- (i) The measure  $\nu$  is called *absolutely continuous with respect to  $\mu$* , denoted  $\nu \ll_{\mathcal{A}} \mu$ , if

$$\forall A \in \mathcal{A}: \mu(A) = 0 \Rightarrow \nu(A) = 0. \quad (3.77)$$

(ii) The measures  $\mu$  and  $\nu$  are called *null-set equivalent*, denoted  $\nu \approx \mu$ , if  $\nu \ll_{\mathcal{A}} \mu$  and  $\mu \ll_{\mathcal{A}} \nu$ , i. e., if

$$\forall A \in \mathcal{A}: \mu(A) = 0 \Leftrightarrow \nu(A) = 0. \quad (3.78)$$

If there is ambiguity about the measurable space, we use the terms *absolutely continuous on  $(\Omega, \mathcal{A})$*  and *null-set equivalent on  $(\Omega, \mathcal{A})$* .

**Remark 3.71 (An Implication)** If there is a density  $f$  of  $\nu$  with respect to  $\mu$ , then  $\nu \ll_{\mathcal{A}} \mu$ . This is a straightforward implication of Lemma 3.45 and (3.72) (see Exercise 3-12).  $\triangleleft$

Vice versa, if  $\nu \ll_{\mathcal{A}} \mu$ , then, according to the following theorem, there is a density  $f$  of  $\nu$  with respect to  $\mu$ , provided that  $\mu$  and  $\nu$  are  $\sigma$ -finite (see Definition 1.63).

**Theorem 3.72 (Radon-Nikodym)**

Let  $\mu$  and  $\nu$  be  $\sigma$ -finite measures on a measurable space  $(\Omega, \mathcal{A})$ .

- (i) Then  $\nu$  has a measurable density with respect to  $\mu$  if and only if  $\nu \ll_{\mathcal{A}} \mu$ . This density is denoted by  $\frac{d\nu}{d\mu}$  and called *Radon-Nikodym derivative*.
- (ii) If  $\nu \ll_{\mathcal{A}} \mu$ , then  $\frac{d\nu}{d\mu}$  is real-valued  $\mu$ -almost everywhere.

For a proof, see Klenke (2013, Corollary 7.34, p. 158) or Bauer (2001, Theorem 17.10, p. 102 and Theorem 17.11, p. 104).

**Remark 3.73 ( $\mu$ -Equivalence of Densities)** Note that, for  $\sigma$ -finite measures, all densities of  $\nu$  with respect to  $\mu$  are pairwise  $\mu$ -equivalent (for a proof see Bauer, 2001, Theorem 17.11, p. 104). If  $\nu$  is a finite measure, which is equivalent to  $\mu$ -integrability of the density  $\frac{d\nu}{d\mu}$ , then Theorem 3.68 (iii) implies  $\mu$ -equivalence of all densities of  $\nu$  with respect to  $\mu$ . The premise that  $\nu$  is a finite holds in particular if  $\nu$  is a probability measure.  $\triangleleft$

**Remark 3.74 (An Implication of the Radon-Nikodym Theorem)** Theorem 3.72 implies for  $\sigma$ -finite measures  $\nu$  and  $\mu$ : If  $\nu$  and  $\mu$  are null-set equivalent, then  $\frac{d\nu}{d\mu}$  and  $\frac{d\mu}{d\nu}$  both exist.  $\triangleleft$

The Radon-Nikodym theorem is used to prove the existence of the conditional expectation [see the proof of Theorem 10.9 (Bauer, 1996, Theorem 15.1, p. 111)]. The following corollary immediately follows from Theorem 3.67 and 3.72.

**Corollary 3.75 (An Implication of the Radon-Nikodym Theorem)**

Let  $\mu$  and  $\nu$  be  $\sigma$ -finite measures on a measurable space  $(\Omega, \mathcal{A})$ , and suppose  $\nu \ll_{\mathcal{A}} \mu$ . Furthermore, let  $g: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a measurable function.

(i) If  $g$  is nonnegative or  $\nu$ -integrable, then

$$\int g d\nu = \int g \cdot \frac{d\nu}{d\mu} d\mu. \quad (3.79)$$

(ii)  $g$  is  $\nu$ -integrable if and only if  $g \cdot \frac{d\nu}{d\mu}$  is  $\mu$ -integrable.

**3.6 Integral With Respect to a Product Measure**

The following theorem shows that integration with respect to a product measure can be decomposed into a two-fold iterated integration where the order of integration is arbitrary.

**Theorem 3.76 (Fubini)**

Let  $(\Omega_i, \mathcal{A}_i, \mu_i)$ ,  $i = 1, 2$ , be  $\sigma$ -finite measure spaces and let  $f: \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$  be  $(\mathcal{A}_1 \otimes \mathcal{A}_2, \overline{\mathcal{B}})$ -measurable. Furthermore, let  $f_i: \Omega_i \rightarrow \overline{\mathbb{R}}$ ,  $i = 1, 2$ , be defined by

$$f_1(\omega_1) := \int f(\omega_1, \omega_2) \mu_2(d\omega_2) \quad \text{and} \quad f_2(\omega_2) := \int f(\omega_1, \omega_2) \mu_1(d\omega_1).$$

If  $f$  is nonnegative or integrable with respect to the product measure  $\mu_1 \otimes \mu_2$ , then the functions  $f_i$  are  $(\mathcal{A}_i, \overline{\mathcal{B}})$ -measurable,  $i = 1, 2$ . Furthermore,

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2) &= \int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) \mu_1 \otimes \mu_2 [d(\omega_1, \omega_2)] \\ &= \int_{\Omega_1} \left( \int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &= \int_{\Omega_2} \left( \int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1) \right) \mu_2(d\omega_2). \end{aligned} \quad (3.80)$$

For a proof, see Klenke (2013, Th. 14.16, p. 278). If  $f = 1_C$  for  $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$ , then this theorem and (3.9) immediately yield the following corollary:

**Corollary 3.77 (Indicators)**

Let  $(\Omega_i, \mathcal{A}_i, \mu_i)$ ,  $i = 1, 2$ , be  $\sigma$ -finite measure spaces and let  $f: \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$  be  $(\mathcal{A}_1 \otimes \mathcal{A}_2, \overline{\mathcal{B}})$ -measurable. Furthermore, let  $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$  and define

$$\begin{aligned}
& \forall \omega_1 \in \Omega_1: C_{\omega_1} := \{\omega_2 \in \Omega_2: (\omega_1, \omega_2) \in C\} \\
& \text{and} \\
& \forall \omega_2 \in \Omega_2: C_{\omega_2} := \{\omega_1 \in \Omega_1: (\omega_1, \omega_2) \in C\}. \\
& \text{Then} \\
& \mu_1 \otimes \mu_2(C) = \int \mu_2(C_{\omega_1}) \mu_1(d\omega_1) = \int \mu_1(C_{\omega_2}) \mu_2(d\omega_2). \quad (3.81) \\
& \hspace{15em} (\text{Proof p. 120})
\end{aligned}$$

**Remark 3.78 (A Special Case)** Choosing  $C = A_1 \times A_2$ , Equation (3.81) yields

$$\begin{aligned}
\mu_1 \otimes \mu_2(A_1 \times A_2) &= \int \mu_2(A_2) \cdot \mathbf{1}_{A_1}(\omega_1) \mu_1(d\omega_1) = \mu_2(A_2) \cdot \int \mathbf{1}_{A_1}(\omega_1) \mu_1(d\omega_1) \\
&= \mu_1(A_1) \cdot \mu_2(A_2),
\end{aligned}$$

which is consistent with Equation (1.50).  $\triangleleft$

### 3.7 Proofs

#### *Proof of Lemma 3.33*

(i) If  $f$  is measurable and  $A \in \mathcal{A}$ , then  $\mathbf{1}_A$  is measurable as well (see Th. 2.57 and Example 2.12). Suppose that  $f$  is quasi- $\mu$ -integrable, i. e., suppose that  $\int f^+ d\mu$  or  $\int f^- d\mu$  are finite. Because

$$(\mathbf{1}_A f)^+ = \mathbf{1}_A f^+ \quad \text{and} \quad (\mathbf{1}_A f)^- = \mathbf{1}_A f^-$$

as well as

$$0 \leq \mathbf{1}_A f^+ \leq f^+ \quad \text{and} \quad 0 \leq \mathbf{1}_A f^- \leq f^-,$$

monotonicity of the integral of nonnegative measurable functions (Lemma 3.26) yields

$$0 \leq \int (\mathbf{1}_A f)^+ d\mu = \int \mathbf{1}_A f^+ d\mu \leq \int f^+ d\mu$$

and

$$0 \leq \int (\mathbf{1}_A f)^- d\mu = \int \mathbf{1}_A f^- d\mu \leq \int f^- d\mu,$$

which implies that  $\int (\mathbf{1}_A f)^+ d\mu$  or  $\int (\mathbf{1}_A f)^- d\mu$  is finite. Hence  $f$  is quasi- $\mu$ -integrable.

(ii) If  $f$  is  $\mu$ -integrable, then  $\int f^+ d\mu < \infty$  and  $\int f^- d\mu < \infty$ . Just like in the proof of (i) this implies  $\int (\mathbf{1}_A f)^+ d\mu < \infty$  and  $\int (\mathbf{1}_A f)^- d\mu < \infty$ . Hence,  $\mathbf{1}_A f$  is  $\mu$ -integrable.

#### *Proof of Theorem 3.36*

(i) Step 1: Let  $\alpha \geq 0$ ,  $f$  be a nonnegative step function, and  $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$  a normal representation (see Rem. 3.8). Then, according to (3.4),

$$\begin{aligned}
\int \alpha f d\mu &= \int \alpha \sum_{i=1}^n \alpha_i 1_{A_i} d\mu && \text{[Rem. 3.8]} \\
&= \int \sum_{i=1}^n (\alpha \alpha_i) 1_{A_i} d\mu \\
&= \sum_{i=1}^n (\alpha \alpha_i) \mu(A_i) && \text{[(3.4)]} \\
&= \alpha \sum_{i=1}^n \alpha_i \mu(A_i) \\
&= \alpha \int f d\mu. && \text{[(3.4)]}
\end{aligned}$$

Step 2: Let  $\alpha \geq 0$ ,  $f$  be a nonnegative measurable function, and  $f_1 \leq f_2 \leq \dots$  an increasing sequence of nonnegative step functions with  $\lim_{n \rightarrow \infty} f_n = f$  (see Th. 3.19). Then, according to Equation (3.22),

$$\begin{aligned}
\int \alpha f d\mu &= \int \alpha \lim_{n \rightarrow \infty} f_n d\mu \\
&= \int \lim_{n \rightarrow \infty} (\alpha f_n) d\mu \\
&= \lim_{n \rightarrow \infty} \int \alpha f_n d\mu && \text{[(3.22)]} \\
&= \lim_{n \rightarrow \infty} \alpha \int f_n d\mu && \text{[Step 1]} \\
&= \alpha \lim_{n \rightarrow \infty} \int f_n d\mu \\
&= \alpha \int f d\mu. && \text{[(3.22)]}
\end{aligned}$$

Step 3: Assume that  $\alpha \geq 0$  and that  $f$  is quasi- $\mu$ -integrable. Because  $\alpha f = \alpha(f^+ - f^-) = \alpha f^+ - \alpha f^-$ ,

$$\begin{aligned}
\int \alpha f d\mu &= \int \alpha f^+ d\mu - \int \alpha f^- d\mu && \text{[(3.27)]} \\
&= \alpha \int f^+ d\mu - \alpha \int f^- d\mu && \text{[Step 2]} \\
&= \alpha \left( \int f^+ d\mu - \int f^- d\mu \right) \\
&= \alpha \int f d\mu. && \text{[(3.27)]}
\end{aligned}$$

This proves Equation (3.32) for  $\alpha \geq 0$ . For  $\alpha < 0$ , note that

$$(\alpha f)^+ = -\alpha f^- \quad \text{and} \quad (\alpha f)^- = -\alpha f^+. \quad (3.82)$$

Therefore,

$$\int \alpha f d\mu = \int (\alpha f)^+ d\mu - \int (\alpha f)^- d\mu \quad \text{[(3.27)]}$$

$$\begin{aligned}
&= \int (-\alpha) f^- d\mu - \int (-\alpha) f^+ d\mu && [(3.82)] \\
&= -\alpha \int f^- d\mu - (-\alpha) \int f^+ d\mu && [-\alpha > 0, \text{ first part of Step 3}] \\
&= \alpha \left( \int f^+ d\mu - \int f^- d\mu \right) \\
&= \alpha \int f d\mu. && [(3.27)]
\end{aligned}$$

This shows that  $\int \alpha f d\mu = \alpha \int f d\mu$  holds for all  $\alpha \in \mathbb{R}$ , all quasi- $\mu$ -integrable  $f$ , and therefore also for all integrable  $f$ . This also implies that  $\alpha f$  is quasi- $\mu$ -integrable or  $\mu$ -integrable if  $f$  is quasi- $\mu$ -integrable or  $\mu$ -integrable, respectively.

(ii) Step 1: Let  $f$  and  $g$  be nonnegative step functions and let  $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ ,  $g = \sum_{j=1}^m \beta_j \mathbf{1}_{B_j}$  be normal representations (see Rem. 3.8) with  $\bigcup_{i=1}^n A_i = \Omega$  and  $\bigcup_{j=1}^m B_j = \Omega$ . (Note that these latter requirements can always be fulfilled using  $A_n := \Omega \setminus \bigcup_{i=1}^{n-1} A_i$  and  $\alpha_n := 0$ , if  $f = \sum_{i=1}^{n-1} \alpha_i \mathbf{1}_{A_i}$  is already a normal representation.) Then  $f + g = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} + \sum_{j=1}^m \beta_j \mathbf{1}_{B_j}$  is again a nonnegative step function (see Def. 3.9) and

$$f + g = \sum_{i=1}^n \sum_{j=1}^m \gamma_{ij} \mathbf{1}_{C_{ij}}$$

is a normal representation, where  $C_{ij} = A_i \cap B_j$  and  $\gamma_{ij} = \alpha_i + \beta_j$ . Note that some of these sets  $C_{ij}$  may be empty. Now,

$$\begin{aligned}
\int f + g d\mu &= \sum_{i=1}^n \sum_{j=1}^m \gamma_{ij} \mu(C_{ij}) && [(3.4)] \\
&= \sum_{i=1}^n \sum_{j=1}^m (\alpha_i + \beta_j) \mu(A_i \cap B_j) \\
&= \sum_{i=1}^n \sum_{j=1}^m \alpha_i \mu(A_i \cap B_j) + \sum_{i=1}^n \sum_{j=1}^m \beta_j \mu(A_i \cap B_j) \\
&= \sum_{i=1}^n \alpha_i \sum_{j=1}^m \mu(A_i \cap B_j) + \sum_{j=1}^m \beta_j \sum_{i=1}^n \mu(A_i \cap B_j) \\
&= \sum_{i=1}^n \alpha_i \mu(A_i) + \sum_{j=1}^m \beta_j \mu(B_j) && [\text{Rem. 1.47}] \\
&= \int f d\mu + \int g d\mu. && [(3.4)]
\end{aligned}$$

Step 2: Let  $f, g$  be nonnegative measure functions and  $f_1 \leq f_2 \leq \dots$ ,  $g_1 \leq g_2 \leq \dots$  increasing sequences of nonnegative step functions with  $\lim_{n \rightarrow \infty} f_n = f$  and  $\lim_{n \rightarrow \infty} g_n = g$ , respectively (see Th. 3.19). Then  $f_1 + g_1 \leq f_2 + g_2 \leq \dots$  is an increasing sequence of nonnegative step functions with  $\lim_{n \rightarrow \infty} (f_n + g_n) = \lim_{n \rightarrow \infty} f_n + \lim_{n \rightarrow \infty} g_n = f + g$ . Then

$$\begin{aligned}
\int f + g \, d\mu &= \int \lim_{n \rightarrow \infty} (f_n + g_n) \, d\mu \\
&= \lim_{n \rightarrow \infty} \int (f_n + g_n) \, d\mu && [(3.22)] \\
&= \lim_{n \rightarrow \infty} \left( \int f_n \, d\mu + \int g_n \, d\mu \right) && [\text{Step 1}] \\
&= \lim_{n \rightarrow \infty} \int f_n \, d\mu + \lim_{n \rightarrow \infty} \int g_n \, d\mu \\
&= \int f \, d\mu + \int g \, d\mu. && [(3.22)]
\end{aligned}$$

Step 3: Assume that  $f$  is quasi- $\mu$ -integrable and  $g$  is  $\mu$ -integrable. Then

$$\begin{aligned}
f + g &= f^+ - f^- + g^+ - g^-, \\
f + g &= (f + g)^+ - (f + g)^-.
\end{aligned}$$

This implies

$$(f + g)^+ - (f + g)^- = f^+ - f^- + g^+ - g^-,$$

which is equivalent to

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$

Applying the result of Step 2 yields

$$\begin{aligned}
&\int (f + g)^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu \\
&= \int (f + g)^- \, d\mu + \int f^+ \, d\mu + \int g^+ \, d\mu.
\end{aligned} \tag{3.83}$$

If  $g$  is  $\mu$ -integrable, then  $\int g^+ \, d\mu$  and  $\int g^- \, d\mu$  are finite, and if  $f$  is quasi- $\mu$ -integrable, then at most one of  $\int f^+ \, d\mu$  and  $\int f^- \, d\mu$  is infinite, the other one is finite. Furthermore,  $(f + g)^+ \leq f^+ + g^+$  and  $(f + g)^- \leq f^- + g^-$ . Hence, Lemma 3.26 implies

$$\int (f + g)^+ \, d\mu \leq \int f^+ + g^+ \, d\mu = \int f^+ \, d\mu + \int g^+ \, d\mu \tag{3.84}$$

and

$$\int (f + g)^- \, d\mu \leq \int f^- + g^- \, d\mu = \int f^- \, d\mu + \int g^- \, d\mu.$$

Therefore, at most one of the integrals  $\int (f + g)^+ \, d\mu$  and  $\int (f + g)^- \, d\mu$  is infinite and this implies that  $f + g$  is quasi- $\mu$ -integrable. If  $\int (f + g)^+ \, d\mu = \infty$ , then

$$\int (f + g) \, d\mu = \int (f + g)^+ \, d\mu - \int (f + g)^- \, d\mu = \infty,$$

and, according to (3.84),  $\int f^+ d\mu = \infty$ . This implies

$$\int f d\mu + \int g d\mu = \int f^+ d\mu - \int f^- d\mu + \int g d\mu = \infty.$$

Analogously, if  $\int (f+g)^- d\mu = \infty$ , then

$$\int (f+g) d\mu = -\infty = \int f d\mu + \int g d\mu.$$

If both,  $\int (f+g)^+ d\mu$  and  $\int (f+g)^- d\mu$  are finite, then (3.83) is equivalent to

$$\int (f+g)^+ d\mu - \int (f+g)^- d\mu = \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu,$$

which in turn is equivalent to

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu.$$

### ***Proof of Corollary 3.38***

Because  $|f| = f^+ + f^-$ , this proposition immediately follows from the definition of integrability (see Def. 3.28) and linearity of the integral [see Eq. (3.34)].

### ***Proof of Lemma 3.40***

If  $f$  is  $\mathcal{C}$ -measurable, then  $f^+$  and  $f^-$  are  $\mathcal{C}$ -measurable as well (see Th. 2.66). Furthermore,  $f^+$  and  $f^-$  can be represented as limits of increasing sequences of nonnegative step functions on  $(\Omega, \mathcal{C})$  [see Th. 3.19 (i)]. Hence, according to Equations (3.22) and (3.4), the values of the integrals  $\int f^+ d\mu$  and  $\int f^- d\mu$  only depend on the values  $\mu(A)$ ,  $A \in \mathcal{C}$ . Therefore, if  $\mu$  and  $\nu$  are identical on  $\mathcal{C}$ , then  $\int f d\mu = \int f d\nu$ , for all  $\mathcal{C}$ -measurable functions that are nonnegative or  $\mu$ -integrable.

### ***Proof of Lemma 3.41***

Define

$$A_+ := \{\omega \in \Omega: f(\omega) = \infty\} \in \mathcal{A} \quad \text{and} \quad A_- := \{\omega \in \Omega: f(\omega) = -\infty\} \in \mathcal{A}.$$

If  $\mu(A_+) > 0$ , then define the increasing sequence  $g_n: \Omega \rightarrow [0, \infty)$ ,  $n \in \mathbb{N}$ , by  $g_n = n \cdot 1_{A_+}$ . Because  $1_{A_+} \cdot f^+ = \lim_{n \rightarrow \infty} g_n$ ,

$$\begin{aligned}
\int f^+ d\mu &= \int 1_{A_+} \cdot f^+ d\mu + \int 1_{\Omega \setminus A_+} \cdot f^+ d\mu && [(3.36)] \\
&= \lim_{n \rightarrow \infty} \int g_n d\mu + \int 1_{\Omega \setminus A_+} \cdot f^+ d\mu && [\text{Def. 3.24}] \\
&= \lim_{n \rightarrow \infty} \int n \cdot 1_{A_+} d\mu + \int 1_{\Omega \setminus A_+} \cdot f^+ d\mu \\
&= \lim_{n \rightarrow \infty} n \cdot \mu(A_+) + \int 1_{\Omega \setminus A_+} \cdot f^+ d\mu && [(3.4)] \\
&= \infty.
\end{aligned}$$

Analogously we can prove that  $\int f^- d\mu = \infty$ , if  $\mu(A_-) > 0$  replacing  $f^+$  by  $f^-$  and  $A_+$  by  $A_-$ . Therefore, if  $f$  is quasi- $\mu$ -integrable and  $\int f^+ d\mu = \infty$ , then  $\int f^- d\mu$  is finite (see Def. 3.28) and  $\int f d\mu = \int f^+ d\mu - \int f^- d\mu = \infty$ . This proves (3.38). Analogously, if  $f$  is quasi- $\mu$ -integrable and  $\int f^- d\mu = \infty$ , then  $\int f^+ d\mu$  is finite (see Def. 3.28) and  $\int f d\mu = \int f^+ d\mu - \int f^- d\mu = -\infty$ , which proves (3.39). Finally, if  $\mu(A_+ \cup A_-) > 0$  and hence,  $\mu(A_+) > 0$  or  $\mu(A_-) > 0$ , then, according to (3.38) and (3.39),  $\int f d\mu$  is not defined or not finite. Thus, by contraposition, if  $f$  is  $\mu$ -integrable, then  $\mu(A_+ \cup A_-) = 0$ , i. e.,  $f$  is real-valued,  $\mu$ -almost everywhere.

#### **Proof of Lemma 3.44**

If  $f(\omega) > 0$ , for all  $\omega \in A$ , then  $1_A \cdot f: (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  is a nonnegative measurable function (see Th. 2.57). Hence,  $\int 1_A \cdot f d\mu \geq 0$  (see Defs. 3.24 and 3.10). Because  $\mu(\{\omega \in \Omega: (1_A \cdot f(\omega) > 0)\}) = \mu(A)$ , the assumption  $\mu(A) > 0$  implies that  $1_A \cdot f \stackrel{\mu}{=} 0$  does *not hold*. Therefore, according to Equation (3.40),  $\int 1_A \cdot f d\mu \neq 0$ , and we can conclude:  $\int 1_A \cdot f d\mu > 0$ .

#### **Proof of Lemma 3.45**

If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  is measurable and  $A \in \mathcal{A}$  with  $\mu(A) = 0$ , then  $1_A \cdot f$  is measurable (see Th. 2.57) and  $1_A \cdot f \stackrel{\mu}{=} 0$  (see Exercise 3-9). This implies  $(1_A \cdot f)^+ \stackrel{\mu}{=} 0$  and  $(1_A \cdot f)^- \stackrel{\mu}{=} 0$ . Now Equation (3.40) yields  $\int (1_A \cdot f)^+ d\mu = 0$  and  $\int (1_A \cdot f)^- d\mu = 0$ . Hence,  $\int 1_A \cdot f d\mu$  exists (see Def. 3.28) and Equation (3.27) implies

$$\int 1_A \cdot f d\mu = \int (1_A \cdot f)^+ d\mu - \int (1_A \cdot f)^- d\mu = 0.$$

#### **Proof of Lemma 3.47**

Define  $A := \{\omega \in \Omega: f(\omega) \neq g(\omega)\}$  and note that  $A \in \mathcal{A}$  [see Rem. 2.67 (c)]. Therefore,  $f \stackrel{\mu}{=} g$  implies  $\mu(A) = 0$ . Hence,

$$\begin{aligned}
\int f d\mu &= \int_{\Omega \setminus A} f d\mu && [(3.43)] \\
&= \int_{\Omega \setminus A} g d\mu && [\text{Def. of } A] \\
&= \int g d\mu. && [(3.43)]
\end{aligned}$$

**Proof of Theorem 3.48**

(a)

$f, g$  are  $\mu$ -integrable and  $f \stackrel{\mu}{=} g$

$$\Rightarrow \forall A \in \mathcal{A}: 1_A \cdot f \stackrel{\mu}{=} 1_A \cdot g \text{ and } 1_A \cdot f, 1_A \cdot g \text{ are } \mu\text{-integrable} \quad [(2.33), (3.29)]$$

$$\Rightarrow \forall A \in \mathcal{A}: \int_A f d\mu = \int_A g d\mu. \quad [(3.44)]$$

(b) If  $f, g$  are  $\mu$ -integrable, then  $f, g$  are real-valued  $\mu$ -a.e. (see Lemma 3.41). Hence, for

$$B := \{\omega \in \Omega: f(\omega) \in \{-\infty, \infty\}\} \cup \{\omega \in \Omega: g(\omega) \in \{-\infty, \infty\}\},$$

$B \in \mathcal{A}$  and  $\mu(B) = 0$ . Now define

$$A_{>} := \{\omega \in \Omega: f(\omega) > g(\omega)\} \quad \text{and} \quad A_{<} := \{\omega \in \Omega: f(\omega) < g(\omega)\}.$$

According to Remark 2.67 (b) and (a),  $A_{>} \in \mathcal{A}$  and  $A_{<} \in \mathcal{A}$ . Then

$$\begin{aligned}
\forall A \in \mathcal{A}: \int_A f d\mu &= \int_A g d\mu \\
\Rightarrow \forall A \in \mathcal{A}: \int_{A \cap B^c} f d\mu &= \int_{A \cap B^c} g d\mu && [(3.43)] \\
\Rightarrow \forall A \in \mathcal{A}: \int_{A \cap B^c} (f - g) d\mu &= 0 && [f, g \text{ } \mu\text{-integrable, (3.34)}] \\
\Rightarrow \int_{A_{>} \cap B^c} (f - g) d\mu &= 0 \quad \text{and} \quad \int_{A_{<} \cap B^c} (f - g) d\mu = 0 && [A_{>}, A_{<} \in \mathcal{A}] \\
\Rightarrow \mu(A_{>} \cap B^c) &= 0 \text{ and } \mu(A_{<} \cap B^c) = 0 && [(3.41)] \\
\Rightarrow \mu(A_{>} \cup A_{<}) &= \mu((A_{>} \cap B^c) \cup (A_{>} \cap B) \cup (A_{<} \cap B^c) \cup (A_{<} \cap B)) = 0 && [\text{Box 1.1 (ii)}] \\
\Rightarrow f &\stackrel{\mu}{=} g. && [\text{Def. 2.68}]
\end{aligned}$$

**Proof of Theorem 3.52**

(i) Define  $A := \{\omega \in \Omega: f(\omega) > g(\omega)\}$ . According to Remark 2.67 (b),  $A \in \mathcal{A}$ . Therefore,  $f \stackrel{\mu}{\leq} g$  implies  $\mu(A) = 0$ . Furthermore, define

$$A_{-\infty} := \{\omega \in \Omega \setminus A: f(\omega) = -\infty\} \quad \text{and} \quad A_{\infty} := \{\omega \in \Omega \setminus A: f(\omega) = g(\omega) = \infty\}.$$

If  $\mu(A_{-\infty}) > 0$ , then  $\int f d\mu = -\infty$  [see (3.39)]. Therefore,  $\int f d\mu \leq \int g d\mu$ . If  $\mu(A_{\infty}) > 0$ , then  $\int f d\mu = \int g d\mu = \infty$  [see (3.38)], and therefore  $\int f d\mu \leq \int g d\mu$ .

If  $\mu(A_{-\infty}) = \mu(A_{\infty}) = 0$ , then  $f, g$  are real-valued  $\mu$ -a.e. Now define  $B := (\Omega \setminus A) \setminus (A_{\infty} \cup A_{-\infty})$ , which implies  $\mu(\Omega \setminus B) = 0$ , and

$$\forall \omega \in B: f(\omega), g(\omega) \text{ are finite, } f(\omega) \leq g(\omega).$$

If  $\int (1_B f)^- d\mu = \infty$ , then  $\int_B f d\mu = -\infty$ . Hence,  $\int f d\mu \leq \int g d\mu$ . If  $\int (1_B f)^+ d\mu = \infty$ , then  $1_B f \leq 1_B g$  implies  $(1_B f)^+ \leq (1_B g)^+$  and  $\int (1_B g)^+ d\mu = \infty$  [see Eq. (3.24)]. Therefore,  $\int f d\mu = \int g d\mu = \infty$ , which implies  $\int f d\mu \leq \int g d\mu$ .

Now, if all  $(1_B f)^+, (1_B f)^-, (1_B g)^+, (1_B g)^-$  are  $\mu$ -integrable, then

$$\begin{aligned} \int g d\mu &= \int_B g d\mu && [(3.43)] \\ &= \int_B (f + g - f) d\mu && [1_B f, 1_B g \text{ are real-valued}] \\ &= \int_B f d\mu + \int_B (g - f) d\mu && [(3.34)] \\ &\geq \int_B f d\mu && [1_B(g - f) \geq 0, (3.24)] \\ &= \int f d\mu. && [(3.43)] \end{aligned}$$

(ii) Define

$$B := \{\omega \in \Omega: f(\omega) \in \{-\infty, \infty\}\} \cup \{\omega \in \Omega: g(\omega) \in \{-\infty, \infty\}\}.$$

If  $f, g$  are  $\mu$ -integrable, then Lemma 3.41 implies  $\mu(B) = 0$ . Furthermore, define  $A := \{\omega \in \Omega \setminus B: f(\omega) \geq g(\omega)\}$ . Then  $f \leq_\mu g$  implies  $\mu(A) = 0$ . Hence,  $\mu(A \cup B) = 0$  [see Box 1.1 (xi)]. Now,

$$\begin{aligned} \int g d\mu &= \int_{\Omega \setminus (A \cup B)} g d\mu && [(3.43)] \\ &= \int_{\Omega \setminus (A \cup B)} (f + g - f) d\mu && [1_{\Omega \setminus (A \cup B)} f, 1_{\Omega \setminus (A \cup B)} g \text{ are real-valued}] \\ &= \int_{\Omega \setminus (A \cup B)} f d\mu + \int_{\Omega \setminus (A \cup B)} (g - f) d\mu && [(3.34)] \\ &> \int_{\Omega \setminus (A \cup B)} f d\mu. && [1_{\Omega \setminus (A \cup B)}(g - f) > 0, \mu(\Omega \setminus (A \cup B)) > 0, \text{Lem. 3.44}] \end{aligned}$$

### **Proof of Corollary 3.59**

Assume that  $f: (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is measurable with a finite number of *positive* values  $\alpha_1, \dots, \alpha_m > 0$  and a finite number of *negative* values  $\alpha_{m+1}, \dots, \alpha_n < 0$ . By

convention, if  $m = n$ , then  $\sum_{i=m+1}^n \alpha_i 1_{A_i} = 0$ , and if  $m = 0$ , then  $\sum_{i=1}^m \alpha_i 1_{A_i} = 0$ . Then

$$f = \sum_{i=1}^n \alpha_i 1_{A_i} = \sum_{i=1}^m \alpha_i 1_{A_i} + \sum_{i=m+1}^n \alpha_i 1_{A_i}$$

and  $f^+ = \sum_{i=1}^m \alpha_i 1_{A_i}$  as well as  $f^- = -\sum_{i=m+1}^n \alpha_i 1_{A_i} = \sum_{i=m+1}^n -\alpha_i 1_{A_i}$ . Therefore,

$$\int f^+ d\mu = \sum_{i=1}^m \alpha_i \mu(A_i) \quad \text{and} \quad \int f^- d\mu = \sum_{i=m+1}^n -\alpha_i \mu(A_i),$$

and  $\int f^+ d\mu$  as well as  $\int f^- d\mu$  are finite if and only if  $\mu(A_i) < \infty$ , for all  $i = 1, \dots, n$ . Now,  $\mu(A_i) < \infty$ , for all  $i = 1, \dots, n$ , implies

$$\begin{aligned} \int f d\mu &= \int f^+ d\mu - \int f^- d\mu && \text{[Def. 3.28]} \\ &= \sum_{i=1}^m \alpha_i \mu(A_i) + \sum_{i=m+1}^n \alpha_i \mu(A_i) \\ &= \sum_{i=1}^n \alpha_i \mu(A_i) \\ &= \sum_{i=1}^n \alpha_i \mu_f(\{\alpha_i\}) && \text{[Def. 2.80, (2.10)].} \end{aligned}$$

Note that, in the last equation, we used the assumption that the  $\alpha_1, \dots, \alpha_n$  are pairwise different.

### ***Proof of Theorem 3.68***

(i) This is the proposition of Lemma 3.47.

(ii) This proposition is Theorem 3.48.

(iii) If  $f, g$  are  $\mu$ -integrable and nonnegative, then it suffices to show: (b)  $\Leftrightarrow$  (c) [see (ii)]. Now

$$\begin{aligned} \int_A f d\mu &= \int_A g d\mu, \quad \forall A \in \mathcal{A} \\ \Leftrightarrow \int 1_A f d\mu &= \int 1_A g d\mu, \quad \forall A \in \mathcal{A} && \text{[Def. 3.35]} \\ \Leftrightarrow \int 1_A df \circ \mu &= \int 1_A dg \circ \mu, \quad \forall A \in \mathcal{A} && \text{[Th. 3.67 (i)]} \\ \Leftrightarrow f \circ \mu(A) &= g \circ \mu(A), \quad \forall A \in \mathcal{A} && \text{[(3.9)]} \\ \Leftrightarrow f \circ \mu &= g \circ \mu. \end{aligned}$$

(iv) If  $f, g$  are  $\mu$ -integrable and nonnegative, then the equivalence of (a), (b), and (c) follows from (ii) and (iii). Hence, it suffices to show: (c)  $\Leftrightarrow$  (d). Because  $\mu$ -integrability of  $f$  and  $g$  implies that  $f \circ \mu$  and  $g \circ \mu$  are finite measures, applying Theorem 1.72 completes the proof.

**Proof of Corollary 3.77**

Note that

$$\forall (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2: 1_C(\omega_1, \omega_2) = 1_{C_{\omega_1}}(\omega_2) = 1_{C_{\omega_2}}(\omega_1). \quad (3.85)$$

Now,

$$\mu_1 \otimes \mu_2(C) = \int 1_C d(\mu_1 \otimes \mu_2) \quad [(3.9)]$$

$$= \int \int 1_C(\omega_1, \omega_2) \mu_2(d\omega_2) \mu_1(d\omega_1) \quad [(3.80)]$$

$$= \int \int 1_{C_{\omega_1}}(\omega_2) \mu_2(d\omega_2) \mu_1(d\omega_1) \quad [(3.85)]$$

$$= \int \mu_2(C_{\omega_1}) \mu_1(d\omega_1). \quad [(3.9)]$$

The proof of the second equation is analog.

**3.8 Exercises**

▷ **Exercise 3-1** Construct a representation of the identity function on  $\Omega = \{1, \dots, n\}$  as a weighted sum of indicators of elements of  $\mathcal{A}$ , where  $n \in \mathbb{N}$  and  $(\Omega, \mathcal{A})$  with  $\mathcal{A} = \mathcal{P}(\Omega)$ .

▷ **Exercise 3-2** Prove that, for every nonnegative step function, there exists a normal representation (see Rem. 3.8).

▷ **Exercise 3-3** Consider the measure space  $(\Omega, \mathcal{A}, \lambda)$ , where  $\lambda$  is the Lebesgue measure. Show that the number  $\sum_{i=1}^n \alpha_i \lambda(A_i)$  assigned to the nonnegative step function  $f$  defined in Example 3.7 is identical for the four specified representations of  $f$ , two of which are nonnormal representations.

▷ **Exercise 3-4** Let  $(\Omega, \mathcal{A})$  be a measurable space and let  $A \in \mathcal{A}$ . Show that if  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  is a normal representation of a nonnegative step function, then the product  $1_A f$  of the indicator  $1_A$  and  $f$  is also a normal representation of a nonnegative step function, and  $1_A f = \sum_{i=1}^n \alpha_i 1_{A \cap A_i}$ .

▷ **Exercise 3-5** Prove Equation (3.8).

▷ **Exercise 3-6** Compute the integral of the identity mapping  $id: \Omega \rightarrow \Omega$  with respect to the counting measure  $\mu_{\#}$  on  $\mathcal{P}(\Omega)$ , where  $\Omega = \{1, \dots, n\}$ . Then look at it for  $n = 5$ .

▷ **Exercise 3-7** Compute the integrals  $\int f_1 d\lambda$  and  $\int f_2 d\lambda$  of the functions  $f_1$  and  $f_2$  defined in Example 3.17.

▷ **Exercise 3-8** Prove the propositions of Example 3.39.

▷ **Exercise 3-9** Let  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be measurable and  $A \in \mathcal{A}$ . Show that  $1_A f \stackrel{\mu}{=} 0$  if  $\mu(A) = 0$ .

▷ **Exercise 3-10** Prove Equation (3.52).

▷ **Exercise 3-11** Compute the integral of the function  $f(x) = 10 - x^2$  considered in Example 3.64.

▷ **Exercise 3-12** Prove the proposition of Remark 3.71.

## Solutions

▷ **Solution 3-1** The identity function  $id: \Omega \rightarrow \Omega$  on  $\Omega = \{1, \dots, n\}$  is defined by

$$id(i) = i, \quad \forall i \in \Omega.$$

According to Example 2.9 it is  $(\mathcal{A}, \mathcal{A}_0)$ -measurable for all  $\sigma$ -algebras  $\mathcal{A}_0 \subset \mathcal{A}$ . Now consider the set  $\{1, \dots, n\}$  of values of  $id$  and the partition  $\{\{1\}, \dots, \{n\}\}$  of  $\Omega$ . Then

$$id = \sum_{i=1}^n i \cdot 1_{\{i\}} = 1 \cdot 1_{\{1\}} + \dots + n \cdot 1_{\{n\}}.$$

▷ **Solution 3-2** Let  $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$  be a nonnegative step function, with  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$ . Define, for all nonempty  $J \subset \{1, \dots, n\}$ ,

$$B_J := \left( \bigcap_{i \in J} A_i \right) \cap \left( \bigcap_{i \notin J} A_i^c \right).$$

These are  $2^n - 1$  sets, where several of them may be empty, and all are pairwise disjoint. Then

$$f = \sum_{J: B_J \neq \emptyset} \left( \sum_{i \in J} \alpha_i \right) \cdot 1_{B_J}$$

is a normal representation of  $f$ .

▷ **Solution 3-3** We compute the sum for all four representations of  $f$ . The first one is:

$$\begin{aligned} \sum_{i=1}^4 \alpha_i \lambda(A_i) &= 2 \cdot (1-0) + 5 \cdot (2-1) + 4 \cdot (3-2) + 1 \cdot (4-3) \\ &= 2 \cdot 1 + 5 \cdot 1 + 4 \cdot 1 + 1 \cdot 1 = 12. \end{aligned}$$

The sum for the second representation of  $f$  is:

$$\begin{aligned} \sum_{i=1}^5 \gamma_i \lambda(C_i) &= 2 \cdot (.5-0) + 2 \cdot (1-.5) + 5 \cdot (2-1) + 4 \cdot (3-2) + 1 \cdot (4-3) \\ &= 2 \cdot .5 + 2 \cdot .5 + 5 \cdot 1 + 4 \cdot 1 + 1 \cdot 1 = 12. \end{aligned}$$

The sum for the third representation of  $f$ , which is nonnormal, is:

$$\begin{aligned} \sum_{i=1}^3 \beta_i \lambda(B_i) &= 2 \cdot (2-0) + 3 \cdot (3-1) + 1 \cdot (4-2) \\ &= 2 \cdot 2 + 3 \cdot 2 + 1 \cdot 2 = 12. \end{aligned}$$

The sum for the fourth representation of  $f$ , which is also nonnormal, is:

$$\begin{aligned} \sum_{i=1}^4 \delta_i \lambda(D_i) &= 1 \cdot (4-0) + 1 \cdot (3-0) + 2 \cdot (3-1) + 1 \cdot (2-1) \\ &= 1 \cdot 4 + 1 \cdot 3 + 2 \cdot 2 + 1 \cdot 1 = 12. \end{aligned}$$

Obviously, all four sums are identical.

▷ **Solution 3-4** Let  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$ , where  $A_1, \dots, A_n \in \mathcal{A}$ . This implies  $A \cap A_1, \dots, A \cap A_n \in \mathcal{A}$ , for  $A \in \mathcal{A}$ . Therefore, and because of  $1_A \cdot f = \sum_{i=1}^n \alpha_i 1_{A \cap A_i}$ , the function  $1_A \cdot f$  is a non-negative step function. If  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  is a normal representation, then  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , which implies  $(A \cap A_i) \cap (A \cap A_j) = A \cap (A_i \cap A_j) = A \cap \emptyset = \emptyset$ , for  $i \neq j$ . Therefore,  $1_A \cdot f = \sum_{i=1}^n \alpha_i 1_{A \cap A_i}$  is a normal representation as well.

▷ **Solution 3-5**  $\int_A \alpha d\mu = \int \alpha 1_A d\mu = \alpha \mu(A)$  [see Eq. (3.4)].

▷ **Solution 3-6** Consider the elements  $\{1\}, \dots, \{n\}$  of  $\mathcal{A} = \mathcal{P}(\Omega)$  and  $id = \sum_{\omega=1}^n \omega \cdot 1_{\{\omega\}}$ . According to Definition 3.10,

$$\begin{aligned} \int id d\mu_{\#} &= \sum_{\omega=1}^n \omega \cdot \mu_{\#}(\{\omega\}) \\ &= 1 \cdot \mu_{\#}(\{1\}) + 2 \cdot \mu_{\#}(\{2\}) + \dots + n \cdot \mu_{\#}(\{n\}) \\ &= \sum_{i=1}^n i = \frac{n(n+1)}{2} \end{aligned} \quad (3.86)$$

is the integral of  $id$  over  $\Omega$  with respect to the measure  $\mu_{\#}$ . Hence, in this example, the integral  $\int id d\mu_{\#}$  is the sum over all elements in  $\Omega$ . For  $n = 5$  this formula yields  $\int id d\mu_{\#} = 15$ .

▷ **Solution 3-7** In Example 3.17 we considered the measure space  $(\mathbb{R}, \mathcal{B}, \lambda)$ , where  $\lambda$  is the Lebesgue (or length) measure on  $\mathcal{B}$ . Remember, the Lebesgue measure satisfies

$$\lambda(]a, b]) = \lambda([a, b]) = b - a,$$

for  $a < b$  [see Eq. (1.54)]. We also considered  $f_1 = \alpha_1 1_{A_1}$  with  $A_1 = [0, (1 - .50)^{1/2}]$  and  $\alpha_1 = .50$ . Hence,  $f_1 = .50 \cdot 1_{A_1}$ . Therefore,

$$\int f_1 d\lambda = \alpha_1 \cdot \lambda(A_1) = .50 \cdot \lambda(A_1) = .50 \cdot (1 - .50)^{1/2} = .50 \cdot .50^{1/2} \approx .3536.$$

This is the area shaded in the left part of Figure 3.3.

Similarly, in Example 3.17 we also considered  $f_2 = \sum_{i=1}^3 \beta_i 1_{B_i}$  with the three intervals

$$B_1 = [0, (1 - .75)^{1/2}], \quad B_2 = ](1 - .75)^{1/2}, (1 - .50)^{1/2}], \quad B_3 = ](1 - .50)^{1/2}, (1 - .25)^{1/2}].$$

Again note that  $B_1, B_2, B_3$  is a sequence of elements of  $\mathcal{A}$ . Furthermore,  $f_2 = \sum_{i=1}^3 \beta_i 1_{B_i}$  with  $\beta_1 = .75$ ,  $\beta_2 = .50$ , and  $\beta_3 = .25$ . Hence, the integral of  $f_2 = \sum_{i=1}^3 \beta_i 1_{B_i}$  with respect to  $\lambda$  is

$$\begin{aligned} \int f_2 d\lambda &= \sum_{i=1}^3 \beta_i \cdot \lambda(B_i) = .75 \cdot \lambda(B_1) + .50 \cdot \lambda(B_2) + .25 \cdot \lambda(B_3) \\ &= .75 \cdot [(1 - .75)^{1/2}] + .50 \cdot [(1 - .50)^{1/2} - (1 - .75)^{1/2}] + .25 \cdot [(1 - .25)^{1/2} - (1 - .50)^{1/2}] \\ &\approx .75 \cdot .50 + .50 \cdot .2071 + .25 \cdot .1589 \approx 0.3750 + .1036 + .0397 = .5183. \end{aligned}$$

This is the area shaded in the middle part of Figure 3.3. The integral of  $f_3 = \sum_{i=1}^7 \gamma_i 1_{C_i}$  can be computed correspondingly. It is the area shaded in the right part of Figure 3.3.

▷ **Solution 3-8** If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is  $\mu$ -integrable and  $A, B \subset \mathcal{A}$ , then  $1_{A \cup B} f$  is  $\mu$ -integrable (see Lemma 3.33) and

$$\int_{A \cup B} f \, d\mu = \int 1_{A \cup B} f \, d\mu \quad [(3.30)]$$

$$= \int (1_A + 1_B - 1_{A \cap B}) f \, d\mu \quad [(1.34)]$$

$$= \int (1_A f + 1_B f - 1_{A \cap B} f) \, d\mu$$

$$= \int 1_A f \, d\mu + \int 1_B f \, d\mu - \int 1_{A \cap B} f \, d\mu \quad [(3.34)]$$

$$= \int_A f \, d\mu + \int_B f \, d\mu - \int_{A \cap B} f \, d\mu. \quad [(3.30)]$$

If  $A \cap B = \emptyset$  and  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is quasi- $\mu$ -integrable, then  $1_{A \cup B} \cdot f = 1_A \cdot f + 1_B \cdot f$ , and these three functions are quasi- $\mu$ -integrable (see Lem. 3.33). If  $\int f^+ \, d\mu$  is finite, then  $\int 1_{A \cup B} \cdot f^+ \, d\mu$ ,  $\int 1_A \cdot f^+ \, d\mu$ , and  $\int 1_B \cdot f^+ \, d\mu$  are finite as well (see Lem. 3.33). If  $\int f^+ \, d\mu$  is infinite, then quasi- $\mu$ -integrability of  $f$  implies that  $\int f^- \, d\mu$  is finite and, according to Lemma 3.33, also the integrals  $\int 1_{A \cup B} \cdot f^- \, d\mu$ ,  $\int 1_A \cdot f^- \, d\mu$ , and  $\int 1_B \cdot f^- \, d\mu$ . Hence, in both cases

$$\int 1_{A \cup B} \cdot f \, d\mu = \int 1_{A \cup B} \cdot f^+ \, d\mu - \int 1_{A \cup B} \cdot f^- \, d\mu \quad [(3.27)]$$

$$= \int 1_A \cdot f^+ \, d\mu + \int 1_B \cdot f^+ \, d\mu - \left( \int 1_A \cdot f^- \, d\mu + \int 1_B \cdot f^- \, d\mu \right) \quad [(3.34)]$$

$$= \left( \int 1_A \cdot f^+ \, d\mu - \int 1_A \cdot f^- \, d\mu \right) + \left( \int 1_B \cdot f^+ \, d\mu - \int 1_B \cdot f^- \, d\mu \right)$$

$$= \int 1_A \cdot f \, d\mu + \int 1_B \cdot f \, d\mu. \quad [(3.27)]$$

▷ **Solution 3-9**

$$1_A(\omega) \cdot f(\omega) = \begin{cases} 0, & \text{if } \omega \notin A \\ f(\omega), & \text{if } \omega \in A. \end{cases}$$

Therefore,  $\{\omega \in \Omega: 1_A(\omega) \cdot f(\omega) \neq 0\} \subset A$ . Hence,  $\mu(\{\omega \in \Omega: 1_A(\omega) \cdot f(\omega) \neq 0\}) \leq \mu(A) = 0$ .

▷ **Solution 3-10** Let  $\alpha \geq 0$  and remember that the measure  $\alpha\mu$  on  $(\Omega, \mathcal{A})$  is defined by  $(\alpha\mu)(A) = \alpha\mu(A)$  for all  $A \in \mathcal{A}$ . The proof is conducted in three steps: (a) for a nonnegative step function, (b) for a nonnegative numerical measurable function, and (c) for an  $\mu$ -integrable numerical function (see Rem. 3.30).

(a) If  $f$  is a nonnegative step function and  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  a normal representation, then

$$\int f \, d(\alpha\mu) = \sum_{i=1}^n \alpha_i (\alpha\mu)(A_i) = \alpha \sum_{i=1}^n \alpha_i \mu(A_i) = \alpha \int f \, d\mu.$$

(b) If  $f$  is a nonnegative numerical measurable function and  $f_i, i \in \mathbb{N}$ , is an increasing sequence of nonnegative step functions with  $\lim_{i \rightarrow \infty} f_i = f$ , then

$$\begin{aligned} \int f \, d(\alpha\mu) &= \lim_{i \rightarrow \infty} \int f_i \, d(\alpha\mu) \\ &= \lim_{i \rightarrow \infty} \alpha \int f_i \, d\mu && [(a)] \\ &= \alpha \lim_{i \rightarrow \infty} \int f_i \, d\mu = \alpha \int f \, d\mu. \end{aligned}$$

(c) If  $f$  an  $\mu$ -integrable numerical function, then

$$\begin{aligned}
\int f d(\alpha\mu) &= \int f^+ d(\alpha\mu) - \int f^- d(\alpha\mu) \\
&= \alpha \int f^+ d\mu - \alpha \int f^- d\mu \quad [(b)] \\
&= \alpha \int f d\mu.
\end{aligned}$$

▷ **Solution 3-11** Because the derivative  $g'(x)$  of a function  $g(x) = \alpha + \beta x + \gamma x^n$ ,  $\alpha, \beta, \gamma \in \mathbb{R}$ ,  $n \in \mathbb{N}$ , is  $g'(x) = \beta + \gamma n x^{n-1}$ , the indefinite integral of  $f(x)$  is

$$\int f(x) dx = F(x) = 10x - \frac{x^3}{3} + c, \quad c \in \mathbb{R},$$

and therefore,

$$\int_a^b f(x) dx = F(x) \Big|_a^b := F(b) - F(a) = \left(10x - \frac{x^3}{3}\right) \Big|_a^b.$$

For  $a = -5$  and  $b = 5$ , this equation yields

$$\int_{-5}^5 f(x) dx = \left(10x - \frac{x^3}{3}\right) \Big|_{-5}^5 = \left(50 - \frac{125}{3}\right) - \left(-50 + \frac{125}{3}\right) = 100 - \frac{250}{3} = 16.\bar{6}.$$

Hence,

$$\int_{[-5,5]} f d\lambda = \int_{-5}^5 f(x) dx = 16.\bar{6}.$$

▷ **Solution 3-12** Let  $A \in \mathcal{A}$  with  $\mu(A) = 0$  and let  $f$  be a density  $f$  of  $\nu$  with respect to  $\mu$ , i. e.,  $\nu = f \circ \mu$ . Then

$$\begin{aligned}
\nu(A) &= \int 1_A d\nu = \int 1_A d f \circ \mu && [(3.8), \nu = f \circ \mu] \\
&= \int 1_A \cdot f d\mu = 0. && [(3.72), (3.42)]
\end{aligned}$$

Hence,  $\nu \ll_{\mathcal{A}} \mu$ .

**Part II**  
**Probability, Random Variable and its**  
**Distribution**



## Chapter 4

# Probability Measure

In chapter 1 we introduced the concept of a *measure*, treated various examples of measures, and some of their properties. In this chapter, we turn to a special class of examples, called *probability measures*. We start with the definition of a probability measure, then turn to conditional probabilities and the most important theorems related to conditional probability: the *multiplication rule*, the *theorem of total probability*, and *Bayes' Theorem*. Furthermore, we introduce the concept of a *conditional-probability measure*. Next, we define *independence of events* and of *independence of sets of events* with respect to a probability measure. A section on *conditional independence given an event* concludes this chapter.

### 4.1 Probability Measure and Probability Space

Now we introduce the concept of a *probability measure* as defined by Kolmogorov (1933/1977) (for the English version of this book see Kolmogorov, 1956). As we shall see, a probability measure is a special finite measure that is standardized.

#### 4.1.1 Definition

**Definition 4.1 (Probability Measure)**

Let  $(\Omega, \mathcal{A})$  be a measurable space. Then the function  $P: \mathcal{A} \rightarrow [0, 1]$  is called a *probability measure* on  $(\Omega, \mathcal{A})$ , if the following conditions hold:

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

**Remark 4.2 (Probability and Probability Space)** Let  $P$  be a probability measure on  $(\Omega, \mathcal{A})$ . Then the triple  $(\Omega, \mathcal{A}, P)$  is called a *probability space* and a value  $P(A)$  of  $P$  is called the *probability* of  $A$ .  $\triangleleft$

**Remark 4.3 (Elementary Event and Event)** Let  $(\Omega, \mathcal{A}, P)$  be a probability space. Then  $A \in \mathcal{A}$  is called an *event*, and a singleton  $\{\omega\}, \omega \in \Omega$ , is called an *elemen-*

*tary event*, if  $\{\omega\} \in \mathcal{A}$ . Note the distinction between an *outcome*  $\omega \in \Omega$  and an *elementary event*  $\{\omega\} \in \mathcal{A}$  (see Exercise 4-1). Also note that the term *event* is only used in the context of a probability space  $(\Omega, \mathcal{A}, P)$ . Otherwise,  $A \in \mathcal{A}$  is called a *measurable set*.  $\triangleleft$

#### 4.1.2 Formal and Substantive Meaning of Probabilistic Terms

We distinguish between the mathematical or formal meaning of probabilistic terms and the meaning of these terms if used in an application of probability theory to a concrete real-world phenomenon. Often such a real-world phenomenon is a random experiment such as flipping a coin. In such a case, the set  $\Omega$  represents the set of possible outcomes of the random experiment considered and, in this sense,  $\Omega$  has a real-world interpretation. In other words, in this case  $\Omega$  is not only an abstract set anymore. This *real-world meaning* or *substantive meaning* is additional to their mathematical or *formal* meaning, namely being elements of the (abstract) set  $\Omega$ .

The terms ‘probability measure’, ‘probability of an event’ etc. hint at an important area of application of probability theory: real-world phenomena called random experiments. However, formally speaking, a probability measure is simply a label for a measure on a measurable space  $(\Omega, \mathcal{A})$  satisfying  $P(\Omega) = 1$ . If  $\Omega$  is not just an abstract set but represents a concrete random experiment, then the probability of an event  $A \in \mathcal{A}$  corresponds to the common language meaning of the term probability.

**Remark 4.4 (No Time Order Between Events)** The intuitive concept of an event often implies that events are ordered with respect to time. That is, in common language an ‘event’ is prior, simultaneous, or posterior to another ‘event’. In contrast, events as defined in probability theory are not necessarily ordered with respect to time. However, time order between events and sets of events can be introduced with respect to a filtration (see Def. 4.17, Fig. 4.1 and Example 4.19).  $\triangleleft$

**Remark 4.5 (A Priori Perspective)** If we apply probability theory to real-world phenomena, then we consider random experiments from the *a priori perspective*. Hence, the possible outcomes of a random experiment and events are considered *before* they happen. Even if an event already happened, we do *as if* it did not happen when we talk about its probability (see also Rem. 4.13).  $\triangleleft$

#### 4.1.3 Properties of a Probability Measure

Comparing conditions (a) to (c) of the definition of a probability measure to the conditions defining a measure (see Def. 1.43) shows that (b) and (c) are identical; only condition (a) differs. However,  $P(\Omega) = 1$  implies  $P(\emptyset) = 0$ , because  $\sigma$ -additivity of  $P$  yields

$$P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \cup \dots) = P(\Omega) + \sum_{i=1}^{\infty} P(\emptyset).$$

**Box 4.1 Rules of Computation for Probabilities**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space.

If  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad (\sigma\text{-additivity}) \quad (\text{i})$$

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i), \quad \forall n \in \mathbf{N}. \quad (\text{finite additivity}) \quad (\text{ii})$$

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

$$P(A) = P(A \cap B) + P(A \setminus B) \quad (\text{iii})$$

$$P(A^c) = 1 - P(A) \quad (\text{iv})$$

$$P(A) \leq P(B), \quad \text{if } A \subset B \quad (\text{monotonicity}) \quad (\text{v})$$

$$P(A \setminus B) = P(A) - P(A \cap B) \quad (\text{vi})$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (\text{vii})$$

$$P(A) = 1 \Rightarrow P(A \cap B) = P(B) \quad (\text{viii})$$

$$P(A) = 0 \Rightarrow P(A \cup B) = P(B). \quad (\text{ix})$$

Let  $A \in \mathcal{A}$  and let  $\Omega_0 \subset \Omega$  be finite or countable with  $P(\Omega_0) = 1$ .

If, for all  $\omega \in \Omega_0$ ,  $\{\omega\} \in \mathcal{A}$ , then

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

If  $A_1, A_2, \dots \in \mathcal{A}$ , then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i). \quad (\sigma\text{-subadditivity}) \quad (\text{xi})$$

Therefore,  $\sum_{i=1}^{\infty} P(\emptyset) = P(\Omega) - P(\Omega) = 0$  and this yields  $P(\emptyset) = 0$ . This proves the following corollary:

**Corollary 4.6 (A Probability Measure is a Measure)**

A measure  $P$  on  $(\Omega, \mathcal{A})$  is a probability measure on  $\mathcal{A}$  if and only if  $P(\Omega) = 1$ .

A direct implication of this corollary is that all rules of computation for a measure (see Box 1.1) also hold for a probability measure. For convenience, these

rules are explicitly formulated for probability measures in Box 4.1 using the additional property  $P(\Omega) = 1$ .

**Remark 4.7 (Distribution vs. Probability Measure)** A probability measure on  $(\Omega, \mathcal{A})$  is also called a *distribution on  $(\Omega, \mathcal{A})$* . Although this term is preferably used in the context of a random variable (see Def. 5.3), the term ‘distribution’ is well-defined without referring to a random variable.  $\triangleleft$

#### 4.1.4 Examples

**Example 4.8 (Continuous Uniform Distribution)** Let  $\mathcal{B}_2$  denote the Borel  $\sigma$ -algebra on  $\mathbb{R}^2$  and consider a probability space  $(\Omega, \mathcal{A}, P)$ , where  $\Omega \in \mathcal{B}_2$ ,  $\mathcal{A} = \mathcal{B}_2|_{\Omega} := \{\Omega \cap A : A \in \mathcal{B}_2\}$  is the trace of  $\mathcal{B}_2$  in  $\Omega$  (see Example 1.10). Furthermore, let  $\lambda_2$  denote the Lebesgue measure on  $(\mathbb{R}^2, \mathcal{B}_2)$ , assume  $0 < \lambda_2(\Omega) < \infty$ , and define

$$P(A) = \frac{\lambda_2(A)}{\lambda_2(\Omega)}, \quad \forall A \in \mathcal{A}. \quad (4.1)$$

Then  $P$  is the *continuous uniform distribution* over  $\Omega$ . The relative size of the set  $A \in \mathcal{A}$  represents the probability  $P(A)$ , and Figure 1.1 (p. 4) can be used to illustrate some of its properties, e. g., Rules (iii) to (ix) of Box 4.1. This example will be generalized in section 8.2.1.  $\triangleleft$

**Example 4.9 (Joe and Ann With Random Assignment – continued)** In Example 1.9, we specified the set

$$\Omega = \{(Joe, no, -), (Joe, no, +), \dots, (Ann, yes, +)\},$$

which is also presented in the first column of Table 4.1. In this table we also specify the probability measure  $P$  on  $\mathcal{A} = \mathcal{P}(\Omega)$  by the probabilities of the eight elementary events  $\{\omega\} \in \Omega$ . Except for the empty set, which has probability  $P(\emptyset) = 0$ , all  $2^8 = 256$  elements of  $\mathcal{A}$  are either one of the eight elementary events  $\{(Joe, no, -)\}, \{(Joe, no, +)\}, \dots, \{(Ann, yes, +)\}$  or a union of some of these elementary events. Note that elementary events are always pairwise *disjoint*, i. e.,  $\{\omega_i\} \cap \{\omega_j\} = \emptyset$ , if  $\omega_i \neq \omega_j$ . Therefore, the probabilities of their unions can easily be computed using finite additivity of the probability measure [see Rule (ii) of Box 4.1]. In order to illustrate this point, consider the event that *Joe is drawn*,

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

and the event that *the drawn person is successful*,

$$C = \{(Joe, no, +), (Joe, yes, +)\}, \{(Ann, no, +), (Ann, yes, +)\}.$$

The event  $A$  has the probability

$$\begin{aligned} P(A) &= P[\{(Joe, no, -)\}] + P[\{(Joe, no, +)\}] + P[\{(Joe, yes, -)\}] + P[\{(Joe, yes, +)\}] \\ &= .09 + .21 + .04 + .16 = .50. \end{aligned}$$

**Table 4.1.** Joe and Ann With Random Assignment: Probability Measures

Elements of $\Omega$			Probability measures			
Unit	Treatment	Success	$P_{\{\omega\}}$	$P^B_{\{\omega\}}$	$P^A_{\{\omega\}}$	$P^{A^c}_{\{\omega\}}$
(Joe, no, -)			.09	0	.18	0
(Joe, no, +)			.21	0	.42	0
(Joe, yes, -)			.04	.10	.08	0
(Joe, yes, +)			.16	.40	.32	0
(Ann, no, -)			.24	0	0	.48
(Ann, no, +)			.06	0	0	.12
(Ann, yes, -)			.12	.30	0	.24
(Ann, yes, +)			.08	.20	0	.16

Note.  $P, P^B, P^A,$  and  $P^{A^c}$  are probability measures on  $(\Omega, \mathcal{A})$ .

Similarly, the event  $C$  has the probability

$$P(C) = P[\{(Joe, no, +)\}] + P[\{(Joe, yes, +)\}] + P[\{(Ann, no, +)\}] + P[\{(Ann, yes, +)\}] = .21 + .16 + .06 + .08 = .51,$$

and the event *Joe is drawn and is successful*,  $A \cap C = \{(Joe, no, +), (Joe, yes, +)\}$ , has the probability

$$P(A \cap C) = P[\{(Joe, no, +)\}] + P[\{(Joe, yes, +)\}] = .21 + .16 = .37.$$

The probability measures specified in the last three columns of Table 4.1 will be treated in Examples 4.34 and 4.35. ◁

**Remark 4.10 (Other Examples)** In section 8.1 probability measures on the measurable space  $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$  are considered, e. g., the *binomial distribution* (see Def. 8.7), the *Poisson distribution* (see Def. 8.14), and the *geometric distribution* (see Def. 8.20). In all these examples, a probability measure on  $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$  is uniquely defined, if the probabilities of the elementary events  $\{x\}$  are determined for all  $x \in \mathbb{N}_0$  [see Box 4.1 (x) for  $\Omega_0 = \mathbb{N}_0$ ].

The example of the Poisson distribution shows that, even for the countably infinite set  $\mathbb{N}_0$ , there are probability measures  $\mathcal{P}_\lambda$  on  $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$  with  $\mathcal{P}_\lambda(\{x\}) > 0$ , for all  $x \in \mathbb{N}_0$ , and

$$\mathcal{P}_\lambda(\mathbb{N}_0) = \sum_{x \in \mathbb{N}_0} \mathcal{P}_\lambda(\{x\}) = \sum_{x=0}^{\infty} \mathcal{P}_\lambda(\{x\}) = 1.$$

For another example see Exercise 4-2. ◁

**Example 4.11 (Finite Mixture of Probability Measures)** In Example 1.62 we already noted that the weighted sum of measures on a measurable space  $(\Omega, \mathcal{A})$  is again a measure on  $(\Omega, \mathcal{A})$ . With an additional assumption this also applies to probability measures. More precisely, if  $P_1, \dots, P_n$  are probability measures on  $(\Omega, \mathcal{A})$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, n$ , and we additionally assume  $\sum_{i=1}^n \alpha_i = 1$ , then  $\sum_{i=1}^n \alpha_i P_i$  is again a probability measure on  $(\Omega, \mathcal{A})$ . It is called a *finite mixture* of  $P_1, \dots, P_n$ . Such a finite mixture of probability measures will be illustrated by Example 4.35 using conditional-probability measures.  $\triangleleft$

## 4.2 Conditional Probability

Conditional probabilities can be used to describe *dependencies* between two events  $A, B \in \mathcal{A}$  with respect to a probability measure  $P$  on  $\mathcal{A}$ . In section 4.2.7 we will also use this concept in order to introduce the concept of a *conditional-probability measure*.

### 4.2.1 Definition

**Definition 4.12 (Conditional Probability)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space, let  $A, B \in \mathcal{A}$ , and let  $P(B) > 0$ . Then

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \quad (4.2)$$

is called the *conditional probability of A given B with respect to P*.

**Remark 4.13 (A Priori Perspective)** The conditional probability  $P(A|B)$  is the probability of the event  $A$  if it is known that the event  $B$  occurred. In order to compute  $P(A|B)$  according to Equation (4.2), we need the (unconditional) probability  $P(B)$ . Using  $P(B)$  we do as if  $B$  did not yet happen (see Rem. 4.5). The fact that  $B$  occurred is reflected by  $P(B|B) = 1$ .  $\triangleleft$

**Remark 4.14 (Continuous Uniform Distribution – continued)** In Example 4.8 we defined the continuous uniform distribution on  $(\Omega, \mathcal{A})$  by Equation (4.1). Using the area of the ellipses presented in Figure 1.1 (p. 4), the conditional probability  $P(A|B)$  corresponds to the area of the intersection  $A \cap B$  divided by the area of  $B$ .  $\triangleleft$

**Example 4.15 (Flipping a Coin two Times)** Consider the random experiment of flipping a coin two times, the measurable space  $(\Omega, \mathcal{A})$  of which is the same as in subsection 2.2.2 and the probability measure has been specified by Equation 2.43. The conditional probability  $P(B|A)$  that we flip *heads* in the second flip ( $B$ ) given that we flip *heads* in the first flip ( $A$ ) is  $1/2$ , which is equal to the *unconditional* probability  $P(B)$  of flipping *heads* in the second flip. In such a case the

two events  $A$  and  $B$  are *independent* (see section 4.3). Note that the conditional probability  $P(A|B)$  that we flip *heads* in the first flip ( $A$ ) given that we flip *heads* in the second flip ( $B$ ) is also equal to the unconditional probability  $P(A)$  of flipping *heads* in the first flip. This example shows that we may condition on events that occur later in time and that a conditional probability does not necessarily describe a causal dependence. Note, however, that conditional probabilities *can* be used to describe causal dependencies, provided that additional assumptions hold (see Examples 4.16 and 4.36).

As another example consider the event *flipping at least one heads* ( $A$ ) and the event *no heads are flipped in the first flip* ( $B$ ). In this case

$$P(A|B) = \frac{1}{2} \neq P(A) = \frac{3}{4},$$

and the two events are not independent (see section 4.3). ◁

**Example 4.16 (Joe and Ann With Random Assignment – continued)** Consider again Table 4.1 (p. 131), 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*. Furthermore, let

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

denote the event that the *drawn person is treated*. Then Equation (4.2) yields:

$$P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{P(\Omega_U \times \{yes\} \times \{+\})}{P(\Omega_U \times \{yes\} \times \Omega_Y)} = \frac{.16 + .08}{.04 + .16 + .12 + .08} = .60.$$

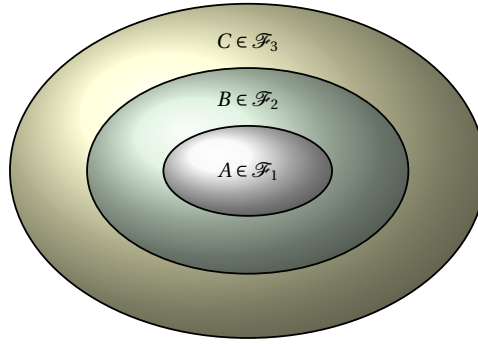
Conditioning on the event  $B^c$  that the *drawn person is not treated* yields

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

In this example, the difference  $P(C|B) - P(C|B^c) = .60 - .45$  *can* be used to evaluate the average effect of the treatment. This will be substantiated in more detail in Example 4.36. ◁

### 4.2.2 Filtration and Time Order Between Events and Sets of Events

As mentioned in Remark 4.4, the definition of an event in probability theory does not presume that there is a time order between events and set of events. However, in many applications of probability theory such a time order is important. In Example 4.16, for instance, it is crucial that the event  $C$  is *posterior* to the event  $B$ . Such a time order is formalized in the theory of stochastic processes (see, e. g., Bauer, 1996; Klenke, 2013) and in the theory of causal effects (see, e. g., Steyer, Mayer, & Fiege, 2014; Mayer, Thoemmes, Rose, Steyer, & West, 2014).



**Figure 4.1.** A filtration with  $T = \{1, 2, 3\}$

**Definition 4.17 (Filtration)**

Let  $(\Omega, \mathcal{A})$  be a measurable space and  $T \subset \mathbb{R}$ . A family  $(\mathcal{F}_t, t \in T)$  of sub- $\sigma$ -algebras  $\mathcal{F}_t$  of  $\mathcal{A}$  is called a *filtration* in  $\mathcal{A}$ , if  $\mathcal{F}_s \subset \mathcal{F}_t$  for all  $s, t \in T$  with  $s \leq t$ .

Referring to such a filtration, time order between events can be introduced as follows.

**Remark 4.18 (Event  $A$  is Prior, Simultaneous, Posterior to Event  $B$ )**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $(\mathcal{F}_t, t \in T)$  a filtration in  $\mathcal{A}$ , and  $A, B \in \mathcal{A}$ .

- (i) The event  $A$  is called *prior to  $B$*  (and  $B$  *posterior to  $A$* ) in  $(\mathcal{F}_t, t \in T)$ , if there is an  $s \in T$  such that  $A \in \mathcal{F}_s, B \notin \mathcal{F}_s$ , and a  $t \in T, t > s$ , such that  $B \in \mathcal{F}_t$ .
- (ii) The event  $A$  is called *simultaneous to  $B$*  in  $(\mathcal{F}_t, t \in T)$ , if there is a  $t \in T$  such that  $A, B \in \mathcal{F}_t$  and no  $s \in T, s < t$ , such that  $A \in \mathcal{F}_s$  or  $B \in \mathcal{F}_s$ .

◁

**Example 4.19 (Joe and Ann With Random Assignment – continued)** In the random experiment described by Table 4.1, the event  $A$  that Joe is drawn (see Example 4.9) is prior to the event

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

that the drawn person is treated, which itself is prior to the event  $C$  that the drawn person is successful. This time order in the real-world can formally be represented by the following filtration:

$$\mathcal{F}_1 := \sigma(\{A\}), \quad \mathcal{F}_2 := \sigma(\{A, B\}), \quad \mathcal{F}_3 := \sigma(\{A, B, C\}) = \mathcal{P}(\Omega),$$

using the concept of a  $\sigma$ -algebra generated by a set system (see Def. 1.13). With respect to the filtration  $(\mathcal{F}_t, t \in T)$ ,  $T = \{1, 2, 3\}$ , the event  $A$  is prior to  $B$ , because  $A \in \mathcal{F}_1, B \notin \mathcal{F}_1$ , but  $B \in \mathcal{F}_2$  (see Fig. 4.1, Rem. 4.18, and Exercise 4-3). ◁

**Remark 4.20 (Formal and Substantive Meaning of Time Order)** As noted in section 4.1.2, we distinguish between the mathematical or formal meaning of a probabilistic term on one side and the meaning of these terms if used in an application of probability theory to a concrete real-world phenomenon on the other side. This also applies to the terms *prior*, *simultaneous*, and *posterior* with respect to a filtration. In applications in which the elements of the set  $T$  represents time points, these terms do not only have a formal meaning that is specified by their mathematical definition, but also a substantive meaning: In the *real world*, an event  $A$  that is prior to  $B$  with respect to  $(\mathcal{F}_t, t \in T)$ , occurs before the event  $B$ . In other applications, the term *prior* may not express the actual substantive meaning. Of course, this applies to the terms *simultaneous* and *posterior* as well.  $\triangleleft$

### 4.2.3 Multiplication Rule

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

**Remark 4.21 (Multiplication Rule for Two and for Three Events)** For two events  $A_1$  and  $A_2$ , the multiplication rule is

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2 | A_1), \quad (4.3)$$

provided that  $P(A_1) > 0$ . This equation directly follows from the definition of the conditional probability  $P(A_2 | A_1)$  [see Eq. (4.2)]. For three events  $A_1$ ,  $A_2$ , and  $A_3$ , the multiplication rule is

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2), \quad (4.4)$$

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

$$P(A_3 | A_1 \cap A_2) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)}, \quad (4.5)$$

inserting Equation (4.3) for  $P(A_1 \cap A_2)$ , and solving the resulting equation for  $P(A_1 \cap A_2 \cap A_3)$ .  $\triangleleft$

For  $n$  events  $A_1, \dots, A_n$ , the multiplication rule is formulated in the following theorem.

#### Theorem 4.22 (Multiplication Rule)

Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $A_1, \dots, A_n \in \mathcal{A}$ , where  $2 \leq n \in \mathbb{N}$ . If  $P(\bigcap_{i=1}^{n-1} A_i) > 0$ , then

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \cdot \prod_{j=2}^n P\left(A_j \mid \bigcap_{i=1}^{j-1} A_i\right). \quad (4.6)$$

(Proof p. 148)

#### 4.2.4 Examples

**Example 4.23 (Joe and Ann With Random Assignment – continued)** Consider again the Example presented in Table 4.1 (p. 131) and let

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

denote the event that Joe is drawn,

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

the event that the drawn person is treated, and

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

the event that there is success, irrespective of the drawn person and treatment received. Then

$$A \cap B \cap C = \{(Joe, yes, +)\}$$

is the event that Joe is drawn, receives the treatment, and is successful. According to Equation (4.4), the probability of this event can be computed by

$$\begin{aligned} P(A \cap B \cap C) &= P(A) \cdot P(B|A) \cdot P(C|A \cap B) \\ &= (.09 + .21 + .04 + .16) \cdot \frac{.04 + .16}{.09 + .21 + .04 + .16} \cdot \frac{.16}{.04 + .16} \\ &= .50 \cdot .40 \cdot .80 = .16 \end{aligned}$$

(see Exercise 4-4). Of course, Equation (4.4) can also be applied to the other seven sets  $A \cap B \cap C^c$  to  $A^c \cap B^c \cap C^c$  in Figure 4.2. In this example,  $P(A \cap B \cap C) = P(\{(Joe, yes, +)\}) = .16$  is the probability of an elementary event (see Table 4.1).  $\triangleleft$

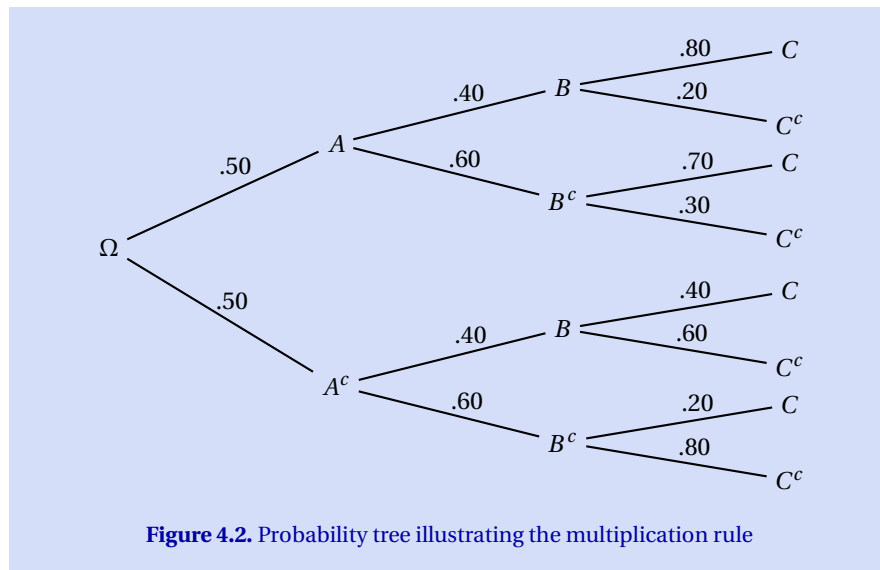
**Example 4.24 (Drawing Three Balls)** Consider drawing three balls without replacement from an urn containing two white balls and four black balls. Furthermore, let us consider the three events  $A_i$  to *draw a black ball at time  $i$* , where  $i = 1, 2, 3$ . According to Theorem 4.22, the probability of drawing three black balls is

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2),$$

where  $P(A_1) = 4/6$ ,  $P(A_2|A_1) = 3/5$ , and  $P(A_3|A_1 \cap A_2) = 2/4$ . Hence,

$$P(A_1 \cap A_2 \cap A_3) = \frac{4}{6} \cdot \frac{3}{5} \cdot \frac{2}{4} = \frac{24}{120} = \frac{1}{5}.$$

$\triangleleft$



#### 4.2.5 Theorem of Total Probability

In our next theorem, called the *theorem of total probability*, we show how the probability of an event  $B \subset A_1 \cup \dots \cup A_n$  can additively be decomposed into the products  $P(B|A_i) \cdot P(A_i)$  of conditional and unconditional probabilities. In this theorem we assume that the events  $A_1, \dots, A_n$  are *pairwise disjoint*, i. e., we assume  $A_i \cap A_j = \emptyset$ , for all  $i, j = 1, \dots, n$ , with  $i \neq j$ .

##### Theorem 4.25 (Theorem of Total Probability)

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

(i) If

(a)  $A_1, \dots, A_n \in \mathcal{A}$  are pairwise disjoint,

(b)  $B \subset \bigcup_{i=1}^n A_i$ ,

then

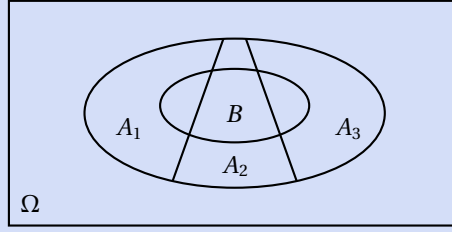
$$P(B) = \sum_{i=1}^n P(B \cap A_i). \quad (4.7)$$

(ii) If (a) and (b) of (i) hold as well as

(c)  $P(A_i) > 0, \quad \forall i = 1, \dots, n$ ,

then

$$P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i). \quad (4.8)$$



**Figure 4.3.** Venn diagram illustrating a partition of a set

(iii) If

- (a)  $A_1, A_2, \dots \in \mathcal{A}$  are pairwise disjoint,  
 (b)  $B \subset \bigcup_{i=1}^{\infty} A_i$ ,

then

$$P(B) = \sum_{i=1}^{\infty} P(B \cap A_i). \quad (4.9)$$

(iv) If (a) and (b) of (iii) hold as well as

- (c)  $P(A_i) > 0, \quad \forall i = 1, 2, \dots$ ,

then

$$P(B) = \sum_{i=1}^{\infty} P(B | A_i) \cdot P(A_i). \quad (4.10)$$

(Proof p. 148)

Equation (4.7) can be illustrated by Figure 4.3. If we assume that  $(\Omega, \mathcal{A}, P)$  is the probability space specified in Example 4.8, 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:

- (a) If the events  $A_1, \dots, A_n$  are pairwise disjoint, then  $B \cap A_1, \dots, B \cap A_n$  are pairwise disjoint as well.  
 (b) The probability measure  $P$  is additive.

#### 4.2.6 Bayes' Theorem

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

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{P(B)}. \quad (4.11)$$

Inserting Equation (4.8) for  $P(B)$  then proves the following theorem.

**Theorem 4.26 (Bayes' Theorem)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $B \in \mathcal{A}$ , and  $P(B) > 0$ . Under the assumptions (a) to (c) of Theorem 4.25 (i) and (ii),

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{\sum_{j=1}^n P(B | A_j) \cdot P(A_j)}, \quad \forall i = 1, \dots, n. \quad (4.12)$$

Analogously, under the assumptions (a) to (c) of Theorem 4.25 (iii) and (iv),

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(B | A_j) \cdot P(A_j)}, \quad \forall i \in \mathbb{N}. \quad (4.13)$$

**Example 4.27 (Joe and Ann With Random Assignment – continued)** 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

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)} \\ &= \frac{.40 \cdot .50}{.40 \cdot .50 + .40 \cdot .50} = .50 \end{aligned}$$

is the conditional probability that *Joe is drawn* given that the *drawn person is treated* (see Table 4.1, p. 131). The corresponding probability that *Ann is drawn* given that the *drawn person is treated* is identical in this example, i. e.,  $P(A^c | B) = .50$ . Hence, given treatment, each person has the same probability to be drawn. This is the *sampling perspective* of a randomized experiment supplementing the *assignment perspective*, according to which the treatment probability is the same for each person, i. e.,  $P(B | A) = P(B | A^c) = .40$  (see again Table 4.1).  $\triangleleft$

### 4.2.7 Conditional-Probability Measure

Just like probabilities, conditional probabilities of events  $A \in \mathcal{A}$  given  $B$  are values of a probability measure.

**Theorem 4.28 (Conditional-Probability Measure)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space. If  $B \in \mathcal{A}$  and  $P(B) > 0$ , then the function  $P^B: \mathcal{A} \rightarrow [0, 1]$  defined by

$$P^B(A) = P(A|B), \quad \forall A \in \mathcal{A}, \quad (4.14)$$

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

(Proof p. 148)

According to this theorem, for each  $B \in \mathcal{A}$  with  $P(B) > 0$ , the triple  $(\Omega, \mathcal{A}, P^B)$  is a probability space.

**Definition 4.29 (Conditional-Probability Measure)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space, let  $B \in \mathcal{A}$ , and let  $P(B) > 0$ . Then the function  $P^B$  defined by (4.14) is called the  $B$ -conditional-probability measure on  $(\Omega, \mathcal{A})$ .

In the following lemma we consider the relationship between conditional probabilities with respect to the measures  $P^B$  and  $P$ .

**Lemma 4.30 (Conditional Probabilities With Respect to  $P^B$ )**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space. If  $A, B, C \in \mathcal{A}$  and  $P(B \cap C) > 0$ , then

$$P^B(A|C) = P(A|B \cap C). \quad (4.15)$$

(Proof p. 149)

**Remark 4.31 (Total Conditional Probability)** Suppose  $A, B, C \in \mathcal{A}$ ,  $P(B \cap C) > 0$ , and  $P(B \cap C^c) > 0$ . This implies  $P(B) > 0$  and  $P^B(C) = P(C|B) = P(C \cap B) / P(B) > 0$ . Applying Equation (4.8) to the measure  $P^B$  then yields

$$P^B(A) = P^B(A|C) \cdot P^B(C) + P^B(A|C^c) \cdot P^B(C^c), \quad (4.16)$$

and Equations (4.14) and (4.15) imply

$$P(A|B) = P(A|B \cap C) \cdot P(C|B) + P(A|B \cap C^c) \cdot P(C^c|B). \quad (4.17)$$

◁

According to the following lemma,  $P^B$  is *absolutely continuous with respect to*  $P$ , i. e.

$$\forall A \in \mathcal{A}: P(A) = 0 \Rightarrow P^B(A) = 0. \quad (4.18)$$

This is denoted by  $P^B \ll_{\mathcal{A}} P$  [see Def. 3.70 (i)]. In contrast,  $P \ll_{\mathcal{A}} P^B$  does not always hold.

**Lemma 4.32 (Absolute Continuity of the Conditional-Probability Measure)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $B \in \mathcal{A}$ , and  $P(B) > 0$ . Then

- (i)  $P^B \ll_{\mathcal{A}} P$ , and  
(ii)  $P^B = \left( \frac{1}{P(B)} \cdot 1_B \right) \odot P$ .

(Proof p. 149)

**Remark 4.33 ( $P^B$  is a Measure With Density)** Proposition (ii) of Lemma 4.32 implies that  $P^B$  is a measure with density  $1_B/P(B)$  with respect to  $P$ . The following equations show how  $P^B(A)$  can be written as an integral in various ways:

$$\begin{aligned} \forall A \in \mathcal{A}: P^B(A) &= \int_A dP^B && [(3.8)] \\ &= \int 1_A d\left(\frac{1}{P(B)} \cdot 1_B \odot P\right) && [(3.30), \text{Lem. 4.32 (ii)}] \\ &= \int 1_A \cdot 1_B \frac{1}{P(B)} dP && [(3.72)] \\ &= \frac{1}{P(B)} \cdot \int 1_{A \cap B} dP. && [(1.33), (3.32)] \end{aligned} \quad (4.19)$$

Note that, according to Theorem 3.72 (i), the density  $1_B/P(B)$  can be written as a Radon-Nikodym derivate of  $P^B$  with respect to  $P$ , i. e.,

$$\frac{1}{P(B)} \cdot 1_B = \frac{dP^B}{dP}. \quad (4.20)$$

◁

**Example 4.34 (Joe and Ann With Random Assignment – continued)** Consider the example presented in Table 4.1 (p. 131). We specify the  $B$ -conditional-probability measure  $P^B: \mathcal{A} \rightarrow [0, 1]$  for the event that the *drawn person is treated*, i. e., for

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

For the first two elementary events,  $P^B(\{(Joe, no, +)\}) = P^B(\{(Joe, no, -)\}) = 0$ , because the intersections  $\{(Joe, no, -)\} \cap B$  and  $\{(Joe, no, +)\} \cap B$  are empty. For the next two elementary events, the  $B$ -conditional probabilities are

$$P^B(\{(Joe, yes, -)\}) = \frac{P(\{(Joe, yes, -)\} \cap B)}{P(B)} = \frac{.04}{.40} = .10$$

and

$$P^B(\{(Joe, yes, +)\}) = \frac{P(\{(Joe, yes, +)\} \cap B)}{P(B)} = \frac{.16}{.40} = .40.$$

For the next two elementary events,  $P^B(\{(Ann, no, -)\}) = P^B(\{(Ann, no, +)\}) = 0$ , because the intersections  $\{(Ann, no, -)\} \cap B$  and  $\{(Ann, no, +)\} \cap B$  are again empty. Finally, for the last two elementary events, the  $B$ -conditional probabilities are

$$P^B(\{(Ann, yes, -)\}) = \frac{P(\{(Ann, yes, -)\} \cap B)}{P(B)} = \frac{.12}{.40} = .30$$

and

$$P^B(\{(Ann, yes, +)\}) = \frac{P(\{(Ann, yes, +)\} \cap B)}{P(B)} = \frac{.08}{.40} = .20.$$

These probability are summarized in the third column of Table 4.1. Except for  $\emptyset$ , all other events 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 the probability measure [see Rule (ii) of Box 4.1 and Exercise 4-5].  $\triangleleft$

**Example 4.35 (Joe and Ann With Random Assignment – continued)** Two other conditional-probability measures on  $(\Omega, \mathcal{A})$  are  $P^A$  and  $P^{A^c}$ , where  $A$  is the event

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

that Joe is sampled and  $A^c$  the event

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

that Ann is sampled. The values of these conditional-probability measures are presented in the last two columns of Table 4.1. These measures can also be used to illustrate a *mixture of two probability measures*. As is easily seen

$$P = .50 \cdot P^A + .50 \cdot P^{A^c},$$

i. e., the measure  $P$  is a mixture of the two conditional-probability measures  $P^A$  and  $P^{A^c}$  (see Examples 4.11 and 1.62).  $\triangleleft$

**Example 4.36 (Joe and Ann With Random Assignment– continued)** In Example 4.16 we computed the two conditional probabilities  $P(C|B) = .60$  and  $P(C|B^c) = .45$  of success given treatment and no treatment, respectively. These are conditional probabilities with respect to the measure  $P$ . Let us now consider the *individual treatment effects* of Joe and of Ann. These individual effects can be computed using the  $P^A$ - and  $P^{A^c}$ -conditional probability measures, respectively. For Joe the individual treatment effect is

$$\begin{aligned} P^A(C|B) - P^A(C|B^c) &= \frac{P^A(\Omega_U \times \{yes\} \times \{+\})}{P^A(\Omega_U \times \{yes\} \times \Omega_Y)} - \frac{P^A(\Omega_U \times \{no\} \times \{+\})}{P^A(\Omega_U \times \{no\} \times \Omega_Y)} \\ &= \frac{.32 + 0}{.08 + .32 + 0 + 0} - \frac{.32 + 0}{.18 + .42 + 0 + 0} \\ &= .80 - .70 = .10, \end{aligned}$$

and for Ann it is

$$\begin{aligned}
P^{A^c}(C|B) - P^{A^c}(C|B^c) &= \frac{P^{A^c}(\Omega_U \times \{\text{yes}\} \times \{+\})}{P^{A^c}(\Omega_U \times \{\text{yes}\} \times \Omega_Y)} - \frac{P^{A^c}(\Omega_U \times \{\text{no}\} \times \{+\})}{P^{A^c}(\Omega_U \times \{\text{no}\} \times \Omega_Y)} \\
&= \frac{.16 + 0}{.24 + .16 + 0 + 0} - \frac{.12 + 0}{.48 + .12 + 0 + 0} \\
&= .40 - .20 = .20.
\end{aligned}$$

Hence, the treatment effect  $P(C|B) - P(C|B^c) = .15$  (see Example 4.16), is just the weighted average  $.50 \cdot .10 + .50 \cdot .20 = .15$  of the two individual treatment effects, where the weights are .50 for Joe and for Ann (see Example 4.35). Note that this property does not always hold [see Table 11.2 (p. 341) and Example 11.28].  $\triangleleft$

### 4.3 Independence

#### 4.3.1 Independence of Events

Independence of two events  $A$  and  $B$  means that the conditional and unconditional probabilities are the same, i. e.,  $P(A|B) = P(A)$  and  $P(B|A) = P(B)$ . This definition 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 and extends the concept of independence to more than two events.

**Definition 4.37 (Independence of Events)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space.

(i) Two events  $A, B \in \mathcal{A}$  are called  $P$ -independent, denoted  $A \perp\!\!\!\perp_P B$ , if

$$P(A \cap B) = P(A) \cdot P(B). \quad (4.21)$$

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

$$P\left(\bigcap_{i \in I_0} A_i\right) = \prod_{i \in I_0} P(A_i), \quad \forall \text{ finite } I_0 \subset I. \quad (4.22)$$

**Remark 4.38 (Pairwise and Triple-Wise Independence)** For  $n$  events  $A_1, \dots, A_n$ ,  $P$ -independence will also be denoted by

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

For three events, for instance, it means that

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j), \quad i \neq j, \quad i, j = 1, 2, 3, \quad (4.23)$$

(pairwise  $P$ -independence) and

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3) \quad (4.24)$$

(triple-wise  $P$ -independence) hold.

Note that pairwise  $P$ -independence of more than two events does not imply  $P$ -independence of these events. Furthermore, triple-wise  $P$ -independence, for instance, does not imply pairwise  $P$ -independence. For more propositions on independence of events see Box 4.2 (p. 147).  $\triangleleft$

**Remark 4.39 (Independence of any event  $A$  with  $\Omega$  and  $\emptyset$ )** For any probability space  $(\Omega, \mathcal{A}, P)$ ,

$$\forall A \in \mathcal{A}: A \perp\!\!\!\perp_P \Omega \quad \text{and} \quad A \perp\!\!\!\perp_P \emptyset. \quad (4.25)$$

(see Exercise 4-7).  $\triangleleft$

### 4.3.2 Independence of Set Systems

Now we extend the concept of  $P$ -independence to *set systems*, i. e., to sets of events, and illustrate independence by an example.

#### Definition 4.40 (Family of Independent Set Systems)

Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $\mathcal{E}_i \subset \mathcal{A}$ ,  $i \in I \neq \emptyset$ . Then  $(\mathcal{E}_i, i \in I)$  is called a family of  $P$ -independent set systems, denoted  $\perp\!\!\!\perp_P (\mathcal{E}_i, i \in I)$ , if  $\perp\!\!\!\perp_P (A_i, i \in I)$  holds for all families  $(A_i, i \in I)$  with  $A_i \in \mathcal{E}_i$ ,  $i \in I$ . If  $I = \{1, 2\}$  we also use the notation  $\mathcal{E}_1 \perp\!\!\!\perp_P \mathcal{E}_2$  instead of  $\perp\!\!\!\perp_P (\mathcal{E}_i, i \in I)$ .

**Remark 4.41 (Independence of an Event and a Set System)** Let  $(\Omega, \mathcal{A}, P)$  be a probability space. An event  $A \in \mathcal{A}$  and a set system  $\mathcal{E} \subset \mathcal{A}$  are called  $P$ -independent, denoted  $A \perp\!\!\!\perp_P \mathcal{E}$ , if  $\{A\} \perp\!\!\!\perp_P \mathcal{E}$ .  $\triangleleft$

**Remark 4.42 (Independence of  $\sigma$ -Algebras)** Note that  $\sigma$ -algebras are special set systems referred to in the definition above. Hence, a family  $(\mathcal{A}_i, i \in I)$  of sub- $\sigma$ -algebras of  $\mathcal{A}$  can be  $P$ -independent as well. This fact will be used introducing the concept of  $P$ -independence of random variables (see section 5.4).  $\triangleleft$

**Example 4.43 (Joe and Ann With Random Assignment – continued)** Suppose  $A = \{\text{Joe}\} \times \Omega_X \times \Omega_Y$  denotes the event that Joe is sampled and  $B = \Omega_U \times \{\text{yes}\} \times \Omega_Y$  the event that the person sampled is treated. Then  $A$  and  $B$  are independent, because

$$P(A \cap B) = P(\{\text{Joe}\} \times \{\text{yes}\} \times \Omega_Y) = .04 + .16 = .20$$

and

$$\begin{aligned} P(A) \cdot P(B) &= P(\{\text{Joe}\} \times \Omega_X \times \Omega_Y) \cdot P(\Omega_U \times \{\text{yes}\} \times \Omega_Y) \\ &= (.09 + .21 + .04 + .16) \cdot (.04 + .16 + .12 + .08) \\ &= .50 \cdot .40 = .20. \end{aligned}$$

Hence,  $P(A \cap B) = P(A) \cdot P(B)$ . This implies that the  $\sigma$ -algebras  $\{A, A^c, \Omega, \emptyset\}$  and  $\{B, B^c, \Omega, \emptyset\}$  are independent as well [see Box 4.2 (iii)]. In fact, this is a special

case of the following theorem, because the set systems  $\mathcal{E}_1 := \{A\}$  and  $\mathcal{E}_2 := \{B\}$  are  $\cap$ -stable (see Def. 1.36) and  $\sigma(\mathcal{E}_1) = \{A, A^c, \Omega, \emptyset\}$  and  $\sigma(\mathcal{E}_2) = \{B, B^c, \Omega, \emptyset\}$  are the  $\sigma$ -algebras generated by  $\mathcal{E}_1$  and  $\mathcal{E}_2$ , respectively (see Def. 1.13 and Example 1.17).  $\triangleleft$

According to the following theorem, it is sufficient to check  $P$ -independence of a family of  $\cap$ -stable generating system in order to check  $P$ -independence of a family of  $\sigma$ -algebras. In this theorem,  $(\sigma(\mathcal{E}_i), i \in I)$  denotes the family of  $\sigma$ -algebras generated by the set systems  $\mathcal{E}_i, i \in I$ .

**Theorem 4.44 ( $\cap$ -Stable Set Systems and Independence)**

If  $(\Omega, \mathcal{A}, P)$  is a probability space and  $\mathcal{E}_i \subset \mathcal{A}, i \in I$ , are  $\cap$ -stable, then

$$\prod_P (\mathcal{E}_i, i \in I) \Rightarrow \prod_P (\sigma(\mathcal{E}_i), i \in I). \quad (4.26)$$

For a proof see Georgii (2008, Theorem 3.19, p. 65).

## 4.4 Conditional Independence 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*.

### 4.4.1 Conditional Independence of Events Given an Event

**Definition 4.45 (Conditional Independence of Two Events)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $A, B, C \in \mathcal{A}$ , and  $P(B) > 0$ . Then the events  $A$  and  $C$  are called  $B$ -conditionally  $P$ -independent, denoted  $A \perp\!\!\!\perp_P C | B$ , if

$$P(A \cap C | B) = P(A | B) \cdot P(C | B). \quad (4.27)$$

**Remark 4.46 (A Condition Equivalent to Conditional Independence)** Suppose  $P(B \cap C) > 0$ . Then Equation (4.27) is equivalent to

$$P(A | B \cap C) = P(A | B) \quad (4.28)$$

[see Box 4.2 (xii)]. Exchanging  $A$  and  $C$  immediately yields: If  $P(A \cap B) > 0$  then Equation (4.27) is equivalent to

$$P(C | A \cap B) = P(C | B). \quad (4.29)$$

$\triangleleft$

**Remark 4.47 (Independence of Events With Respect to  $P^B$ )** Using the conditional-probability measure  $P^B$  defined by (4.14) we can rewrite Equation (4.27):

$$P^B(A \cap C) = P^B(A) \cdot P^B(C). \quad (4.30)$$

This equation shows that  $B$ -conditional  $P$ -independence of  $A$  and  $C$  is equivalent to  $P^B$ -independence of  $A$  and  $C$ , which will also be denoted by  $A \perp\!\!\!\perp_{P^B} C$ .  $\triangleleft$

**Remark 4.48 (Independence and Conditional Independence)** Assume that  $B \in \mathcal{A}$  with  $P(B) > 0$ . Then  $P$ -independence of  $A$  and  $C$  neither implies nor is implied by  $B$ -conditional  $P$ -independence of  $A$  and  $C$  (see Exercise 4-8). However,  $P$ -independence of  $A, B$ , and  $C$  *does* imply  $B$ -conditional  $P$ -independence of  $A$  and  $C$  [see Box 4.2 (x)]. For more propositions on conditional independence of events see Box 4.2 (p. 147), which is proved in Exercise 4-9.  $\triangleleft$

#### 4.4.2 Conditional Independence of Set Systems Given an Event

Now we extend the concept of conditional  $P$ -independence to *set systems*. In Remark 4.47 we already noted that  $B$ -conditional  $P$ -independence of two events  $A$  and  $C$  is equivalent to  $P^B$ -independence of  $A$  and  $C$ . Correspondingly,  $B$ -conditional  $P$ -independence of a family  $(\mathcal{E}_i, i \in I)$  of events is defined as  $P^B$ -independence of  $(\mathcal{E}_i, i \in I)$ .

##### Definition 4.49 (Family of Conditionally Independent Set Systems)

Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $B \in \mathcal{A}$  with  $P(B) > 0$ , and  $\mathcal{E}_i \subset \mathcal{A}$ ,  $i \in I$ . Then  $(\mathcal{E}_i, i \in I)$  is called a *family of  $B$ -conditionally  $P$ -independent set systems*, denoted  $\perp\!\!\!\perp_{P^B}(\mathcal{E}_i, i \in I) | B$ , if  $\perp\!\!\!\perp_{P^B}(\mathcal{E}_i, i \in I)$ .

**Remark 4.50 (Conditional Independence of  $\sigma$ -Algebras)** Again,  $\sigma$ -algebras can be such set systems referred to in the definition above. Hence, a family  $(\mathcal{A}_i, i \in I)$  of sub- $\sigma$ -algebras of  $\mathcal{A}$  can be  $B$ -conditionally  $P$ -independent as well.  $\triangleleft$

**Remark 4.51 (Independence of Set Systems With Respect to  $P^B$ )** According to Theorem 4.44, under the assumptions of Definition 4.49,  $\cap$ -stability of the set systems  $\mathcal{E}_i$ ,  $i \in I$ , implies

$$\perp\!\!\!\perp_{P^B}(\mathcal{E}_i, i \in I) \Rightarrow \perp\!\!\!\perp_{P^B}(\sigma(\mathcal{E}_i), i \in I). \quad (4.31)$$

$\triangleleft$

Together with Definition 4.49, this remark immediately implies the following corollary.

**Box 4.2 Independence and Conditional Independence of Events**

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

$$A \perp\!\!\!\perp_p B \Leftrightarrow P(A \cap B) = P(A) \cdot P(B) \quad (\text{i})$$

$$A \perp\!\!\!\perp_p B \Leftrightarrow A^c \perp\!\!\!\perp_p B \quad (\text{ii})$$

$$A \perp\!\!\!\perp_p B \Leftrightarrow \sigma(\{A\}) \perp\!\!\!\perp_p \sigma(\{B\}). \quad (\text{iii})$$

$$\begin{aligned} \perp\!\!\!\perp_p A, B, C \Leftrightarrow & P(A \cap B) = P(A) \cdot P(B), \\ & P(A \cap C) = P(A) \cdot P(C), \\ & P(B \cap C) = P(B) \cdot P(C), \\ & P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C). \end{aligned} \quad (\text{iv})$$

$$\perp\!\!\!\perp_p A, B, C \Rightarrow A \perp\!\!\!\perp_p B, A \perp\!\!\!\perp_p C, B \perp\!\!\!\perp_p C. \quad (\text{v})$$

If  $P(B) > 0$ , then

$$A \perp\!\!\!\perp_p B \Leftrightarrow P(A|B) = P(A) \quad (\text{vi})$$

$$A \perp\!\!\!\perp_p C|B \Leftrightarrow P(A \cap C|B) = P(A|B) \cdot P(C|B) \quad (\text{vii})$$

$$A \perp\!\!\!\perp_p C|B \Leftrightarrow A \perp\!\!\!\perp_{p^B} C \quad (\text{viii})$$

$$A \perp\!\!\!\perp_p C|B \Leftrightarrow A \perp\!\!\!\perp_p C^c|B \quad (\text{ix})$$

$$\perp\!\!\!\perp_p A, B, C \Rightarrow A \perp\!\!\!\perp_p C|B. \quad (\text{x})$$

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

$$A \perp\!\!\!\perp_p B \Leftrightarrow P(A|B) = P(A|B^c). \quad (\text{xi})$$

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

$$A \perp\!\!\!\perp_p C|B \Leftrightarrow P(A|B \cap C) = P(A|B). \quad (\text{xii})$$

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

$$A \perp\!\!\!\perp_p C|B \Leftrightarrow P(A|B \cap C^c) = P(A|B). \quad (\text{xiii})$$

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

$$A \perp\!\!\!\perp_p C|B \Leftrightarrow P(A|B \cap C) = P(A|B \cap C^c) \quad (\text{xiv})$$

$$B \perp\!\!\!\perp_p C \Rightarrow P(A|B) = P(A|B \cap C) \cdot P(C) + P(A|B \cap C^c) \cdot P(C^c) \quad (\text{xv})$$

$$A \perp\!\!\!\perp_p C|B \Rightarrow P(A|B) = P(A|B \cap C) \cdot P(C) + P(A|B \cap C^c) \cdot P(C^c). \quad (\text{xvi})$$

**Corollary 4.52 ( $\cap$ -Stable Set Systems and Conditional Independence)**

If  $(\Omega, \mathcal{A}, P)$  is a probability space,  $B \in \mathcal{A}$  with  $P(B) > 0$ , and  $(\mathcal{E}_i, i \in I)$  is a family of  $\cap$ -stable set systems  $\mathcal{E}_i \subset \mathcal{A}$ , then

$$\prod_P (\mathcal{E}_i, i \in I) | B \Rightarrow \prod_P (\sigma(\mathcal{E}_i), i \in I) | B. \quad (4.32)$$

**4.5 Proofs****Proof of Theorem 4.22**

In Remark 4.21 we have already shown that Equation (4.6) holds for  $n = 2$  [see Box 4.1 v]. Hence, for an induction over  $n$  it suffices to show that (4.6) holds for  $A_1, \dots, A_n$  if it holds for  $A_1, \dots, A_{n-1}$ . Note that  $P(\bigcap_{i=1}^{n-1} A_i) > 0$  implies  $P(\bigcap_{i=1}^{j-1} A_i) > 0$  for  $2 \leq j \leq n$ . Hence,

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= P\left(\bigcap_{i=1}^{n-1} A_i \cap A_n\right) \\ &= P\left(\bigcap_{i=1}^{n-1} A_i\right) \cdot P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right) && [(4.6)] \\ &= P(A_1) \cdot \left[ \prod_{j=2}^{n-1} P\left(A_j \mid \bigcap_{i=1}^{j-1} A_i\right) \right] \cdot P\left(A_n \mid \bigcap_{i=1}^{n-1} A_i\right) && [(4.6), \text{ ass. of induction}] \\ &= P(A_1) \cdot \prod_{j=2}^n P\left(A_j \mid \bigcap_{i=1}^{j-1} A_i\right). \end{aligned}$$

**Proof of Theorem 4.25**

(i) This equation immediately follows from (1.29).

(ii) If  $P(A_i) > 0$ , then  $P(B \cap A_i) = P(B|A_i) \cdot P(A_i)$  [see Eq. (4.2)]. Hence, (4.7) immediately implies

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

(iii) This proposition immediately follows from (1.30).

(iv) This proposition immediately follows from (iii) inserting  $P(B \cap A_i) = P(B|A_i) \cdot P(A_i)$  [see Eq. (4.2)].

**Proof of Theorem 4.28**

4.1 (a)

$$\begin{aligned}
P^B(\Omega) &= \frac{P(B \cap \Omega)}{P(B)} && [(4.2)] \\
&= \frac{P(B)}{P(B)} && [B \subset \Omega] \\
&= 1.
\end{aligned}$$

4.1 (b) We assume  $P(B) > 0$ . Therefore,  $P(A \cap B) \geq 0$ , for all  $A \in \mathcal{A}$ , implies that  $P^B(A) = P(A \cap B)/P(B) \geq 0$ , for all  $A \in \mathcal{A}$ .

4.1 (c) If  $A_1, A_2, \dots$  are pairwise disjoint, then  $A_1 \cap B, A_2 \cap B, \dots$  are pairwise disjoint. Therefore,

$$\begin{aligned}
P^B\left(\bigcup_{i=1}^{\infty} A_i\right) &= \frac{P\left[\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right]}{P(B)} && [(4.2)] \\
&= \frac{P\left[\bigcup_{i=1}^{\infty} (A_i \cap B)\right]}{P(B)} \\
&= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} && [\text{Def. 4.1 (c)}] \\
&= \sum_{i=1}^{\infty} P^B(A_i). && [(4.2)]
\end{aligned}$$

### ***Proof of Lemma 4.30***

$$\begin{aligned}
P(A|B \cap C) &= \frac{P(A \cap B \cap C)}{P(B \cap C)} && [(4.2)] \\
&= \frac{P(A \cap C|B) \cdot P(B)}{P(C|B) \cdot P(B)} && [(4.2)] \\
&= \frac{P^B(A \cap C)}{P^B(C)} && [(4.14)] \\
&= P^B(A|C). && [(4.2)]
\end{aligned}$$

### ***Proof of Lemma 4.32***

(i) For all  $A \in \mathcal{A}$ ,

$$\begin{aligned}
P(A) = 0 &\Rightarrow P(A \cap B) = 0 && [\text{Box 4.1 (v)}] \\
&\Rightarrow \frac{P(A \cap B)}{P(B)} = 0 && [P(B) > 0] \\
&\Rightarrow P(A|B) = 0 && [(4.2)] \\
&\Rightarrow P^B(A) = 0. && [(4.14)]
\end{aligned}$$

Hence,  $P^B \ll P$  [see Def. 3.70].

(ii) For all  $A \in \mathcal{A}$ ,

$$\begin{aligned}
P^B(A) &= \frac{P(A \cap B)}{P(B)} && [(4.14), (4.2)] \\
&= \frac{1}{P(B)} \int 1_{A \cap B} dP && [(3.9)] \\
&= \frac{1}{P(B)} \int 1_A \cdot 1_B dP && [(1.33)] \\
&= \int_A \frac{1}{P(B)} \cdot 1_B dP. && [(3.30), (3.32)]
\end{aligned}$$

According to Theorem 3.65, this means  $P^B = \left( \frac{1}{P(B)} \cdot 1_B \right) \circledast P$ .

## 4.6 Exercises

▷ **Exercise 4-1** Consider flipping a coin  $n$  times and the event  $A_1 =$  *flipping heads at the first flip*. Specify the set  $\Omega$  of possible outcomes of this random experiment and the set  $A_1$  as a subset of  $\Omega$ . How many elements has  $\Omega$ ? How many elements has the event  $A_1$ ?

▷ **Exercise 4-2** Draw the interval  $[0, 1]$ , cut it in two halves, cut the right-hand piece in two halves, cut the remaining part in two halves, etc. In this way you can visualize the sequence  $1/2, 1/4, 1/8, \dots$  by lengths of intervals. This sequence can also be written:  $1/2^i, i \in \mathbb{N}$ . Note that all terms  $1/2^i$  of this sequence are positive, i. e.,  $1/2^i > 0$  for all  $i \in \mathbb{N}$ . Determine

$$\sum_{i=1}^{\infty} \frac{1}{2^i} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{2^i}.$$

▷ **Exercise 4-3** Consider Example 4.19 and list all elements of the  $\sigma$ -algebras  $\mathcal{F}_1$  and  $\mathcal{F}_2$  referring explicitly to the elements of  $\Omega$ .

▷ **Exercise 4-4** Compute the probabilities  $P(A)$  and  $P(C | A \cap B)$  of the events defined in Example 4.23.

▷ **Exercise 4-5** Compute the  $B$ -conditional probability of the event  $\{(Ann, yes, -), (Ann, yes, +)\}$ , i. e., the event that *Ann is sampled and treated*. Use the results already obtained in Example 4.34.

▷ **Exercise 4-6** In Example 4.16 we computed the conditional probabilities  $P(C | B) = .60$  and  $P(C | B^c) = .45$ , where  $C$  is the event that the *drawn person is successful* and  $B$  is the event that the *drawn person is treated*. What is the conclusion the effect of the treatment if we compare these two conditional probabilities to each other?

▷ **Exercise 4-7** Prove the proposition of Remark 4.39.

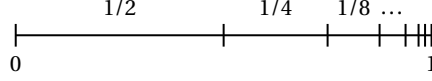
▷ **Exercise 4-8** Show by examples that, for  $B \in \mathcal{A}$  with  $P(B) > 0$ ,  $P$ -independence of  $A$  and  $C$  neither implies nor is implied by  $B$ -conditional  $P$ -independence of  $A$  and  $C$ .

▷ **Exercise 4-9** Prove the propositions of Box 4.2.

## Solutions

▷ **Solution 4-1** The set of possible outcomes is  $\Omega = \{h, t\}^n = \{h, t\} \times \dots \times \{h, t\}$  ( $n$ -times). The event *flipping heads at time 1* is  $A_1 = \{h\} \times \{h, t\}^{n-1}$ . The set  $\Omega$  has  $2^n$  elements and  $A_1$  has  $2^n/2 = 2^{n-1}$  elements.

▷ **Solution 4-2** The picture of this interval is



and this illustrates that  $\sum_{i=1}^{\infty} \frac{1}{2^i} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{2^i} = 1$ .

▷ **Solution 4-3**

$$\begin{aligned} \mathcal{F}_1 &= \sigma(\{A\}) = \{A, A^c, \Omega, \emptyset\} \\ &= \left\{ \{ (Joe, no, -), (Joe, no, +), (Joe, yes, -), (Joe, yes, +) \}, \right. \\ &\quad \left. \{ (Ann, no, -), (Ann, no, +), (Ann, yes, -), (Ann, yes, +) \}, \Omega, \emptyset \right\}. \end{aligned}$$

$$\begin{aligned} \mathcal{F}_2 &= \sigma(\{A, B\}) \\ &= \{A, A^c, B, B^c, (A \cap B) \cup (A^c \cap B^c), (A \cap B^c) \cup (A^c \cap B), \\ &\quad A \cap B, A^c \cap B, A \cap B^c, A^c \cap B^c, \\ &\quad A \cup B, A^c \cup B, A \cup B^c, A^c \cup B^c, \Omega, \emptyset\} \end{aligned}$$

where

$$\begin{aligned} A &= \{ (Joe, yes, -), (Joe, yes, +), (Joe, no, -), (Joe, no, +) \}, \\ A^c &= \{ (Ann, yes, -), (Ann, yes, +), (Ann, no, -), (Ann, no, +) \}, \\ B &= \{ (Joe, yes, -), (Joe, yes, +), (Ann, yes, -), (Ann, yes, +) \}, \\ B^c &= \{ (Joe, no, -), (Joe, no, +), (Ann, no, -), (Ann, no, +) \}, \\ (A \cap B) \cup (A^c \cap B^c) &= \{ (Joe, yes, -), (Joe, yes, +), (Ann, no, -), (Ann, no, +) \}, \\ (A \cap B^c) \cup (A^c \cap B) &= \{ (Joe, no, -), (Joe, no, +), (Ann, yes, -), (Ann, yes, +) \}, \end{aligned}$$

$$\begin{aligned} A \cap B &= \{ (Joe, yes, -), (Joe, yes, +) \}, \\ A^c \cap B &= \{ (Ann, yes, -), (Ann, yes, +) \}, \\ A \cap B^c &= \{ (Joe, no, -), (Joe, no, +) \}, \\ A^c \cap B^c &= \{ (Ann, no, -), (Ann, no, +) \}, \end{aligned}$$

$$\begin{aligned} A \cup B &= \{ (Joe, yes, -), (Joe, yes, +), (Joe, no, -), (Joe, no, +), (Ann, yes, -), (Ann, yes, +) \}, \\ A^c \cup B &= \{ (Ann, yes, -), (Ann, yes, +), (Ann, no, -), (Ann, no, +), (Joe, yes, -), (Joe, yes, +) \}, \\ A \cup B^c &= \{ (Joe, yes, -), (Joe, yes, +), (Joe, no, -), (Joe, no, +), (Ann, no, -), (Ann, no, +) \}, \\ A^c \cup B^c &= \{ (Ann, yes, -), (Ann, yes, +), (Ann, no, -), (Ann, no, +), (Joe, no, -), (Joe, no, +) \}. \end{aligned}$$

▷ **Solution 4-4** Because the four events  $\{(Joe, no, -)\}, \dots, \{(Joe, yes, +)\}$  are pairwise disjoint, we can simply add their probabilities. Hence,  $P(A) = .09 + .21 + .04 + .16 = .5$  (see the second column of Table 4.1, p. 131). In order to compute  $P(C|A \cap B)$  note that  $A \cap B = \{(Joe, yes, -), (Joe, yes, +)\}$  is the event that *Joe is drawn and treated*. Again, because the two elementary events involved are disjoint,  $P(A \cap B) = .04 + .16 = 0.2$ . Furthermore,  $A \cap B \cap C = \{(Joe, yes, +)\}$  is the event that *Joe is drawn, treated, and successful*. Its probability is  $P(A \cap B \cap C) = P(\{(Joe, yes, +)\}) = .16$ . Hence,

$$P(C|A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)} = \frac{.16}{.04 + .16} = .8.$$

▷ **Solution 4-5** In Example 4.34 we already computed the two  $B$ -conditional probabilities  $P^B(\{(Ann, yes, -)\}) = .30$  and  $P^B(\{(Ann, yes, +)\}) = .20$ . Because these elementary events are *disjoint*, the probabilities of their union can easily be computed using the additivity property of the probability measure  $P^B$ . Hence,  $P^B(\{(Ann, yes, -), (Ann, yes, +)\}) = .30 + .20 = .50$ .

▷ **Solution 4-6** Although this question and the concepts needed for an answer are beyond the scope of this book, the difference  $P(C|B) - P(C|B^c) = .60 - .45 = .15$  is the *average total treatment effect* (see Steyer et al., 2014). It is the average of the two *individual total treatment effects* of Joe and of Ann. For Joe this individual treatment effect is  $.80 - .70 = .10$ , (probability of *success* given *Joe* and *treatment* – probability of *success* given *Joe* and no *treatment*) whereas it is  $.40 - .20 = 0$  for Ann.

▷ **Solution 4-7** Let  $(\Omega, \mathcal{A}, P)$  be a probability space. Then for all  $A \in \mathcal{A}$ :  
 $P(\Omega \cap A) = P(A) = 1 \cdot P(A) = P(\Omega) \cdot P(A)$  and  $P(\emptyset \cap A) = P(\emptyset) = 0 = 0 \cdot P(A) = P(\emptyset) \cdot P(A)$ .

▷ **Solution 4-8** Consider Example 2.2.2, let  $A = \{(h, t), (h, h)\}$  denote the event to flip heads with the first coin,  $B = \{(t, t), (h, h)\}$  the event to flip tails or to flip heads with both coins, and  $C = \{(t, h), (h, h)\}$  the event to flip heads with the second coin. All three events have the same probability  $P(A) = P(B) = P(C) = .5$ . Now,

$$P(A \cap C) = P(\{(h, h)\}) = .25 = .5 \cdot .5 = P(A) \cdot P(C)$$

and

$$P(B \cap C) = P(\{(h, h)\}) = .25 = .5 \cdot .5 = P(B) \cdot P(C).$$

Hence,  $A$  and  $C$  as well as  $B$  and  $C$  are  $P$ -independent, which implies  $P(A|B) = .5$  and  $P(C|B) = .5$ . However,

$$\begin{aligned} P(A \cap C|B) &= \frac{P(A \cap C \cap B)}{P(B)} = \frac{.25}{.5} = .5 \\ &\neq .25 = .5 \cdot .5 = P(A|B) \cdot P(C|B), \end{aligned}$$

which shows that  $A$  and  $C$  are not  $B$ -conditionally  $P$ -independent.

Now we present an example in which  $A$  and  $C$  are  $B$ -conditionally  $P$ -independent but not (unconditionally)  $P$ -independent. Consider flipping three coins. This random experiment is represented by the probability space  $(\Omega, \mathcal{A}, P)$ , where  $\Omega = \{h, t\}^3$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$ , and  $P: \mathcal{A} \rightarrow [0, 1]$ , satisfying  $P(\{\omega\}) = .125$  for all  $\omega \in \Omega$ . Furthermore, let  $A = \{(t, t, t), (t, t, h)\}$  denote the event to flip tails with the first two coins,  $B = \{(t, t, t), (t, t, h), (t, h, t), (t, h, h)\}$  the event to flip tails with the first coin, and  $C = \{(t, t, h), (t, h, h)\}$  the event to flip tails with the first coin and heads with the third coin. The two events  $A$  and  $C$  have the same probability  $P(A) = P(C) = .25$  and  $P(B) = .5$ . Because

$$P(A \cap C) = P(\{(t, t, h)\}) = .125 \neq .25 \cdot .25 = P(A) \cdot P(C),$$

$A$  and  $C$  are *not*  $P$ -independent. Further,  $P(A \cap B) = P(\{(t, t, t), (t, t, h)\}) = .25$ ,  $P(C \cap B) = P(\{(t, t, h), (t, h, h)\}) = .25$ , and

$$\begin{aligned} P(A \cap C | B) &= \frac{P(A \cap B \cap C)}{P(B)} = \frac{.125}{.5} = .25 \\ &= .5 \cdot .5 = \frac{P(A \cap B)}{P(B)} \cdot \frac{P(B \cap C)}{P(B)} = P(A | B) \cdot P(C | B). \end{aligned}$$

This shows that  $A$  and  $C$  are  $B$ -conditionally  $P$ -independent.

▷ **Solution 4-9** (i) This is the definition of  $A \perp\!\!\!\perp_P B$ .

$$\begin{aligned} \text{(ii)} \quad P(A^c \cap B) &= P(B \setminus A) \\ &= P(B) - P(A \cap B) && \text{[Box 4.1 (vi)]} \\ &= P(B) - P(A) \cdot P(B) && [A \perp\!\!\!\perp_P B] \\ &= [1 - P(A)] \cdot P(B) \\ &= P(A^c) \cdot P(B) && \text{[Box 4.1 (iv)],} \end{aligned}$$

which is  $A^c \perp\!\!\!\perp_P B$ .

(iii) We have to show that  $A \perp\!\!\!\perp_P B$  implies

$$P(A_i \cap B_j) = P(A_i) \cdot P(B_j), \quad \forall A_i \in \{A, A^c, \Omega, \emptyset\} \text{ and } \forall B_j \in \{B, B^c, \Omega, \emptyset\}.$$

Whenever  $A_i$  or  $B_j$  is  $\Omega$  or  $\emptyset$  this equation holds [see (4.25)]. Furthermore,  $P(A \cap B) = P(A) \cdot P(B)$  is equivalent to our premise  $A \perp\!\!\!\perp_P B$ , and  $P(A^c \cap B) = P(A^c) \cdot P(B)$  is proposition (ii). The corresponding argument holds for  $P(A \cap B^c) = P(A) \cdot P(B^c)$  and  $P(A^c \cap B^c) = P(A^c) \cdot P(B^c)$ , exchanging the roles of  $A$  and  $B$ .

(iv) This is the definition of  $\perp\!\!\!\perp_P A, B, C$ .

(v) This proposition immediately follows from (iv) and (i).

(vi) We assume  $P(B) > 0$ . Then

$$\begin{aligned} A \perp\!\!\!\perp_P B &\Leftrightarrow P(A \cap B) = P(A) \cdot P(B) && \text{[(i)]} \\ &\Leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A) \\ &\Leftrightarrow P(A | B) = P(A). && \text{[(4.2)]} \end{aligned}$$

(vii) This is the definition of  $A \perp\!\!\!\perp_P C | B$ .

(viii) We assume  $P(B) > 0$ . Therefore,

$$\begin{aligned} A \perp\!\!\!\perp_P C | B &\Leftrightarrow P(A \cap C | B) = P(A | B) \cdot P(C | B) && \text{[(vii)]} \\ &\Leftrightarrow P^B(A \cap C) = P^B(A) \cdot P^B(C) && \text{[(4.14)]} \\ &\Leftrightarrow A \perp\!\!\!\perp_{P^B} C && \text{[(i)]} \end{aligned}$$

(ix) We assume  $P(B) > 0$ . Therefore,

$$\begin{aligned} A \perp\!\!\!\perp_P C | B &\Leftrightarrow A \perp\!\!\!\perp_{P^B} C && \text{[(viii)]} \\ &\Leftrightarrow A \perp\!\!\!\perp_{P^B} C^c. && \text{[(ii)]} \end{aligned}$$

(x) We assume  $P(B) > 0$ . Then

$$\begin{aligned}
P(A \cap C | B) &= \frac{P(A \cap B \cap C)}{P(B)} && [(4.2)] \\
&= \frac{P(A) \cdot P(B) \cdot P(C)}{P(B)} && [A \perp_p B, C, \text{ (iv)}] \\
&= P(A|B) \cdot P(C|B). && [(v), (vi)]
\end{aligned}$$

(xi) We assume  $P(B), P(B^c) > 0$ . Then

$$\begin{aligned}
P(A|B) = P(A|B^c) &\Leftrightarrow \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B^c)}{1 - P(B)} && [(4.2), \text{Box 4.1 (iv)}] \\
&\Leftrightarrow P(A \cap B) \cdot [1 - P(B)] = P(A \cap B^c) \cdot P(B) \\
&\Leftrightarrow P(A \cap B) = [P(A \cap B) + P(A \cap B^c)] \cdot P(B) \\
&\Leftrightarrow P(A \cap B) = P(A) \cdot P(B) && [(4.7)] \\
&\Leftrightarrow A \perp_p B. && [(i)]
\end{aligned}$$

(xii) We assume  $P(B \cap C) > 0$ . This implies  $P(B) > 0$  and

$$\begin{aligned}
A \perp_p C | B &\Leftrightarrow P(A \cap C | B) = P(A|B) \cdot P(C|B) && [(vii)] \\
&\Leftrightarrow \frac{P(A \cap B \cap C)}{P(B)} = \frac{P(A \cap B)}{P(B)} \cdot \frac{P(B \cap C)}{P(B)} && [(4.2)] \\
&\Leftrightarrow \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(A \cap B)}{P(B)} \\
&\Leftrightarrow P(A|B \cap C) = P(A|B). && [(4.2)]
\end{aligned}$$

(xiii) We assume  $P(B \cap C^c) > 0$ . This implies  $P(B) > 0$  and

$$\begin{aligned}
A \perp_p C | B &\Leftrightarrow A \perp_{P^B} C && [(viii)] \\
&\Leftrightarrow A \perp_{P^B} C^c && [(ii)] \\
&\Leftrightarrow P^B(A|C^c) = P^B(A) && [(vi)] \\
&\Leftrightarrow P(A|B \cap C^c) = P(A|B) && [(4.15), (4.14)]
\end{aligned}$$

(xiv) We assume  $P(B \cap C), P(B \cap C^c) > 0$ .

$$\begin{aligned}
P(A|B \cap C) = P(A|B \cap C^c) &\Leftrightarrow P^B(A|C) = P^B(A|C^c) && [(4.15)] \\
&\Leftrightarrow A \perp_{P^B} C && [(xi)] \\
&\Leftrightarrow A \perp_p C | B. && [(viii)]
\end{aligned}$$

(xv) We assume  $P(B \cap C), P(B \cap C^c) > 0$ .

$$\begin{aligned}
B \perp_p C &\Rightarrow P(C|B) = P(C), P(C^c|B) = P(C^c) && [(vi), (ii)] \\
&\Rightarrow P(A|B) = P(A|B \cap C) \cdot P(C) + P(A|B \cap C^c) \cdot P(C^c). && [(4.17)]
\end{aligned}$$

(xvi) We assume  $P(B \cap C), P(B \cap C^c) > 0$  and  $A \perp_p C | B$ .

$$\begin{aligned}
P(A|B) &= P(A|B \cap C) && [A \perp_p C | B, (xii)] \\
&= P(A|B \cap C) \cdot [P(C) + P(C^c)] && [\text{Box 4.1 (iv)}] \\
&= P(A|B \cap C) \cdot P(C) + P(A|B \cap C) \cdot P(C^c) \\
&= P(A|B \cap C) \cdot P(C) + P(A|B \cap C^c) \cdot P(C^c). && [(xiv)]
\end{aligned}$$

## Chapter 5

# Random Variable, Distribution, Density, and Distribution Function

In chapter 4 we translated the concepts *measure* and *measure space* to probability theory introducing the notions *probability measure* and *probability space*. In this chapter we define a *random variable* as a measurable mapping and its *distribution* as the image measure of a measurable mapping with respect to a probability measure (see ch. 2). The distribution of a random variable contains the comprehensive information about its properties. It informs us about the probability of each event that can be represented by this random variable. Expectation, variance and other moments of a random variable are determined by its distribution (see ch. 6). For a multivariate random variable, the (joint) distribution also contains the information about the dependencies between its components. It also determines the conditional expectations (see ch. 10). In this chapter, we apply the concept of independence of families of events in order to introduce *independence of random variables* and *families of random variables*. Finally, the last sections of this chapter are devoted to the concept of a *probability function*, and, for a real-valued random variable, the notions of a *distribution function* and a *probability density*, which are very useful for describing a distribution, for calculations (see, e. g., ch. 6), and for providing instructive illustrations of the underlying distributions (see ch. 8).

### 5.1 Random Variable and its Distribution

In section 2.6 we introduced the notation

$$f: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}'),$$

which expresses that  $f: \Omega \rightarrow \Omega'$  is an  $(\mathcal{A}, \mathcal{A}')$ -measurable mapping and that  $\mu$  is a measure on the measurable space  $(\Omega, \mathcal{A})$ . If  $\mu$  is a probability measure, then a measurable mapping is also called a *random variable* and its image measure  $\mu_f$  is also called its *distribution*. This change of terms goes along with a change in notation. Instead of  $f$ ,  $g$ , and  $h$ , we preferably use letters such as  $X$ ,  $Y$ , and  $Z$ .

#### Definition 5.1 (Random Variable)

If  $(\Omega, \mathcal{A}, P)$  is a probability space and  $X: (\Omega, \mathcal{A}) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  a measurable mapping, i. e., if  $X: \Omega \rightarrow \Omega'_X$  satisfies

$$X^{-1}(A') \in \mathcal{A}, \quad \forall A' \in \mathcal{A}'_X, \quad (5.1)$$

then  $X$  is called a random variable on  $(\Omega, \mathcal{A}, P)$  with values in  $(\Omega'_X, \mathcal{A}'_X)$ . If  $(\Omega'_X, \mathcal{A}'_X) = (\mathbb{R}, \mathcal{B})$ , then  $X$  is called real-valued, and if  $(\Omega'_X, \mathcal{A}'_X) = (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ , then  $X$  is called numerical.

**Remark 5.2 (Measurability of Inverse Images)** Equation (5.1) implies that all inverse images

$$X^{-1}(A') := \{\omega \in \Omega: X(\omega) \in A'\}, \quad A' \in \mathcal{A}'_X,$$

are elements of the  $\sigma$ -algebra  $\mathcal{A}$  on  $\Omega$ . Because the measure  $P: \mathcal{A} \rightarrow [0, 1]$  assigns a probability to *all* elements of  $\mathcal{A}$ , the probabilities  $P[X^{-1}(A')]$  of these inverse images are determined by  $P$  (see Exercises 5-1 and 5-2).  $\triangleleft$

**Definition 5.3 (Distribution of a Random Variable)**

Suppose that  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  is a random variable. Then the function  $P_X: \mathcal{A}'_X \rightarrow [0, 1]$  defined by

$$P_X(A') = P[X^{-1}(A')], \quad \forall A' \in \mathcal{A}'_X, \quad (5.2)$$

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

**Remark 5.4 (Notation  $P(X \in A')$  and  $P(X=x)$ )** If  $A' \in \mathcal{A}'_X$ , we use the notation

$$P(X \in A') := P[X^{-1}(A')] \quad (5.3)$$

for the probability of the event  $\{X \in A'\} = X^{-1}(A')$ , i. e., the event that  $X$  takes on a value in the subset  $A'$  of  $\Omega'_X$ . If  $\{x\} \in \mathcal{A}'_X$ ,

$$P(X=x) := P[X^{-1}(\{x\})] \quad (5.4)$$

for the probability of the event  $\{X=x\} = X^{-1}(\{x\}) = \{\omega \in \Omega: X(\omega) = x\}$ . If we write  $P(X=x)$ , then we always assume  $\{x\} \in \mathcal{A}'_X$ , even if not mentioned explicitly.  $\triangleleft$

**Remark 5.5 (A New Probability Space)** Definition 5.1 implies that *every* random variable  $X$  on a probability space  $(\Omega, \mathcal{A}, P)$  has a distribution  $P_X$ . Furthermore,  $P_X: \mathcal{A}'_X \rightarrow [0, 1]$  is also a measure, the *image measure of  $P$  under  $X$*  (see Th. 2.79 and Def. 2.80). Because  $P_X(\Omega'_X) = P(\Omega) = 1$ , we can conclude that  $P_X$  is a probability measure, and  $(\Omega'_X, \mathcal{A}'_X, P_X)$  is also a probability space. Therefore, we use the notation

$$X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X, P_X)$$

expressing that

- (a)  $X: \Omega \rightarrow \Omega'_X$  is a random variable on the probability space  $(\Omega, \mathcal{A}, P)$ ,
- (b)  $\mathcal{A}'_X$  is a  $\sigma$ -algebra on  $\Omega'_X$ , and

(c)  $P_X$  is the distribution of  $X$ .

◁

**Definition 5.6 (Identically Distributed Random Variables)**

Let  $X: (\Omega^{(1)}, \mathcal{A}^{(1)}, P^{(1)}) \rightarrow (\Omega', \mathcal{A}')$  and  $Y: (\Omega^{(2)}, \mathcal{A}^{(2)}, P^{(2)}) \rightarrow (\Omega', \mathcal{A}')$  be random variables. If  $P_X = P_Y$ , then we say that  $X$  and  $Y$  are identically distributed.

Note that, oftentimes,  $(\Omega^{(1)}, \mathcal{A}^{(1)}, P^{(1)}) = (\Omega^{(2)}, \mathcal{A}^{(2)}, P^{(2)})$ . Now we consider the distribution of a composition  $g(X)$  of a random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  and a measurable function  $g: (\Omega'_X, \mathcal{A}'_X) \rightarrow (\Omega', \mathcal{A}')$ . According to Remark 5.5, the mapping  $g$  is a random variable on the probability space  $(\Omega'_X, \mathcal{A}'_X, P_X)$ . Furthermore, according to the following lemma,  $g(X)$  is a random variable on  $(\Omega, \mathcal{A}, P)$  and the distribution of  $g(X)$  is the image measure of  $P_X$  under  $g$ . The notation of this image measure is  $(P_X)_g$ .

**Lemma 5.7 (Distribution of a Composition)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable and  $g: (\Omega'_X, \mathcal{A}'_X) \rightarrow (\Omega', \mathcal{A}')$  a measurable function. Then the composition  $g(X): (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  is a random variable and

$$(P_X)_g = P_{g(X)}. \tag{5.5}$$

(Proof p. 188)

**Example 5.8 (Indicator (Variable) of an Event)** If  $(\Omega, \mathcal{A}, P)$  is a probability space and  $A \in \mathcal{A}$ , then the mapping  $1_A: (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  is a random variable. It is called the *indicator (variable) of  $A$* . The distribution of  $1_A$  is

$$P_{1_A}(\{0\}) = P(A^c), \quad P_{1_A}(\{1\}) = P(A), \quad P_{1_A}(\{0, 1\}) = P(\Omega) = 1, \quad P_{1_A}(\emptyset) = P(\emptyset) = 0.$$

If we consider the same event  $A$  and the measurable space  $(\mathbb{R}, \mathcal{B})$ , then we can also write  $1_A: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  in order to express that  $1_A$  is also  $(\mathcal{A}, \mathcal{B})$ -measurable. Note, however, that now the distribution of  $1_A$  is a probability measure on  $(\mathbb{R}, \mathcal{B})$ , and for all  $B \in \mathcal{B}$ ,

$$\begin{aligned} P_{1_A}(B) &= P[1_A^{-1}(B)] = P[\{\omega \in \Omega: 1_A(\omega) \in B\}] && [(5.2), (2.2)] \\ &= \begin{cases} P(\emptyset) = 0, & \text{if } 0 \notin B, 1 \notin B, \\ P(A), & \text{if } 0 \notin B, 1 \in B, \\ P(A^c), & \text{if } 0 \in B, 1 \notin B, \\ P(\Omega) = 1, & \text{if } \{0, 1\} \subset B. \end{cases} \end{aligned}$$

◁

**Example 5.9 (Indicator of an Inverse Image)** If  $(\Omega, \mathcal{A}, P)$  is a probability space,  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  a random variable, and  $A' \in \mathcal{A}'_X$ , then  $1_{X^{-1}(A')}: (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  is a random variable on  $(\Omega, \mathcal{A}, P)$  and

$$1_{X \in A'} := 1_{X^{-1}(A')} = 1_{A'}(X) = 1_{A'} \circ X \quad (5.6)$$

(see Exercise 5-3). The distribution of  $1_{X \in A'}$  is

$$\begin{aligned} P_{1_{X \in A'}}(\{0\}) &= P(X \notin A'), & P_{1_{X \in A'}}(\{1\}) &= P(X \in A'), \\ P_{1_{X \in A'}}(\{0, 1\}) &= P(\Omega) = 1, & P_{1_{X \in A'}}(\emptyset) &= P(\emptyset) = 0. \end{aligned}$$

◁

**Example 5.10 (Dichotomous Random Variable)** Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable on  $(\Omega, \mathcal{A}, P)$ . Then  $X$  is called *dichotomous with values  $x_1$  and  $x_2$*  if  $\{x_1\}, \{x_2\} \in \mathcal{A}'_X$ ,  $P(X \in \{x_1, x_2\}) = 1$  and  $0 < P(X=x_1) < 1$ . If  $X$  is dichotomous with values 0 and 1, then  $X \stackrel{p}{=} 1_{X=1}$ . ◁

**Example 5.11 (Flipping two Coins – continued)** In Examples 2.2.2 and 2.83, we considered flipping two coins and defined  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{P}(\Omega'_X))$ , a random variable representing with its values the *number of flipping heads*. Its possible values are 0, 1, or 2. Hence, we can choose  $\Omega'_X := \{0, 1, 2\}$  and

$$\begin{aligned} P(X=0) &= P_X(\{0\}) = P[X^{-1}(\{0\})] = P[\{(t, t)\}] = \frac{1}{4}, \\ P(X=1) &= P_X(\{1\}) = P[X^{-1}(\{1\})] = P[\{(h, t), (t, h)\}] = \frac{1}{2}, \\ P(X=2) &= P_X(\{2\}) = P[X^{-1}(\{2\})] = P[\{(h, h)\}] = \frac{1}{4} \end{aligned}$$

are the probabilities assigned to the singletons  $\{0\}$ ,  $\{1\}$ , and  $\{2\}$ , whereas

$$\begin{aligned} P(X \in \{0, 1\}) &= P_X(\{0, 1\}) = P[X^{-1}(\{0, 1\})] = P[\{(t, t), (h, t), (t, h)\}] = \frac{3}{4}, \\ P(X \in \{0, 2\}) &= P_X(\{0, 2\}) = P[X^{-1}(\{0, 2\})] = P[\{(t, t), (h, h)\}] = \frac{2}{4}, \\ P(X \in \{1, 2\}) &= P_X(\{1, 2\}) = P[X^{-1}(\{1, 2\})] = P[\{(h, t), (t, h), (h, h)\}] = \frac{3}{4} \end{aligned}$$

are the probabilities assigned to the sets  $\{0, 1\}$ ,  $\{0, 2\}$ , and  $\{1, 2\}$ , which consist of two elements of  $\Omega'_X$ . Finally,  $P_X(\Omega'_X) = P[X^{-1}(\Omega'_X)] = P(\Omega) = 1$  and  $P_X(\emptyset) = P[X^{-1}(\emptyset)] = P(\emptyset) = 0$ . ◁

**Example 5.12 (Tom, Jim, and Kate)** Now we consider an example that is similar to the experiment with Joe and Ann. However, the set of persons is now  $\Omega_U := \{Tom, Jim, Kate\}$ , and we consider three treatments, the elements of the set  $\Omega_X := \{Con, BTh, PTh\}$ , where *Con* could be *no treatment*. The random experiment consists of: drawing a unit  $u$  from the set  $\Omega_U$ , assigning it to one of the three

**Table 5.1.** Tom, Jim, and Kate

Elements of $\Omega$			Random variables			
Unit	Treatment	Success	Probabilities of elementary events $P(\{\omega\})$	Person variable $U$	Treatment variable $X$	Outcome variable $Y$
<i>(Tom, Con, -)</i>			10/99	<i>Tom</i>	0	0
<i>(Tom, Con, +)</i>			10/99	<i>Tom</i>	0	1
<i>(Tom, BTh, -)</i>			2/99	<i>Tom</i>	1	0
<i>(Tom, BTh, +)</i>			6/99	<i>Tom</i>	1	1
<i>(Tom, PTh, -)</i>			1/99	<i>Tom</i>	2	0
<i>(Tom, PTh, +)</i>			4/99	<i>Tom</i>	2	1
<i>(Jim, Con, -)</i>			5/99	<i>Jim</i>	0	0
<i>(Jim, Con, +)</i>			15/99	<i>Jim</i>	0	1
<i>(Jim, BTh, -)</i>			3/99	<i>Jim</i>	1	0
<i>(Jim, BTh, +)</i>			5/99	<i>Jim</i>	1	1
<i>(Jim, PTh, -)</i>			2/99	<i>Jim</i>	2	0
<i>(Jim, PTh, +)</i>			3/99	<i>Jim</i>	2	1
<i>(Kate, Con, -)</i>			12/99	<i>Kate</i>	0	0
<i>(Kate, Con, +)</i>			8/99	<i>Kate</i>	0	1
<i>(Kate, BTh, -)</i>			5/99	<i>Kate</i>	1	0
<i>(Kate, BTh, +)</i>			3/99	<i>Kate</i>	1	1
<i>(Kate, PTh, -)</i>			4/99	<i>Kate</i>	2	0
<i>(Kate, PTh, +)</i>			1/99	<i>Kate</i>	2	1

treatment conditions *Con*, *BTh*, or *PTh* and observing whether (+) or not (-) a success criterion is reached. Hence, the set of possible outcomes of this random experiment is

$$\Omega := \Omega_U \times \Omega_X \times \Omega_Y = \{ (Tom, Con, -), (Tom, Con, +), \dots, (Kate, PTh, +) \}.$$

It consists of the  $3 \cdot 3 \cdot 2 = 18$  triples  $(u, \omega_X, \omega_Y)$  listed in the first column of Table 5.1. As the set of possible events  $\mathcal{A}$  we consider the power set  $\mathcal{P}(\Omega)$ . This set has  $2^{18} = 262144$  elements, where 18 is the number of elements of  $\Omega$ . The probabilities of the 18 elementary events  $\{\omega\}$ ,  $\omega \in \Omega$ , are displayed in the second column of the table. With these specifications, the probabilities  $P(A)$  of all  $2^{18}$  elements  $A \in \mathcal{A}$  are determined [see Rule (x) of Box 4.1]. Hence, the probability space  $(\Omega, \mathcal{A}, P)$  is completely specified.

Table 5.1 also displays the values of the three random variables  $U: (\Omega, \mathcal{A}, P) \rightarrow [\Omega_U, \mathcal{P}(\Omega_U)]$ ,  $X: (\Omega, \mathcal{A}, P) \rightarrow [\Omega'_X, \mathcal{P}(\Omega'_X)]$ , and  $Y: (\Omega, \mathcal{A}, P) \rightarrow [\Omega'_Y, \mathcal{P}(\Omega'_Y)]$ , where

$\Omega'_X := \{0, 1, 2\}$  and  $\Omega'_Y := \{0, 1\}$ . For the singletons  $\{x\}$ ,  $x \in \Omega'_X$ , the values  $P_X(\{x\}) = P[X^{-1}(\{x\})]$  of the distribution of  $X$  are

$$P_X(\{0\}) = 60/99, \quad P_X(\{1\}) = 24/99, \quad P_X(\{2\}) = 15/99,$$

for the sets that consist of two elements of  $\Omega'_X$ , they are

$$P_X(\{0, 1\}) = 84/99, \quad P_X(\{0, 2\}) = 75/99, \quad P_X(\{1, 2\}) = 39/99,$$

and for  $\Omega'_X$  and  $\emptyset$ , they are  $P_X(\Omega'_X) = 1$  and  $P_X(\emptyset) = 0$ .

For the singletons  $\{u\}$ ,  $u \in \Omega_U$ , the values  $P_U(\{u\}) = P[U^{-1}(\{u\})]$  of the distribution of  $U$  are

$$P_U(\{Tom\}) = P_U(\{Jim\}) = P_U(\{Kate\}) = 1/3,$$

for the sets that consist of two elements of  $\Omega_U$  they are

$$P_U(\{Tom, Jim\}) = P_U(\{Tom, Kate\}) = P_U(\{Jim, Kate\}) = 2/3,$$

and for  $\Omega_U$  and  $\emptyset$ , they are  $P_U(\Omega_U) = 1$  and  $P_U(\emptyset) = 0$ . ◁

### Time Order Between Random Variables

In Example 5.12 and also in the examples with Joe and Ann, there is a time order between the random variables involved. Obviously, the person variable  $U$  represents events that are *prior* to the events represented by the treatment variable  $X$  and to the events represented by the outcome variable  $Y$ . In the following definition we extend the definitions introduced in section 4.2.2 to random variables.

#### Definition 5.13 (*X is Prior, Simultaneous, Posterior to Y*)

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$ ,  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_Y, \mathcal{A}'_Y)$  be random variables and  $(\mathcal{F}_t, t \in T)$  a filtration in  $\mathcal{A}$ . Then:

- (i)  $X$  is called *prior to Y* (and  $Y$  posterior to  $X$ ) in  $(\mathcal{F}_t, t \in T)$ , if there is an  $s \in T$  such that  $\sigma(X) \subset \mathcal{F}_s$ ,  $\sigma(Y) \not\subset \mathcal{F}_s$ , and a  $t \in T$ ,  $t > s$ , such that  $\sigma(Y) \subset \mathcal{F}_t$ .
- (ii)  $X$  is called *simultaneous to Y* in  $(\mathcal{F}_t, t \in T)$ , if there is a  $t \in T$  such that  $\sigma(X), \sigma(Y) \subset \mathcal{F}_t$  and no  $s \in T$ ,  $s < t$ , such that  $\sigma(X) \subset \mathcal{F}_s$  or  $\sigma(Y) \subset \mathcal{F}_s$ .

**Remark 5.14 (Filtration Generated by a Family of Random Variables)** Definition 5.13 does not presume that all pairs of random variables can be ordered and compared to each other with respect to a filtration  $(\mathcal{F}_t, t \in T)$ . However, given a family  $(X_i, i \in I)$  of random variables where  $I \subset \mathbb{R}$ , we can define a filtration  $(\mathcal{F}_i, i \in I)$  by  $\mathcal{F}_i := \sigma(X_j, j \leq i)$  for all  $i \in I$ . This filtration is called the *filtration generated by*  $(X_i, i \in I)$  (see Klenke, 2013, p. 191). In this filtration  $X_i$  is prior to  $X_j$  if and only if  $i < j$ , where  $i, j \in I$ . ◁

**Example 5.15 (Flipping a Coin Twice)** If we consider the random experiment of flipping the same coin twice, the probability space  $(\Omega, \mathcal{A}, P)$  is identical to the one specified in section 2.2.2 and Example 2.83, where we considered the random experiment of flipping two coins. Hence, the set of possible outcomes is

$$\Omega = \{(h, h), (h, t), (t, h), (t, t)\}.$$

The possible outcome  $(t, h)$  now represents to obtain tails in the first flip and heads in the second flip. Now, for  $i = 1, 2$ , define the random variables  $X_i: \Omega \rightarrow \{0, 1\}$  by

$$X_i[(a_1, a_2)] := \begin{cases} 1, & \text{if } a_i = h, \\ 0, & \text{if } a_i = t, \end{cases} \quad \forall (a_1, a_2) \in \Omega. \quad (5.7)$$

Hence, the value 1 of  $X_i$  indicates that the outcome of the  $i$ th flip is heads. If we define the filtration  $(\mathcal{F}_1, \mathcal{F}_2)$  by  $\mathcal{F}_1 := \sigma(X_1)$  and  $\mathcal{F}_2 := \sigma(X_1, X_2)$ , then  $X_1$  is prior to  $X_2$  in the filtration  $(\mathcal{F}_1, \mathcal{F}_2)$ . Hence, this filtration serves to introduce time order between the first and the second flip of the coin. It can be shown, e. g., that  $X_1$  is also prior to  $X_1 \cdot X_2$  in  $(\mathcal{F}_1, \mathcal{F}_2)$  (see Exercise 5-4).  $\triangleleft$

**Example 5.16 (Joe and Ann With Random Assignment – continued)** In Example 5.37, the random variable  $U$  is prior to  $X$  in the filtration  $(\mathcal{F}_t, t \in \{1, 2, 3\})$  specified in Example 4.19, because  $\sigma(U) \subset \mathcal{F}_1$ ,  $\sigma(X) \not\subset \mathcal{F}_1$ , and  $\sigma(X) \subset \mathcal{F}_2$ . Analogously it can be shown that, in this example,  $X$  is prior to  $Y$ .  $\triangleleft$

## 5.2 Equivalence of Two Random Variables With Respect to a Probability Measure

### 5.2.1 Identical and $P$ -Equivalent Random Variables

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  be two random variables. Then  $X$  and  $Y$  are called *identical* if

$$\forall \omega \in \Omega: X(\omega) = Y(\omega). \quad (5.8)$$

**Remark 5.17 ( $P$ -Equivalent Random Variables)** Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  be two random variables. Then  $X$  and  $Y$  are *almost surely identical* with respect to  $P$  or  *$P$ -equivalent*, denoted  $X \stackrel{P}{=} Y$ , if

$$\exists A \in \mathcal{A}: (\forall \omega \in \Omega \setminus A: X(\omega) = Y(\omega) \quad \text{and} \quad P(A) = 0) \quad (5.9)$$

(see Def. 2.68). Another notation for  $X \stackrel{P}{=} Y$  is  $X(\omega) \stackrel{P\text{-a.a.}}{=} Y(\omega)$ , which is a shortcut for

$$X(\omega) = Y(\omega), \quad \text{for } P\text{-a.a. } \omega \in \Omega, \quad (5.10)$$

meaning that the values of  $X$  and  $Y$  are identical for  $P$ -almost all  $\omega \in \Omega$  (see Rem. 2.70).  $\triangleleft$

**Remark 5.18 (Singleton With a Positive Probability)** If  $X \stackrel{P}{=} Y$  or, equivalently, if  $X(\omega) \stackrel{P-a.a.}{=} Y(\omega)$ , and  $\{\omega^*\} \in \mathcal{A}$ , with  $P(\{\omega^*\}) > 0$ , then  $X(\omega^*) = Y(\omega^*)$  [see Rem. 2.71].  $\triangleleft$

**Example 5.19 (Indicator of a Null Set)** Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $A \in \mathcal{A}$ . If  $P(A) = 0$ , then

$$1_A \stackrel{P}{=} 0 \quad \text{and} \quad 1_{A^c} = 1 - 1_A \stackrel{P}{=} 1 \quad (5.11)$$

(see Example 5.8).  $\triangleleft$

**Remark 5.20 (Q-Equivalence)** Note that the definition of equivalence of two random variables  $X$  and  $Y$  with respect to a probability measure only presumes that  $X$  and  $Y$  are measurable with respect to a  $\sigma$ -algebra on  $\Omega$  and that the measure considered is a probability measure on this  $\sigma$ -algebra. Hence, we can consider the equivalence of  $X$  and  $Y$  with respect to different probability measures, say  $P$  and  $Q$  and study their relationship.  $\triangleleft$

In the following lemma we consider the relationship between  $P$ -equivalence and  $Q$ -equivalence, presuming  $Q \ll_{\mathcal{A}} P$  (absolute continuity), i. e., presuming

$$\forall C \in \mathcal{C}: P(C) = 0 \Rightarrow Q(C) = 0$$

(see Def. 3.70).

**Lemma 5.21 ( $P$ -Equivalence and  $Q$ -Equivalence)**

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  be random variables. If  $Q$  is a probability measure on  $(\Omega, \mathcal{A})$  such that  $Q \ll_{\mathcal{A}} P$ , then  $X \stackrel{P}{=} Y$  implies  $X \stackrel{Q}{=} Y$ .

(Proof p. 188)

According to Lemma 4.32 (i),  $P^B \ll_{\mathcal{A}} P$ , provided that  $B \in \mathcal{A}$  is an event for which  $P(B) > 0$ . Hence, Lemma 5.21 immediately implies the following corollary:

**Corollary 5.22 ( $P$ -Equivalence Implies  $P^B$ -Equivalence)**

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  be two random variables and let  $B \in \mathcal{A}$  with  $P(B) > 0$ . Then  $X \stackrel{P}{=} Y$  implies  $X \stackrel{P^B}{=} Y$ .

**Example 5.23 ( $P^B$ -Equivalence Does not Imply  $P$ -Equivalence)** Consider the set  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$  with the  $\sigma$ -algebra  $\mathcal{A} = \mathcal{P}(\Omega)$ , and the set  $\Omega' = \{a, b, c, d\}$ , with the  $\sigma$ -algebra  $\mathcal{A}' = \mathcal{P}(\Omega')$ . Furthermore, let  $P: \mathcal{A} \rightarrow [0, 1]$  satisfy  $P(\{\omega_1\}) = .25$ ,  $P(\{\omega_2\}) = .25$ ,  $P(\{\omega_3\}) = 0$ , and  $P(\{\omega_4\}) = .50$ . Finally, define  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  by

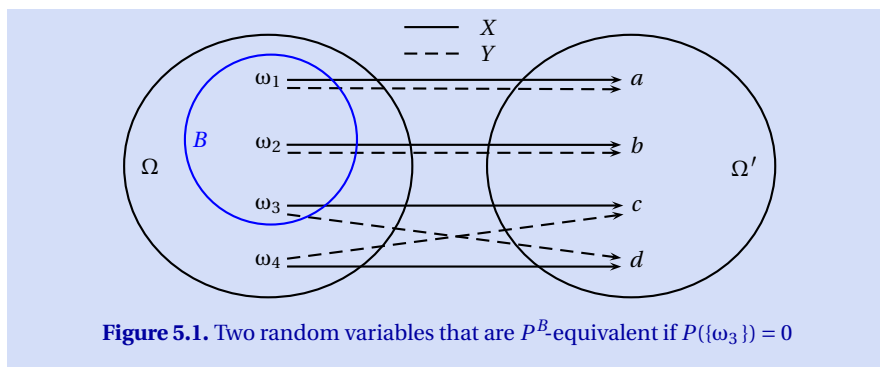


Figure 5.1. Two random variables that are  $P^B$ -equivalent if  $P(\{\omega_3\}) = 0$

$$X(\omega) = \begin{cases} a, & \text{if } \omega = \omega_1 \\ b, & \text{if } \omega = \omega_2 \\ c, & \text{if } \omega = \omega_3 \\ d, & \text{if } \omega = \omega_4 \end{cases} \quad Y(\omega) = \begin{cases} a, & \text{if } \omega = \omega_1 \\ b, & \text{if } \omega = \omega_2 \\ d, & \text{if } \omega = \omega_3 \\ c, & \text{if } \omega = \omega_4 \end{cases}$$

(see Fig. 5.1). If  $B = \{\omega_1, \omega_2, \omega_3\}$ , then  $X \stackrel{P^B}{=} Y$ , but neither  $X=Y$  nor  $X \stackrel{P}{=} Y$  (see Exercise 5-5). Therefore, equivalence with respect to  $P^B$  does not imply equivalence with respect to  $P$ .  $\triangleleft$

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 5.24 ( $P$ -Equivalence Implies Equal Distributions)**

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  be random variables with distributions  $P_X$  and  $P_Y$ , respectively. If  $X \stackrel{P}{=} Y$ , then  $P_X = P_Y$ .

In other words, if  $X$  and  $Y$  are  $P$ -equivalent, then they are identically distributed. Note, however, that identical distributions of  $X$  and  $Y$  do not imply that  $X$  and  $Y$  are  $P$ -equivalent.

In chapter 6 we shall see that Corollary 5.24 implies that also the expectations, variances, and other moments of  $X$  and  $Y$  are identical if  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  are  $P$ -equivalent numerical random variables, provided that the expectations, variances, and other moments of  $X$  and  $Y$  exist.

The following corollary is an immediate implication of Theorem 2.86.

**Corollary 5.25 ( $P$ -Equivalence of Compositions)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable with distribution  $P_X$  and let  $g, g^*: (\Omega'_X, \mathcal{A}'_X) \rightarrow (\mathbb{R}, \mathcal{B})$  be measurable functions. Then:

$$(i) \quad g(X) \stackrel{P}{=} g^*(X) \Leftrightarrow g \stackrel{P_X}{=} g^*.$$

$$(ii) \quad g(X) \stackrel{P}{<} g^*(X) \Leftrightarrow g \stackrel{P_X}{<} g^*.$$

$$(iii) \quad g(X) \stackrel{P}{\leq} g^*(X) \Leftrightarrow g \stackrel{P_X}{\leq} g^*.$$

(Proof p. 188)

**Remark 5.26 (Alternative Notations)** Note that

$$g \stackrel{P_X}{=} g^* \Leftrightarrow g(x) = g^*(x), \quad \text{for } P_X\text{-a.a. } x \in \Omega'_X, \quad (5.12)$$

$$g \stackrel{P_X}{<} g^* \Leftrightarrow g(x) < g^*(x), \quad \text{for } P_X\text{-a.a. } x \in \Omega'_X, \quad (5.13)$$

$$g \stackrel{P_X}{\leq} g^* \Leftrightarrow g(x) \leq g^*(x), \quad \text{for } P_X\text{-a.a. } x \in \Omega'_X. \quad (5.14)$$

&lt;

### 5.2.2 $P$ -Equivalence, $P^B$ -Equivalence, and Absolute Continuity

Now we consider the relationship between equivalence of two random variables  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  with respect to  $P$  and  $P^B$ , and absolute continuity of  $P_X$  with respect to  $P_X^B$ , the distribution of  $X$  with respect to the conditional-probability measure  $P^B$ . Remember, for  $B \in \mathcal{A}$  and  $P(B) > 0$ , we defined the  $B$ -conditional probability measure  $P^B$  (see Def. 4.29). Referring to such a measure,  $X \stackrel{P^B}{=} Y$  means

$$\exists A \in \mathcal{A}: \left( \forall \omega \in \Omega \setminus A: X(\omega) = Y(\omega) \quad \text{and} \quad P^B(A) = 0 \right), \quad (5.15)$$

[see (5.9)]. If  $B$  denotes the event  $\{X=x\} := \{\omega \in \Omega: X(\omega)=x\}$ , then we define  $P^{X=x} := P^B$  and call it the  $(X=x)$ -conditional probability measure on  $(\Omega, \mathcal{A})$ .

#### Lemma 5.27 (An Implication of Absolute Continuity)

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  be random variables that are measurable with respect to the  $\sigma$ -algebra  $\mathcal{C} \subset \mathcal{A}$ , and assume  $\{X \neq Y\} \in \mathcal{C}$ . Furthermore, let  $B \in \mathcal{A}$  with  $P(B) > 0$ . If  $X \stackrel{P^B}{=} Y$  and  $P \ll_{\mathcal{C}} P^B$ , then  $X \stackrel{P}{=} Y$ .

(Proof p. 189)

**Example 5.28 (No Treatment For Joe)** Consider Table 5.2. In this example we define the set  $\Omega_U = \{Joe, Jim, Ann\}$  and

$$\mathcal{A}_U = \mathcal{P}(\Omega_U) = \{\{Joe\}, \{Jim\}, \{Ann\}, \{Joe, Jim\}, \{Joe, Ann\}, \{Jim, Ann\}, \Omega_U, \emptyset\}.$$

Using these sets, not only  $U: (\Omega, \mathcal{A}, P) \rightarrow (\Omega_U, \mathcal{A}_U)$  is a random variable, but also  $U^*: (\Omega, \mathcal{A}, P) \rightarrow (\Omega_U, \mathcal{A}_U)$  defined in Table 5.2. Now the distribution of  $U$  is specified by  $P_U(\{Joe\}) = .5$ ,  $P_U(\{Jim\}) = 0$ , and  $P_U(\{Ann\}) = .5$ . The probabilities of the

other five elements of  $\mathcal{A}_U$  are obtained using Rule (x) of Box 4.1. Furthermore,  $P_{U^*} = P_U$ .

Considering the measure  $P^{X=0}$  we find  $U \underset{P^{X=0}}{=} U^*$ , because

$$P^{X=0}(\{U \neq U^*\}) = P^{X=0}(\{(Joe, yes, +), (Joe, yes, -)\}) = 0.$$

Furthermore, there are only two sets  $A \in \sigma(U)$  with  $P^{X=0}(A) = 0$ . These are the sets  $U^{-1}(\{Jim\}) = \{(Joe, yes, +), (Joe, yes, -)\}$  and  $\emptyset$ , and for these sets we find  $P(U^{-1}(\{Jim\})) = P(\emptyset) = 0$ . Hence,  $P \ll_{\sigma(U)} P^{X=0}$ , and according to Lemma 5.27 this implies  $U \underset{P}{=} U^*$ . In fact, we find

$$P(\{U \neq U^*\}) = P(\{(Joe, yes, +), (Joe, yes, -)\}) = 0.$$

◁

### Lemma 5.29 (Absolute Continuity)

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable and  $B \in \mathcal{A}$  with  $P(B) > 0$ . Then

$$P \ll_{\sigma(X)} P^B \Leftrightarrow P_X \ll_{\mathcal{A}'_X} P_X^B. \quad (5.16)$$

(Proof p. 189)

**Example 5.30 (No Treatment For Joe – continued)** In Example 5.28, we already found

$$P \ll_{\sigma(U)} P^{X=0}.$$

There are only two sets  $A' \in \mathcal{A}_U$  with  $P_U^{X=0}(A') = 0$ , the sets  $\{Jim\}$  and  $\emptyset$ , and for these sets we find  $P_U(\{Jim\}) = P_U(\emptyset) = 0$ . Hence, in this example,

$$P_U \ll_{\mathcal{A}_U} P_U^{X=0}$$

holds as well.

◁

Lemmas 5.29 and 4.32 immediately imply the following corollary.

### Corollary 5.31 (Null-Set Equivalence)

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable and  $B \in \mathcal{A}$  with  $P(B) > 0$ . Then  $P$  and  $P^B$  are null-set equivalent on  $(\Omega, \sigma(X))$  if and only if  $P_X$  and  $P_X^B$  are null-set equivalent on  $(\Omega'_X, \mathcal{A}'_X)$ .

According to Lemma 4.32, absolute continuity of  $P_X^B$  with respect to  $P_X$  always holds. In other words,  $P_X^B \ll_{\mathcal{A}'_X} P_X$ , which is equivalent to

$$\forall A' \in \mathcal{A}'_X: P_X(A') = 0 \Rightarrow P_X^B(A') = 0, \quad (5.17)$$

always holds. In contrast,  $P_X \ll_{\mathcal{A}'_X} P_X^B$  is *not necessarily true*.

**Table 5.2.** No Treatment for Joe

Outcomes $\omega$			Observables						
Unit	Treatment	Success							
			$P(\{\omega\})$	Person variable $U$	Person variable $U^*$	Treatment variable $X$	Outcome variable $Y$		
							$P^{X=0}(\{\omega\})$ (rounded)	$P^{X=1}(\{\omega\})$	
(Joe, no, -)			.152	Joe	Joe	0	0	.245	0
(Joe, no, +)			.348	Joe	Joe	0	1	.561	0
(Joe, yes, -)			0	Joe	Jim	1	0	0	0
(Joe, yes, +)			0	Joe	Jim	1	1	0	0
(Ann, no, -)			.096	Ann	Ann	0	0	.155	0
(Ann, no, +)			.024	Ann	Ann	0	1	.039	0
(Ann, yes, -)			.228	Ann	Ann	1	0	0	.60
(Ann, yes, +)			.152	Ann	Ann	1	1	0	.40

**Example 5.32 (No Treatment For Joe – continued)** Table 5.2 displays an example illustrating absolute continuity of  $P_U$  with respect to  $P_U^B$  for a discrete random variable  $U$ . Consider the event  $B = \{X=1\} = \{\omega \in \Omega: X(\omega) = 1\}$ . Using this notation,  $P_U$  is not absolutely continuous with respect to  $P_U^{X=1}$ , i. e.,  $P_U \ll_{\mathcal{A}_U} P_U^{X=1}$  does *not hold*. In contrast,  $P_U \ll_{\mathcal{A}_U} P_U^{X=0}$  *does hold*.

In this example, the eight elements of  $\Omega$  are listed in the first column of the table. Furthermore, we choose  $\mathcal{A} = \mathcal{P}(\Omega)$  and the probability measure on  $(\Omega, \mathcal{A})$  is specified by the probabilities of the singletons  $\{\omega\}$  specified in the second column of the table [see Box 4.1 (x)]. The random variables  $U: (\Omega, \mathcal{A}, P) \rightarrow (\Omega_U, \mathcal{A}_U)$ , with  $\Omega_U = \{Joe, Ann\}$ ,  $\mathcal{A}_U = \mathcal{P}(\Omega_U)$ , and  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{P}(\Omega'))$  with  $\Omega' = \{0, 1\}$ , are specified in columns 3, 5, and 6 of Table 5.2. (The random variable  $U^*$  has been used in Example 5.28.) Note that the distribution of  $U$  is:

$$\begin{aligned}
 P_U(\{Joe\}) &= P(\{(Joe, no, -), (Joe, no, +), (Joe, yes, +), (Joe, yes, -)\}) = .5, \\
 P_U(\{Ann\}) &= P(\{(Ann, no, -), (Ann, no, +), (Ann, yes, +), (Ann, yes, -)\}) = .5, \\
 P_U(\Omega_U) &= 1, \quad P_U(\emptyset) = 0.
 \end{aligned}$$

Now, we compute the  $(X=1)$ -conditional probabilities of the elementary events:

$$\begin{aligned}
 P^{X=1}(\{\omega_1\}) &= P^{X=1}(\{(Joe, no, -)\}) = \frac{P(\{(Joe, no, -)\} \cap \{X=1\})}{P(X=1)} \\
 &= 0 / (.228 + .152) = 0,
 \end{aligned}$$

and the same result is obtained for  $\omega_2$  to  $\omega_6$ . In contrast,

$$\begin{aligned} P^{X=1}(\{\omega_7\}) &= P^{X=1}(\{(Ann, yes, -)\}) = \frac{P(\{(Ann, yes, -)\} \cap \{X=1\})}{P(X=1)} \\ &= .228 / (.228 + .152) = .60, \end{aligned}$$

and

$$\begin{aligned} P^{X=1}(\{\omega_8\}) &= P^{X=1}(\{(Ann, yes, +)\}) = \frac{P(\{(Ann, yes, +)\} \cap \{X=1\})}{P(X=1)} \\ &= .152 / (.228 + .152) = .40. \end{aligned}$$

These results are displayed in the last column of Table 5.2 and the last but one column shows the probabilities  $P^{X=0}(\{\omega\})$  of the singletons with respect to  $P^{X=0}$ .

Now consider the element  $\{Joe\} \in \mathcal{A}_U$ . Inspecting the last and the second columns of Table 5.2 shows that

$$P_U^{X=1}(\{Joe\}) = 0 \quad \text{and} \quad P_U(\{Joe\}) = .5.$$

According to Definition 3.70 (i), this implies that  $P_U \ll_{\mathcal{A}_U} P_U^{X=1}$  does *not hold*. In contrast, none of the four elements  $A' \in \mathcal{A}_U$  satisfies

$$P_U^{X=0}(A') = 0 \quad \text{and} \quad P_U(A') \neq 0.$$

Therefore, in this example,  $P_U \ll_{\mathcal{A}_U} P_U^{X=0}$  *does hold*. ◁

### 5.3 Multivariate Random Variable

Univariate random variables take on their values in sets such as  $\Omega' \subset \bar{\mathbb{R}}$ ,  $\Omega' = \{male, female\}$ , or  $\Omega' = \{low, medium, high\}$ , whereas bivariate random variables take on their values in sets such as  $\Omega' \subset \bar{\mathbb{R}}^2$  or

$$\Omega' = \{male, female\} \times \{low, medium, high\}.$$

The values of bivariate random variables are pairs such as (5,8) or (male, low). The values of  $n$ -variate random variables are  $n$ -tuples. If  $X$  takes on values such as *male* or (male, low), then we call  $X$  *qualitative*. If  $X$  takes on values in a subset of  $\mathbb{R}^n$ ,  $n \in \mathbb{N}$ , we call it  *$n$ -variate real-valued*. If  $X$  takes on values in a subset of  $\bar{\mathbb{R}}^n$ ,  $n \in \mathbb{N}$ , we call it  *$n$ -variate numerical*.

**Remark 5.33 (Joint and Marginal Distributions)** Definition 5.1 also applies to an  $n$ -variate random variable  $X$ , i. e., to a random variable

$$X = (X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow \left( \prod_{i=1}^n \Omega'_i, \bigotimes_{i=1}^n \mathcal{A}'_i \right) \quad (5.18)$$

that consists of  $n$  univariate random variables  $X_i: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_i, \mathcal{A}'_i)$ . Hence,

$$X(\omega) = [X_1(\omega), \dots, X_n(\omega)], \quad \forall \omega \in \Omega. \quad (5.19)$$

The distribution  $P_X = P_{X_1, \dots, X_n}$  of  $X$  is also called the *joint distribution* of the random variables  $X_i, i = 1, \dots, n$ .

Because  $\pi_i(X_1, \dots, X_n) = X_i$ ,

$$P_{X_i} = P_{\pi_i(X_1, \dots, X_n)}, \quad i = 1, \dots, n, \quad (5.20)$$

[see Eq. (2.20) defining the projection  $\pi_i$ ]. In this context,  $P_{X_i}$  is called the (one-dimensional) *marginal distribution of  $X_i$* . Equation (5.20) shows that the joint distribution uniquely determines all marginal distributions, but not vice versa! More specifically, for  $i = 1, \dots, n$ ,

$$P_{X_i}(A'_i) = P_{X_1, \dots, X_n}(\Omega'_1 \times \dots \times \Omega'_{i-1} \times A'_i \times \Omega'_{i+1} \times \dots \times \Omega'_n), \quad \forall A'_i \in \mathcal{A}'_i. \quad (5.21)$$

Analogously, we may also describe the marginal distribution of  $(X_{i_1}, \dots, X_{i_m})$ , where  $\{i_1, \dots, i_m\} \subset \{1, \dots, n\}$ .  $\triangleleft$

**Remark 5.34 (Joint Distribution vs. Other Quantities)** The joint distribution of a multivariate random variable contains the essential information about the random variables  $X_1, \dots, X_n$ . All other quantities such as expectations  $E(X_i)$ , variances  $\text{Var}(X_i)$ , covariances  $\text{Cov}(X_i, X_j)$ , or conditional expectations such as  $E(X_1 | X_2, \dots, X_n)$ , which are introduced in succeeding chapters, are determined by the joint distribution, and usually they contain less information. Nevertheless, these other quantities often reveal certain properties of a multivariate random variable more clearly than the joint distribution.  $\triangleleft$

**Example 5.35 (Flipping two Coins – continued)** In Examples 2.2.2 and 2.83, we considered the random experiment of flipping two coins and defined a random variable  $X$  representing with its values the number of flipping heads. Additional to  $X$  we may also define the random variables  $X_1, X_2: (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  by

$$X_1(\omega) = \begin{cases} 1, & \text{if } \omega \in \{(h, t), (h, h)\} \\ 0, & \text{if } \omega \in \{(t, h), (t, t)\} \end{cases} \quad (5.22)$$

and

$$X_2(\omega) = \begin{cases} 1, & \text{if } \omega \in \{(t, h), (h, h)\} \\ 0, & \text{if } \omega \in \{(h, t), (t, t)\}. \end{cases} \quad (5.23)$$

They indicate with their value 1 if *heads* are flipped at the first and second flip, respectively. Obviously,  $X = X_1 + X_2$ . Furthermore,

$$(X_1, X_2): (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\} \times \{0, 1\}, \mathcal{P}(\{0, 1\}) \otimes \mathcal{P}(\{0, 1\}))$$

is a two-dimensional random variable with values  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$ , and  $(1, 1)$ . The joint distribution  $P_{X_1, X_2}$  is uniquely defined by

$$P_{X_1, X_2}(\{(i, j)\}) = \frac{1}{4}, \quad \forall i, j = 0, 1.$$

The marginal distribution of  $X_1$  is

$$P_{X_1}(\{i\}) = P_{X_1, X_2}(\{(i, 0)\}) + P_{X_1, X_2}(\{(i, 1)\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \quad i = 0, 1,$$

$P_{X_1}(\{0, 1\}) = 1$ , and  $P_{X_1}(\emptyset) = 0$ . Obviously,  $P_{X_1}$  is completely determined by the joint distribution  $P_{X_1, X_2}$ , and the same applies to the marginal distribution  $P_{X_2}$ .  $\triangleleft$

**Example 5.36 (Tom, Jim, and Kate – continued)** The second column of Table 5.1 also displays the probabilities  $P_{U, X, Y}(\{(u, x, y)\}) = P(\{\omega\})$  of the three-dimensional random variable  $(U, X, Y)$  that maps the elements  $\omega \in \Omega$  onto the set

$$\Omega' := \{Tom, Jim, Kate\} \times \{0, 1, 2\} \times \{0, 1\}$$

on which we consider the  $\sigma$ -algebra

$$\mathcal{A}' := \mathcal{P}(\Omega') = \mathcal{P}(\{Tom, Jim, Kate\}) \otimes \mathcal{P}(\{0, 1, 2\}) \otimes \mathcal{P}(\{0, 1\}).$$

The probabilities  $P_{U, X, Y}(\{(u, x, y)\})$ ,  $(u, x, y) \in \Omega'$ , uniquely determine the joint distribution  $P_{U, X, Y}$  as well as the one-dimensional marginal distributions  $P_U, P_X, P_Y$ , and the two-dimensional marginal distributions  $P_{U, X}, P_{U, Y}$ , and  $P_{X, Y}$ .  $\triangleleft$

For another example of a joint distribution, which refers to Example 2.34 (p. 53) see Exercise 5-6.

## 5.4 Independence of Random Variables

The concepts of independence of events and of set systems, i. e., of sets of events, which have been introduced in Definition 4.40, can be used to define *stochastic independence of random variables*. Remember that

$$\sigma(X) := X^{-1}(\mathcal{A}') := \{X^{-1}(A') : A' \in \mathcal{A}'\}$$

is a  $\sigma$ -algebra on  $\Omega$ , called the  $\sigma$ -algebra generated by  $X$  (see Def. 2.26). Hence, we can define the random variables  $X_1: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_1, \mathcal{A}'_1)$  and  $X_2: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_2, \mathcal{A}'_2)$  to be  $P$ -independent if  $X_1^{-1}(\mathcal{A}'_1)$  and  $X_2^{-1}(\mathcal{A}'_2)$  are  $P$ -independent. In other words,  $X_1$  and  $X_2$  are defined to be  $P$ -independent, if and only if

$$P(A \cap B) = P(A) \cdot P(B), \quad \forall (A, B) \in X_1^{-1}(\mathcal{A}'_1) \times X_2^{-1}(\mathcal{A}'_2). \quad (5.24)$$

Using the notation introduced in Remark 5.4 and

$$P(X_1 \in A', X_2 \in B') := P(\{X_1 \in A'\} \cap \{X_2 \in B'\}), \quad (5.25)$$

this equation is equivalent to

$$P(X_1 \in A', X_2 \in B') = P(X_1 \in A') \cdot P(X_2 \in B'), \quad \forall (A', B') \in \mathcal{A}'_1 \times \mathcal{A}'_2. \quad (5.26)$$

Independence of the random variables  $X_1$  and  $X_2$  with respect to  $P$  is denoted by  $X_1 \perp\!\!\!\perp_P X_2$ .

**Example 5.37 (Joe and Ann With Random Assignment – continued)** In Table 2.2 we presented the random experiment of drawing a person from a set of persons,  $\Omega_U = \{Joe, Ann\}$ , performing a randomized assignment of the drawn person to one of two treatment conditions represented by the elements of the set  $\Omega_X = \{yes, no\}$ , and observing success or failure, represented by the elements of the set  $\Omega_Y = \{-, +\}$ . Hence, the set of possible outcomes of this random experiment is

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

which consists of the eight triples listed in the first column of Table 2.2. In that table we defined the three random variables

$$U: (\Omega, \mathcal{P}(\Omega)) \rightarrow (\Omega_U, \mathcal{P}(\Omega_U)) \quad \text{and} \quad X, Y: (\Omega, \mathcal{P}(\Omega)) \rightarrow (\Omega', \mathcal{A}'),$$

where  $\Omega' = \{0, 1\}$  and  $\mathcal{A}' = \{\{0\}, \{1\}, \Omega', \emptyset\}$ . In order to check if Equation (5.26) actually holds, we choose the two sets  $\{Joe\} \in \mathcal{P}(\Omega_U)$  and  $\{0\} \in \mathcal{A}'$  and compare the probability

$$P(X=0, U=Joe) = P(\{(Joe, no, -), (Joe, no, +)\}) = .3$$

(see the first two rows of Table 2.2) to the product of the two probabilities

$$P(X=0) = P(\{(Joe, no, -), (Joe, no, +), (Ann, no, -), (Ann, no, +)\}) = .6$$

and

$$P(U=Joe) = P(\{(Joe, no, -), (Joe, no, +), (Joe, no, -), (Joe, no, +)\}) = .5.$$

Obviously, Equation (5.26) holds for the pair  $(\{0\}, \{Joe\}) \in \mathcal{A}' \times \mathcal{P}(\Omega_U)$ . Repeating the corresponding comparisons for all pairs of  $\mathcal{A}' \times \mathcal{P}(\Omega_U)$  shows that Equation (5.26) actually holds in this example (see also Exercise 5-7).  $\triangleleft$

**Remark 5.38 (A Methodological Note on Random Assignment)** In random experiments such as the one presented in Example 5.37, with a randomized assignment of the drawn person to one of several treatment conditions, we create independence of  $X$  and the person variable  $U$ . According to Equation 5.26, this implies that we create independence of  $X$  and all  $U$ -measurable mappings  $f(U)$ , because  $\sigma[f(U)] \subset \sigma(U)$  (see Cor. 2.53). More generally, randomized assignment of an observational unit (such as a person) creates independence of  $X$  and all pretreatment variables.  $\triangleleft$

Using definition 4.40, the following definition extends the concept of independence of two random variables to a family of random variables. This includes a finite sequence of random variables  $X_i, i \in I := \{1, \dots, n\}$ , an infinite sequence of random variables  $X_i, i \in I := \{1, 2, \dots\}$ , and a family  $(X_i, i \in I)$  of random variables in which the index set  $I$  may be *any* set, including, e. g.,  $I \subset \mathbb{R}$ .

**Definition 5.39 (Family of Independent Random Variables)**

A family  $(X_i, i \in I)$  of random variables  $X_i: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_i, \mathcal{A}'_i)$  is called a family of  $P$ -independent random variables, denoted  $\perp\!\!\!\perp_P (X_i, i \in I)$ , if  $(X_i^{-1}(\mathcal{A}'_i), i \in I)$  is a family of  $P$ -independent  $\sigma$ -algebras.

**Remark 5.40 (Independence of Three Random Variables)** Hence, three random variables  $X_1, X_2, X_3$  are *independent*, denoted  $\perp\!\!\!\perp_P X_1, X_2, X_3$ , if and only if

$$\begin{aligned} P(X_1 \in A', X_2 \in B', X_3 \in C') &= P(X_1 \in A') \cdot P(X_2 \in B') \cdot P(X_3 \in C'), \\ \forall (A', B', C') \in \mathcal{A}'_1 \times \mathcal{A}'_2 \times \mathcal{A}'_3 \end{aligned} \quad (5.27)$$

(see Rem. 4.38). Note that pairwise independence of  $X_1, X_2, X_3$  follows from choosing  $A' = \Omega'_1$ ,  $B' = \Omega'_2$ , or  $C' = \Omega'_3$ , respectively.  $\triangleleft$

**Remark 5.41 (Independence of  $n$  Random Variables)** Correspondingly, the random variables  $X_1, \dots, X_n$  are *independent*, denoted  $\perp\!\!\!\perp_P X_1, \dots, X_n$ , if and only if

$$\begin{aligned} P(X_1 \in A'_1, \dots, X_n \in A'_n) &= P(X_1 \in A'_1) \cdot \dots \cdot P(X_n \in A'_n), \\ \forall (A'_1, \dots, A'_n) \in \mathcal{A}'_1 \times \dots \times \mathcal{A}'_n. \end{aligned} \quad (5.28)$$

$\triangleleft$

**Remark 5.42 (Sample)** Oftentimes, we assume that  $X_1, \dots, X_n$  is a sequence of independent and identically distributed (abbreviated i. i. d.) random variables (see, e. g., chs. 6 and 8). In statistics, a sequence  $X_1, \dots, X_n$  of i. i. d. random variables is called a *random sample*.

An important example of i. i. d. random variables is treated in the section on Bernoulli trials (see section 8.1.2).  $\triangleleft$

**Remark 5.43 (Independence With Respect to a Probability Measure)** If there is no ambiguity we also use the term *independence* of events, sets of events, random variables, and sets of random variables. Note that, if  $Q$  is another probability measure on  $(\Omega, \mathcal{A})$ , then events, sets of events, and random variables can be  $P$ -independent although they are not  $Q$ -independent.  $\triangleleft$

**Remark 5.44 (A Random Variable and a Set System)** Independence of a set system and a random variable is defined in the same way. A set system  $\mathcal{E} \subset \mathcal{A}$  and a random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  are called *independent*, denoted  $\mathcal{E} \perp\!\!\!\perp_P X$ , if  $\mathcal{E}$  and  $\sigma(X)$  are independent. Of course,  $\mathcal{E}$  can also be a sub- $\sigma$ -algebra of  $\mathcal{A}$ .  $\triangleleft$

**Remark 5.45 (A Random Variable and a Family of Random Variables)** Independence of a random variable  $X$  and a family  $(Y_i, i \in I)$  of random variables, denoted  $X \perp\!\!\!\perp_P (Y_i, i \in I)$ , is defined by  $X \perp\!\!\!\perp_P \sigma(Y_i, i \in I)$  [see Eq. (2.18)]. Note that  $X \perp\!\!\!\perp_P (Y_i, i \in I)$  implies  $X \perp\!\!\!\perp_P \sigma(Y_i)$ , for all  $i \in I$ .  $\triangleleft$

**Remark 5.46 (Equivalent Propositions)** Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $B \in \mathcal{A}$ , and  $\mathcal{C} \subset \mathcal{A}$ . Then the following propositions are equivalent to each other:

$$1_B \perp\!\!\!\perp_P \mathcal{C}, \quad \sigma(\{B\}) \perp\!\!\!\perp_P \mathcal{C}, \quad \{B\} \perp\!\!\!\perp_P \mathcal{C}, \quad B \perp\!\!\!\perp_P \mathcal{C}$$

(see Rem. 4.41 and Exercise 5-8). ◁

In Corollary 5.24 we noted that  $P$ -equivalent random variables have identical distributions. According to the following lemma this also has implications for independence of random variables.

**Lemma 5.47 ( $P$ -Equivalence and Independence)**

Let  $X_i: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ ,  $i = 1, 2$ , and  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega_Y, \mathcal{A}_Y)$  be random variables. Then

$$(X_1 \stackrel{P}{\sim} X_2 \wedge X_1 \perp\!\!\!\perp_P Y) \Rightarrow X_2 \perp\!\!\!\perp_P Y. \quad (5.29)$$

(Proof p. 189)

Now we consider the probability measure  $P$  and the  $B$ -conditional-probability measure  $P^B$  on  $(\Omega, \mathcal{A})$  (see Def. 4.29). In Lemma 4.32 we have shown that  $P^B$  is absolutely continuous (see Def. 3.70) with respect to  $P$  on  $(\Omega, \mathcal{A})$ . In the following lemma we show that  $P$  is absolutely continuous with respect to  $P^B$  on  $(\Omega, \mathcal{C})$ ,  $\mathcal{C} \subset \mathcal{A}$ , provided that  $B$  and  $\mathcal{C}$  are independent.

**Lemma 5.48 (Independence and Absolute Continuity)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space, let  $\mathcal{C} \subset \mathcal{A}$  be a  $\sigma$ -algebra, and  $B \in \mathcal{A}$  with  $P(B) > 0$ . Then  $1_B \perp\!\!\!\perp_P \mathcal{C}$  implies  $P \ll_{\mathcal{C}} P^B$ .

(Proof p. 189)

In the following lemma,  $P_{X_1} \otimes \dots \otimes P_{X_n}$  denotes the product measure of the marginal distributions (see Def. 1.67 and Rem. 5.33).

**Lemma 5.49 (Independence and Product Measure)**

Let  $X = (X_1, \dots, X_n)$  be an  $n$ -variate random variable as specified in (5.18). Then

$$\perp\!\!\!\perp_P X_1, \dots, X_n \Leftrightarrow P_{X_1, \dots, X_n} = P_{X_1} \otimes \dots \otimes P_{X_n}. \quad (5.30)$$

(Proof p. 190)

**Example 5.50 (Tom, Jim, and Kate – continued)** In example 5.12 we considered the random variables  $X$  and  $U$ , which have been constructed such that they are

independent. All  $8 \cdot 8 = 64$  pairs  $(A, B)$  of elements  $A \in X^{-1}(\mathcal{A}'_X)$  and  $B \in U^{-1}(\mathcal{A}'_U)$  satisfy  $P(A \cap B) = P(A) \cdot P(B)$ . Let us consider, e. g.,  $A_1 := X^{-1}(\{0\})$  and  $B_1 := U^{-1}(\{Tom\})$ ,  $B_2 := U^{-1}(\{Jim\})$ , and  $B_3 := U^{-1}(\{Kate\})$ . Then

$$P(A_1 \cap B_j) = \frac{20}{99}, \quad j = 1, 2, 3$$

and

$$P(A_1) \cdot P(B_j) = \frac{60}{99} \cdot \frac{1}{3} = \frac{20}{99}, \quad j = 1, 2, 3.$$

Similarly, considering the events  $A_2 := X^{-1}(\{1\})$  and  $B_j$ ,

$$P(A_2 \cap B_j) = \frac{8}{99}, \quad j = 1, 2, 3$$

and

$$P(A_2) \cdot P(B_j) = \frac{24}{99} \cdot \frac{1}{3} = \frac{8}{99}, \quad j = 1, 2, 3.$$

Finally, considering the events  $A_3 := X^{-1}(\{2\})$  and  $B_j$ , yields

$$P(A_3 \cap B_j) = \frac{5}{99}, \quad j = 1, 2, 3$$

and

$$P(A_3) \cdot P(B_j) = \frac{15}{99} \cdot \frac{1}{3} = \frac{5}{99}, \quad j = 1, 2, 3.$$

Because  $\emptyset$  and all sets  $A \in \mathcal{A}$  are independent, this implies that independence holds for all pairs  $(A, B) \in \{A_1, A_2, A_3, \emptyset\} \times \{B_1, B_2, B_3, \emptyset\}$ . Furthermore, because

- (a)  $\mathcal{E}_1 := \{A_1, A_2, A_3, \emptyset\}$  and  $\mathcal{E}_2 := \{B_1, B_2, B_3, \emptyset\}$  are  $\cap$ -stable set systems on  $\mathcal{A}$ ,
- (b)  $\sigma(\mathcal{E}_1) = X^{-1}(\mathcal{A}'_X)$  and  $\sigma(\mathcal{E}_2) = U^{-1}(\mathcal{A}'_U)$ ,

we can conclude that  $P(A \cap B) = P(A) \cdot P(B)$  holds for *all* elements  $A \in X^{-1}(\mathcal{A}'_X)$  and  $B \in U^{-1}(\mathcal{A}'_U)$  (see Th. 4.44). Therefore, according to Equation (5.24),  $X$  and  $U$  are independent.  $\triangleleft$

**Lemma 5.51 (Independence of a Constant and a Set of Events)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable and  $\mathcal{C} \subset \mathcal{A}$ . If  $X \stackrel{P}{=} \alpha$  and  $\{\alpha\} \in \mathcal{A}'_X$ , then  $X$  and  $\mathcal{C}$  are independent.

(Proof p. 190)

Now we consider mappings of independent random variables. If two random variables  $X_1$  and  $X_2$  are independent and  $f_i: (\Omega'_i, \mathcal{A}'_i) \rightarrow (\Omega''_i, \mathcal{A}''_i)$ ,  $i = 1, 2$ , are measurable mappings, then the two random variables  $f_1(X_1)$  and  $f_2(X_2)$  are independent as well. More generally, if  $f_i: (\Omega'_i, \mathcal{A}'_i) \rightarrow (\Omega''_i, \mathcal{A}''_i)$ ,  $i = 1, \dots, n$ , is a sequence of measurable mappings, then

$$\perp\!\!\!\perp_P f_1(X_1), \dots, f_n(X_n),$$

i. e., then  $f_1(X_1), \dots, f_n(X_n)$  is a sequence of independent random variables on  $(\Omega, \mathcal{A}, P)$ , provided that  $X_1, \dots, X_n$  are independent. In the following theorem we generalize this proposition.

**Theorem 5.52 (Mappings of Families of Independent Random Variables)**

Let  $X_i: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_i, \mathcal{A}'_i)$ ,  $i = 1, \dots, n$ , be random variables and, for  $m \in \mathbb{N}$ , let  $I_1 = \{1, \dots, i_1\}$ ,  $I_2 = \{i_1 + 1, \dots, i_2\}, \dots, I_m = \{i_{m-1} + 1, \dots, n\}$ . Furthermore, let

$$f_j: \left( \prod_{i \in I_j} \Omega'_i, \bigotimes_{i \in I_j} \mathcal{A}'_i \right) \rightarrow (\Omega''_j, \mathcal{A}''_j), \quad j = 1, \dots, m,$$

be measurable mappings. If  $X_1, \dots, X_n$  are independent, then

$$f_1(X_1, \dots, X_{i_1}), f_2(X_{i_1+1}, \dots, X_{i_2}), \dots, f_m(X_{i_{m-1}+1}, \dots, X_n)$$

are independent.

For a generalization and a proof see Bauer (1996, Theorem 9.6, p. 63).

**Example 5.53 (Sums of Independent Random Variables)** Let  $X_1, \dots, X_{2n}$ ,  $n \in \mathbb{N}$ , be independent real-valued random variables, then the  $n$  random variables

$$X_1 + X_2, X_3 + X_4, \dots, X_{2n-1} + X_{2n}$$

are independent as well (see Th. 2.57).  $\triangleleft$

**Example 5.54 (Tom, Jim, and Kate – continued)** In Example 5.50 we showed that the random variables  $X$  and  $U$  are independent. Now we consider the mappings  $f: \{0, 1, 2\} \rightarrow \{0, 1\}$  and  $g: \Omega_U \rightarrow \{male, female\}$  defined by

$$f(x) = \begin{cases} 0, & \text{if } x = 0 \\ 1, & \text{if } x = 1 \text{ or } x = 2 \end{cases}$$

and

$$g(u) = \begin{cases} male, & \text{if } u = Tom \text{ or } u = Jim \\ female, & \text{if } u = Kate, \end{cases}$$

respectively. According to Theorem 5.52 the mappings  $f(X)$  (control vs. any of the two treatments) and  $g(U)$  (*sex*) are independent as well (see Exercise 5-9).  $\triangleleft$

**Remark 5.55 (Conditional Independence of Random Variables Given an Event)**

In chapter 4 we also considered *conditional* independence of events and families of events *given an event*  $B$ . If, for random variables  $X_1, \dots, X_n$  (or, more generally, families of random variables) we consider the set systems  $\sigma(X_1), \dots, \sigma(X_n)$ , then we can use Definition 4.49 in order to define conditional independence of  $X_1, \dots, X_n$  given an event  $B$ , presuming  $P(B) > 0$ . According to Definitions 4.49 and 5.39, conditional independence given  $B$  is equivalent to independence with respect to the probability measure  $P^B$ . In chapter 16 we generalize this concept and study it in more detail.  $\triangleleft$

## 5.5 Probability Function of a Discrete Random Variable

The distribution of a discrete random variable can be described by its *probability function* that is now introduced. Remember, if  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  is a random variable, then the distribution  $P_X$  of  $X$  is a probability measure on  $(\Omega'_X, \mathcal{A}'_X)$ .

### Definition 5.56 (Discrete Random Variable and its Probability Function)

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable and assume that  $\Omega'_0 \subset \Omega'_X$  is finite or countable with  $P_X(\Omega'_0) = 1$  and  $\{x\} \in \mathcal{A}'_X$  for all  $x \in \Omega'_0$ . Then  $X$  and its distribution  $P_X$  are called *discrete*, and the function  $p_X: \Omega'_X \rightarrow [0, 1]$  defined by

$$p_X(x) = \begin{cases} P_X(\{x\}), & \text{if } x \in \Omega'_0, \\ 0, & \text{if } x \in \Omega'_X \setminus \Omega'_0, \end{cases} \quad (5.31)$$

is called the *probability function* of  $X$ .

**Remark 5.57 (Notation)** Note that  $P(X=x) = p_X(x)$ , using the notation introduced in Remark 5.4.  $\triangleleft$

**Remark 5.58 (Probability Function vs. Distribution)** The *distribution*  $P_X$  is defined for *every* random variable, whereas the *probability function*  $p_X$  only applies to *discrete* random variables. While  $P_X$  assigns probabilities to *subsets* of the codomain  $\Omega'_X$  of  $X$ , the probability function  $p_X$  assigns a probability to each *element*  $x$  in  $\Omega'_X$ . Note that  $p_X$  is a real-valued random variable on the probability space  $(\Omega'_X, \mathcal{A}'_X, P_X)$ .  $\triangleleft$

### Remark 5.59 (The Probability Function Uniquely Determines the Distribution)

Note that  $\sigma$ -additivity of the probability measure  $P_X$  implies that  $P_X$  is uniquely determined by the probability function  $p_X$  [see Rule (x) in Box 4.1]. Vice versa, according to Definition 5.56,  $P_X$  defines  $p_X$ . Hence, if  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  are discrete random variables, then

$$p_X = p_Y \Leftrightarrow P_X = P_Y. \quad (5.32)$$

$\triangleleft$

**Remark 5.60 (Probability Function of a Discrete Distribution)** Note that (5.32) allows us to use the term *probability function of a discrete distribution* instead of *probability function of a discrete random variable*.  $\triangleleft$

### Lemma 5.61 (Characterizations of a Discrete Random Variable)

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable.

- (i) Then  $X$  is discrete if and only if there is a finite or countable  $\Omega'_0 \subset \Omega'_X$  such that  $\{x\} \in \mathcal{A}'_X$  for all  $x \in \Omega'_0$ ,  $P_X(\Omega'_0) = 1$ , and

$$1_{X=x_0} \stackrel{p}{=} 1 - \sum_{x \in \Omega'_0 \setminus \{x_0\}} 1_{X=x}, \quad \forall x_0 \in \Omega'_0. \quad (5.33)$$

(ii) Now assume that  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  is real-valued. Then  $X$  is discrete if and only if there is a finite or countable  $\Omega' \subset \mathbb{R}$  such that

$$X \stackrel{p}{=} \sum_{x \in \Omega'} x \cdot 1_{X=x}. \quad (5.34)$$

(Proof p. 191)

**Remark 5.62 (A Caveat)** Note that Equation (5.33) is equivalent to  $P(\Omega'_0) = 1$ . In proposition (ii), we can choose  $\Omega'$  such that  $0 \notin \Omega'$  even if  $P(X=0) > 0$ . In this case, the set  $\Omega'_0$  referred to in (i) can be chosen such that  $\Omega'_0 := \Omega' \cup \{0\}$ .  $\triangleleft$

**Corollary 5.63 (Discrete Real-Valued Random Variable)**

Assume that  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  is a real-valued random variable. Then  $X$  is discrete if and only if the following two conditions hold:

(a)  $\Omega'_> := \{x \in \mathbb{R} : P(X=x) > 0\}$  is finite or countable.

(b)  $X \stackrel{p}{=} \sum_{x \in \Omega'_>} x \cdot 1_{X=x}$ .

(Proof p. 191)

Condition (b) may equivalently be written

$$X \stackrel{p}{=} \sum_{\substack{x \in \mathbb{R} \\ P(X=x) > 0}} x \cdot 1_{X=x}. \quad (5.35)$$

**Example 5.64 (Flipping two Coins – continued)** Consider again Example 5.35 and let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  denote the number of flipping heads. If we assume that  $P(\{\omega\}) = \frac{1}{4}$  for all  $\omega \in \Omega$ , then

$$p_X(0) = P_X(\{0\}) = \frac{1}{4} \quad p_X(1) = P_X(\{1\}) = \frac{1}{2} \quad p_X(2) = P_X(\{2\}) = \frac{1}{4}$$

are the values of the probability function  $p_X$  of  $X$ . They are the probabilities of the events that  $X$  takes on the value 0, 1, and 2, respectively. For simplicity, we also denote these probabilities by  $P(X=0)$ ,  $P(X=1)$ , and  $P(X=2)$ . In this example, we may choose different measurable spaces  $(\Omega'_X, \mathcal{A}'_X)$ . If we choose  $(\Omega'_X, \mathcal{A}'_X) = (\{0, 1, 2\}, \mathcal{P}(\{0, 1, 2\}))$ , then  $\Omega'_0 = \Omega'_X$  (see Def. 5.56). If we choose  $(\Omega'_X, \mathcal{A}'_X) = (\mathbb{R}, \mathcal{B})$ , then  $\Omega'_0 = \{0, 1, 2\}$  and  $\mathcal{A}'_X|_{\Omega'_0} = \mathcal{B}|_{\Omega'_0} = \mathcal{P}(\Omega'_0)$  (see Remark 1.29). According to Equation (5.34),

$$X = 0 \cdot 1_{X=0} + 1 \cdot 1_{X=1} + 2 \cdot 1_{X=2}.$$

This example is a special case of a random variable with a binomial distribution. The general case is treated in Definition 8.7. Other examples of a discrete

random variable and their probability function are random variables that have a *Poisson distribution* or a *geometric distribution*. In both cases, the random variable considered takes on an infinite and countable number of values, each of which has a probability greater than 0. These examples are treated in chapter 8 (see Defs. 8.14 and 8.20).  $\triangleleft$

**Example 5.65 (Flipping two Coins – continued)** In Example 5.35 we introduced the random variables  $X_1$  and  $X_2$ , which indicate if we flip *heads* in the first and second trial, respectively. The probability function of the bivariate random variable  $X = (X_1, X_2)$  is

$$p_{X_1, X_2}(x_1, x_2) = \frac{1}{4}, \quad \forall (x_1, x_2) \in \{0, 1\}^2.$$

 $\triangleleft$ 

**Lemma 5.66 (Probability Function of a Marginal Distribution)**

Consider a multivariate random variable  $X = (X_1, \dots, X_n)$  as specified in (5.18) and assume that there is a finite or countable set  $\Omega'_0 \subset \prod_{i=1}^n \Omega'_i$  with  $P(\Omega'_0) = 1$  and  $\{x\} \in \otimes_{i=1}^n \mathcal{A}'_i$  for all  $x \in \Omega'_0$ . Furthermore, for all  $x_i \in \Omega'_i$ , define

$$\Omega'_{0, x_i} := \{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) : (x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \in \Omega'_0\}.$$

Then, for all  $x_i \in \Omega'_i$ ,

$$p_{X_i}(x_i) = \sum_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \Omega'_{0, x_i}} p_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n), \quad (5.36)$$

where  $p_{X_i}$  denotes the probability function of  $X_i$ ,  $i = 1, \dots, n$ , which is also called the *marginal probability function* of  $X_i$ .

(Proof p. 192)

Now we turn to a condition that is equivalent to independence of discrete random variables.

**Remark 5.67 (Support Sets of Discrete Random Variables)** Under the assumptions of Lemma 5.66 we define the support sets of the distributions  $P_{X_i}$ , i. e., the sets

$$\Omega'_{0, i} := \{x_i \in \Omega'_i : p_{X_i}(x_i) > 0\}, \quad i = 1, \dots, n. \quad (5.37)$$

Obviously,  $\Omega'_{0, i}$  is finite or countable for all  $i = 1, \dots, n$ . Hence,  $\Omega'_{sn} := \prod_{i=1}^n \Omega'_{0, i}$  is finite or countable as well. Furthermore,  $P(X \in \Omega'_{sn}) = 1$ , because, for  $(x_1, \dots, x_n) \in \Omega'_0 \setminus \Omega'_{sn}$ , there is at least one  $i$  such that  $X_i \notin \Omega'_{0, i}$  and therefore  $p_{X_i} = P(X_i = x_i) = 0$ , which implies  $P(X_1 = x_1, \dots, X_i = x_i, \dots, X_n = x_n) = 0$ .  $\triangleleft$

**Lemma 5.68 (A Condition Equivalent to Independence)**

Let  $X$  be a multivariate random variable as specified in (5.18) and assume that there is a finite or countable set  $\Omega'_0 \subset \prod_{i=1}^n \Omega'_i$  with  $P(\Omega'_0) = 1$  and  $\{x\} \in \mathcal{A}'_X$ , for all  $x \in \Omega'_0$ . Furthermore, let  $p_X, p_{X_1}, \dots, p_{X_n}$  denote the probability functions of  $X, X_1, \dots, X_n$ , respectively, and let  $\Omega'_{0,1}, \dots, \Omega'_{0,n}$  be the sets defined in (5.37). Then  $X_1, \dots, X_n$  are independent if and only if

$$p_X(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n), \quad \forall (x_1, \dots, x_n) \in \prod_{i=1}^n \Omega'_{0,i}. \quad (5.38)$$

*(Proof p. 192)*

Note that, in Lemmas 5.66 and 5.68, the set  $\Omega'_0$  is not necessarily a Cartesian product. We only require that it is a *subset* of a Cartesian product.

In the following section we shall see that a probability function is a special probability density (see Th. 5.77).

## 5.6 Probability Density With Respect to a Measure

Some probability measures can also be described by a density with respect to the Lebesgue measure on  $(\mathbb{R}, \mathcal{B})$  or the counting measure on  $(\Omega, \mathcal{P}(\Omega))$ , where  $\Omega$  is a finite or countable set (see Th. 3.65). Such a density is useful for explicit numerical calculations and comparing distributions to each other. We start by translating some concepts and results of chapter 3 to probability measures.

### 5.6.1 General Concepts and Properties

According to Theorem 3.65 and Definition 3.66, a nonnegative measurable function  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \mathcal{B})$  is called a *density of  $\nu$  with respect to  $\mu$* , if

$$\nu(A) = \int_A f \, d\mu, \quad \forall A \in \mathcal{A}. \quad (5.39)$$

The function  $\nu: \mathcal{A} \rightarrow \overline{\mathbb{R}}$  defined by (5.39) is a measure, which is also denoted by  $f \circ \mu$ . Hence,  $f \circ \mu(A) = \int_A f \, d\mu, \forall A \in \mathcal{A}$ . Theorem 3.65 implies the following corollary.

**Corollary 5.69 (Probability Measure With Density)**

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space. If  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \mathcal{B})$  is a nonnegative  $\mu$ -integrable function with  $\int f \, d\mu = 1$ , then  $P := f \circ \mu$  is a probability measure on  $(\Omega, \mathcal{A})$ .

Corollary 5.69 justifies the following definition.

**Definition 5.70 (Probability Density)**

Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $\mu$  a measure on  $(\Omega, \mathcal{A})$ . If  $P = f \circ \mu$ , then  $f$  is called a probability density of  $P$  with respect to  $\mu$ .

**Remark 5.71 (Probability Density of a Random Variable)** Consider the random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  and let  $\mu$  be a measure on  $(\Omega'_X, \mathcal{A}'_X)$ . If  $P_X = f_X \circ \mu$ , then  $f_X$  is also referred to as a probability density of  $X$  with respect to  $\mu$ .  $\triangleleft$

Applying Equation (5.39) yields the following corollary.

**Corollary 5.72 (Characterizing the Probability Measure by a Density)**

Let  $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$  be  $\mu$ -integrable and nonnegative and  $P$  a probability measure on  $(\Omega, \mathcal{A})$ . Then  $f$  is a (probability) density of  $P$  with respect to  $\mu$  if and only if it satisfies

$$P(A) = \int_A f \, d\mu, \quad \forall A \in \mathcal{A}. \quad (5.40)$$

Theorem 3.68 (a) and (c) imply the following corollary.

**Corollary 5.73 (Probability Densities are  $\mu$ -Equivalent)**

Let  $\mu$  and  $P$  be measures on the measurable space  $(\Omega, \mathcal{A})$ , where  $P$  is a probability measure. If  $f, f^*: (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$  are probability densities of  $P$  with respect to  $\mu$ , then  $f \stackrel{\mu}{=} f^*$ .

Now we translate the Radon-Nikodym Theorem (see Th. 3.72), which yields the following corollary.

**Corollary 5.74 (An Implication of the Radon-Nikodym Theorem)**

Let  $(\Omega, \mathcal{A}, \mu)$  be a measure space. If  $\mu$  is  $\sigma$ -finite,  $P$  is a probability measure on  $(\Omega, \mathcal{A})$ , and  $P \ll \mu$ , then there is a probability density  $f$  of  $P$  with respect to  $\mu$  (also called a Radon-Nikodym derivative), i. e.,

$$f = \frac{dP}{d\mu}. \quad (5.41)$$

**Example 5.75 (Conditional-Probability Measure)** In Lemma 4.32 (ii) we showed that  $1_B/P(B)$  is a density of  $P^B$  with respect to  $P$ , and according to Remark 5.71 it is a probability density of  $P^B$ .  $\triangleleft$

### 5.6.2 Density of a Discrete Random Variable

As a special case, we consider a discrete random variable (see section 5.5).

**Remark 5.76 (A Sum of Dirac Measures)** Let the assumptions of Definition 5.56 hold, i. e., let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable and suppose  $\Omega'_0 \subset \Omega'_X$  is finite or countable with  $P_X(\Omega'_0) = 1$  and  $\{x\} \in \mathcal{A}'_X$  for all  $x \in \Omega'_0$ . Furthermore, define

$$\mu := \sum_{x \in \Omega'_0} \delta_x, \quad (5.42)$$

the sum of Dirac measures at  $x$ ,  $x \in \Omega'_0$ . According to Example 1.57,  $\mu$  is a measure on  $(\Omega'_X, \mathcal{A}'_X)$ , which is  $\sigma$ -finite.  $\triangleleft$

The following theorem asserts that the probability function  $p_X$  is a density of  $P_X$  with respect to  $\mu$ .

#### Theorem 5.77 (The Probability Function is a Density)

Let the assumptions of Definition 5.56 be satisfied and let  $\mu$  be defined by Equation (5.42). Then:

- (i)  $P_X \ll_{\mathcal{A}'_X} \mu$ .
- (ii) The probability function  $p_X$  is a density of  $P_X$  with respect to  $\mu$ , i. e.,

$$p_X = \frac{dP_X}{d\mu} \quad (5.43)$$

and, for all  $A' \in \mathcal{A}'_X$ ,

$$P_X(A') = \int_{A'} p_X d\mu \quad (5.44)$$

$$= \sum_{x \in \Omega'_0} \mathbf{1}_{A'}(x) \cdot p_X(x) \quad (5.45)$$

$$= \sum_{x \in A'} p_X(x). \quad (5.46)$$

(Proof p. 193)

Hence, each probability  $P_X(A')$ ,  $A' \in \mathcal{A}'_X$ , can be computed from the probability function  $p_X$ .

### 5.6.3 Density of a Bivariate Random Variable

Now we consider bivariate random variables. However, extending the following notation and propositions to general multivariate random variables is straightforward.

**Lemma 5.78 (Absolute Continuity of Marginal Distributions)**

Let  $(X, Y): (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X \times \Omega'_Y, \mathcal{A}'_X \otimes \mathcal{A}'_Y)$  be a random variable and suppose that  $(\Omega'_X, \mathcal{A}'_X, \mu)$  and  $(\Omega'_Y, \mathcal{A}'_Y, \nu)$  are  $\sigma$ -finite measure spaces. Then

$$P_{X,Y} \ll_{\mathcal{A}'_X \otimes \mathcal{A}'_Y} \mu \otimes \nu \Rightarrow P_X \ll_{\mathcal{A}'_X} \mu \text{ and } P_Y \ll_{\mathcal{A}'_Y} \nu.$$

(Proof p. 193)

Let

$$f_{X,Y} := \frac{dP_{X,Y}}{d\mu \otimes \nu}, \quad f_X := \frac{dP_X}{d\mu}, \quad f_Y := \frac{dP_Y}{d\nu}$$

denote Radon-Nikodym derivatives (see Th. 3.72). In the following lemma, we use the notation ' $\stackrel{\mu}{\equiv}$ ' introduced in Remark 2.70.

**Lemma 5.79 (Marginal Densities)**

Let  $(X, Y): (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X \times \Omega'_Y, \mathcal{A}'_X \otimes \mathcal{A}'_Y)$  be a random variable,  $(\Omega'_X, \mathcal{A}'_X, \mu)$ ,  $(\Omega'_Y, \mathcal{A}'_Y, \nu)$  be measure spaces,  $\mu, \nu$  be  $\sigma$ -finite, and assume  $P_{X,Y} \ll_{\mathcal{A}'_X \otimes \mathcal{A}'_Y} \mu \otimes \nu$ .

Then

$$f_X(x) \stackrel{\mu}{\equiv} \int f_{X,Y}(x, y) \nu(dy), \quad (5.47)$$

and

$$f_Y(y) \stackrel{\nu}{\equiv} \int f_{X,Y}(x, y) \mu(dx). \quad (5.48)$$

The functions  $f_X$  and  $f_Y$  are also called *marginal densities* of  $X$  and  $Y$ , respectively.

(Proof p. 194)

**Remark 5.80 (Marginal and Joint Density)** Suppose that  $X$  and  $Y$  are real-valued. Then for the Lebesgue measure  $\mu = \nu = \lambda$  and a Riemann integrable density  $f_{X,Y}$ , Equations (5.47) and (5.48) yield

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \quad (5.49)$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx, \quad (5.50)$$

respectively (see Th. 3.62).  $\triangleleft$ 

Note that, for a discrete random variable  $X = (X_1, \dots, X_n)$ , the probability function  $p_X$  is a density of  $P_X$  with respect to the measure  $\mu$  specified in Equation (5.42), and the marginal probability functions  $p_{X_i}$ ,  $i = 1, \dots, n$ , are marginal densities. In section 5.7.4 we consider multivariate densities with respect to the Lebesgue measure  $\mu = \nu = \lambda$ .

## 5.7 Uni- or Multivariate Real-Valued Random Variable

The remaining sections of this chapter show how to describe distributions of *real-valued* uni- and multivariate random variables.

### 5.7.1 Distribution Function of a Univariate Real-Valued Random Variable

If we consider a univariate real-valued random variable  $X$ , then the *distribution function*  $F_X$  assigns to each  $x \in \mathbb{R}$  the probability  $P(X \leq x)$  of the event  $\{X \leq x\} = \{\omega \in \Omega: X(\omega) \leq x\}$  that  $X$  takes on a value *smaller or equal* than  $x$ . As we shall see, the distribution function uniquely determines the distribution  $P_X$ .

#### Definition 5.81 (Distribution Function)

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}, P_X)$  denote a real-valued random variable. Then the (cumulative) distribution function  $F_X: \mathbb{R} \rightarrow [0, 1]$  of  $X$  is defined by:

$$F_X(x) := P_X(]-\infty, x]) = P(X \leq x), \quad \forall x \in \mathbb{R}. \quad (5.51)$$

**Remark 5.82 (Probabilities of Intervals)** This definition implies that we can compute the probability  $P(a < X \leq b)$  of  $X$  taking a value in the interval  $]a, b]$  by

$$P(a < X \leq b) = F_X(b) - F_X(a), \quad \text{if } a < b, \quad (5.52)$$

because

$$P(a < X \leq b) = P_X(]-\infty, b] \setminus ]-\infty, a]) = P_X(]-\infty, b]) - P_X(]-\infty, a])$$

[see Rule (vi), Box 4.1]. ◁

**Remark 5.83 (The Distribution Function Determines the Distribution)** Every random variable  $X$  has a distribution  $P_X$ . Therefore, the distribution function  $F_X$  exists for all real-valued random variables. The distribution function uniquely determines the distribution  $P_X$  of a real-valued random variable, because a finite measure on  $(\Omega, \mathcal{A})$  is already uniquely specified by its values on a  $\cap$ -stable generating system of  $\mathcal{A}$  (see Th. 1.72) and the set system  $\{]-\infty, x]: x \in \mathbb{R}\}$  is a  $\cap$ -stable generating system of  $\mathcal{B}$ , the Borel  $\sigma$ -algebra on  $\mathbb{R}$  [see Eq. (1.19)]. Hence,  $P_X$  uniquely determines  $F_X$ , which implies the following theorem. ◁

#### Theorem 5.84 (Uniqueness)

Let  $P_X, P_Y$  denote the distributions and  $F_X, F_Y$  the distribution functions of two real-valued random variables  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ . Then

$$F_X = F_Y \Leftrightarrow P_X = P_Y. \quad (5.53)$$

This theorem facilitates describing distributions and calculations involving distributions considerably, because the distribution function is defined on the set  $\mathbb{R}$  of real numbers, whereas  $P_X$  is defined on a much more complex domain, the Borel  $\sigma$ -algebra  $\mathcal{B}$ .

**Example 5.85 (Flipping two Coins – continued)** In Example 5.11, we considered flipping two coins and specified the distribution  $P_X$  of  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{P}(\Omega'_X))$ , representing with its values  $x$  the number of flipping heads. The distribution  $P_X$  assigns a probability to all  $2^3 = 8$  subsets of  $\Omega'_X := \{0, 1, 2\}$ . Because  $\{0, 1, 2\} \subset \mathbb{R}$ , the random variable  $X$  is also a random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  with values in  $\mathbb{R}$ . In this case,  $F_X$  is a step function defined by

$$F_X(x) = P_X([-\infty, x]) = P(X \leq x) = \begin{cases} 0, & \text{if } x < 0 \\ 1/4, & \text{if } 0 \leq x < 1 \\ 3/4, & \text{if } 1 \leq x < 2 \\ 1, & \text{if } x \geq 2. \end{cases}$$

According to Theorem 5.84 the distribution  $P_X$  is uniquely defined by these four values. In other words, if we know these four values, then we know the probabilities  $P_X(A')$  for all elements  $A'$  of the Borel  $\sigma$ -algebra  $\mathcal{B}$  (see Exercise 5-10).  $\triangleleft$

Now we turn to the *quantile function*, which, in some cases, is the inverse of the distribution function. Sometimes this function is also called the *pseudo-inverse* of  $F_X$ . It assigns to each  $p \in [0, 1]$  the smallest real number  $x$  for which  $P(X \leq x) = F_X(x) \geq p$ .

**Definition 5.86 (Quantile Function)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a real-valued random variable with distribution function  $F_X$ . Then the quantile function  $Q_X: [0, 1] \rightarrow \bar{\mathbb{R}}$  of  $X$  is defined by:

$$\begin{aligned} Q_X(p) &= \inf \{x \in \mathbb{R} : F_X(x) \geq p\}, \quad \forall p \in ]0, 1[, \\ Q_X(0) &= \begin{cases} \inf \{x \in \mathbb{R} : F_X(x) > 0\}, & \text{if } \exists x \in \mathbb{R} \text{ with } F_X(x) = 0 \\ -\infty, & \text{if } F_X(x) > 0, \forall x \in \mathbb{R}, \end{cases} \quad (5.54) \\ Q_X(1) &= \begin{cases} \inf \{x \in \mathbb{R} : F_X(x) = 1\}, & \text{if } \exists x \in \mathbb{R} \text{ with } F_X(x) = 1 \\ \infty, & \text{if } F_X(x) < 1, \forall x \in \mathbb{R}. \end{cases} \end{aligned}$$

A value  $Q_X(p)$ ,  $p \in [0, 1]$ , is called the *quantile of  $p$  with respect to  $F_X$* .

**Remark 5.87 (Inverse Function of  $F_X$ )** If  $F_X$  is continuous and strictly monotone, i. e., if  $x_1 < x_2$  implies  $F_X(x_1) < F_X(x_2)$ , then

$$Q_X(p) = F_X^{-1}(p), \quad \forall p \in ]0, 1[. \quad (5.55)$$

where  $F_X^{-1}$  denotes the inverse function of  $F_X$ .  $\triangleleft$

**Example 5.88 (Flipping two Coins – continued)** In Example 5.85, we specified the distribution function of  $X := \text{number of flipping heads}$  for the random experiment of flipping two coins. The corresponding quantile function takes on the three values

$$Q_X(p) = \begin{cases} 0, & \text{if } 0 \leq p \leq 1/4 \\ 1, & \text{if } 1/4 < p \leq 3/4 \\ 2, & \text{if } 3/4 < p \leq 1. \end{cases}$$

&lt;

### 5.7.2 Distribution Function of a Multivariate Real-Valued Random Variable

Now we extend the concept of a distribution function to the multivariate case. In the following definition we use the notation introduced in Equation (5.25).

**Definition 5.89 (Joint Distribution Function)**

Let  $(X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  be a random variable. Its distribution function  $F_{X_1, \dots, X_n}: \mathbb{R}^n \rightarrow [0, 1]$  is defined by

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) := P(X_1 \leq x_1, \dots, X_n \leq x_n), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n. \quad (5.56)$$

It is also called the joint distribution function of  $X_1, \dots, X_n$ .

**Example 5.90 (Flipping two Coins – continued)** In example 5.35, we considered flipping two coins and defined the random variables  $X_1$  and  $X_2$  indicating whether or not we flip *heads* at first and second flip, respectively. In this example, the bivariate distribution function  $F_{X_1, X_2}$  takes on the values

$$F_{X_1, X_2}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 < 0 \text{ or } x_2 < 0, \\ 1/4 & \text{if } 0 \leq x_1 < 1, 0 \leq x_2 < 1, \\ 2/4 & \text{if } x_1 \geq 1, 0 \leq x_2 < 1, \\ 2/4 & \text{if } 0 \leq x_1 < 1, x_2 \geq 1, \\ 1 & \text{if } x_1 \geq 1, x_2 \geq 1. \end{cases}$$

&lt;

Just like in Theorem 5.84 we can prove uniqueness, using a  $\cap$ -stable generating system for  $\mathcal{B}_n$ , now referring to Equation (1.21).

**Theorem 5.91 (Uniqueness)**

Let  $P_X, P_Y$  denote the distributions and  $F_X, F_Y$  the distribution functions of two  $n$ -variate real-valued random variables  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$ . Then

$$F_X = F_Y \Leftrightarrow P_X = P_Y. \quad (5.57)$$

As an implication of Equation (5.21) with  $A_i' = ] - \infty, X_i]$  and  $\Omega_j' = \mathbb{R}$ ,  $i \neq j$ , we obtain the next corollary. In the special case of a bivariate real-valued random variable  $(X_1, X_2)$ , this corollary asserts that the value of the *marginal distribution function of  $X_1$*  for the argument  $x_1$ , i. e.,  $\lim_{x_2 \rightarrow \infty} F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1)$ . In this corollary,

$$\lim_{\substack{x_j \rightarrow \infty \\ j \neq i}} F_{X_1, \dots, X_n}(x_1, \dots, x_n), \quad (5.58)$$

denotes the limit of the distribution function of  $(X_1, \dots, X_n)$  for  $x_j \rightarrow \infty$ , for all  $j = 1, \dots, i-1, i+1, \dots, n$ . This limit is the value of the *marginal distribution function of  $X_i$*  for the argument  $x_i$ , and the corollary asserts that this limit is identical to the value  $F_{X_i}(x_i)$  of the distribution function of  $X_i$  for the argument  $x_i$ .

**Corollary 5.92 (Joint and Marginal Distribution Function)**

Let  $(X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  be a random variable. Then

$$F_{X_i}(x_i) = \lim_{\substack{x_j \rightarrow \infty \\ j \neq i}} F_{X_1, \dots, X_n}(x_1, \dots, x_n), \quad \forall x_i \in \mathbb{R}. \quad (5.59)$$

(Proof p. 194)

The next corollary shows how independence of the random variables  $X_1, \dots, X_n$  can be formulated in terms of their distribution functions.

**Corollary 5.93 (Independence and Joint Distribution Function)**

Let  $(X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  be a random variable. Then the following two propositions are equivalent to each other:

- (a)  $X_1, \dots, X_n$  are independent
- (b)  $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n)$ ,  $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$ .

(Proof p. 194)

**5.7.3 Density of a Continuous Univariate Real-Valued Random Variable**

As a special case, we consider a random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  for which there is a nonnegative measurable function  $f_X: (\mathbb{R}, \mathcal{B}, P_X) \rightarrow (\mathbb{R}, \mathcal{B})$  such that

$$P_X(B) = \int_B f_X d\lambda, \quad \forall B \in \mathcal{B}, \quad (5.60)$$

where  $\lambda$  denotes the Lebesgue measure on  $(\mathbb{R}, \mathcal{B})$ . According to Theorem 3.68 (ii), this equation is equivalent to  $P_X = f_X \circ \lambda$ .

The following definition is a special case of Definition 5.70.

**Definition 5.94 (Continuous Random Variable and its Density)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a real-valued random variable with distribution  $P_X$ . We call  $X$  continuous, if there is a nonnegative function  $f_X: \mathbb{R} \rightarrow \mathbb{R}$  that is integrable with respect to the Lebesgue measure  $\lambda$  and

$$P_X = f_X \circ \lambda. \quad (5.61)$$

A function  $f_X$  satisfying (5.61) is called a (probability) density of  $X$ .

Note that Equation (5.61) is equivalent to

$$F_X(x) = \int_{]-\infty, x]} f_X d\lambda, \quad \forall x \in \mathbb{R}, \quad (5.62)$$

because

$$F_X(x) = P_X(]-\infty, x]) = \int 1_{]-\infty, x]} dP_X = \int_{]-\infty, x]} f_X d\lambda, \quad \forall x \in \mathbb{R}, \quad (5.63)$$

(see Th. 3.67). Furthermore, Theorem 3.62 immediately implies the following corollary:

**Corollary 5.95 (Riemann Integral of the Density)**

If  $f_X$  is a density of the random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  and  $f_X$  is Riemann integrable, then

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad \forall x \in \mathbb{R}. \quad (5.64)$$

**Remark 5.96 (Interpretation of Densities)** Note that the term  $f_X(t)$  in Equation (5.64) is not a probability, instead it is a value of the density for  $t \in \mathbb{R}$ . However, the probability  $P(a < X \leq b)$  that  $X$  takes on a value in the interval  $]a, b]$  can be computed using Equation (5.52) and the density  $f_X$ , provided that it exists and that it is Riemann integrable:

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx, \quad \text{if } a < b. \quad (5.65)$$

This probability can be represented as the area between the density and the  $x$ -axis above the interval  $[a, b]$  (see Fig. 3.6).  $\triangleleft$

**Remark 5.97 (Continuity of  $X$  Implies  $P(X = x) = 0$ )** Consider a continuous random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ . Definition 5.94 and Remark 3.71 imply that  $P_X \ll \lambda$ . Because  $\lambda(\{x\}) = 0$  [see Eq. (1.53)] we can conclude  $P(X=x) = 0$ , for all  $x \in \mathbb{R}$ . Hence, additivity of  $P$  yields, for all  $a, b \in \mathbb{R}$ ,  $a < b$ ,

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b), \quad (5.66)$$

provided that  $X$  is continuous.  $\triangleleft$

**Example 5.98 (Continuous Random Variables and Their Densities)** In section 8.2 we present some examples of continuous random variables and their densities, such as the densities of normal distributions, central  $\chi^2$ -distributions, central  $t$ -distributions, and central  $F$ -distributions.  $\triangleleft$

### 5.7.4 Density of a Continuous Multivariate Real-Valued Random Variable

**Remark 5.99 (Multivariate Case)** Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  be a multivariate random variable with distribution  $P_X$ . If  $f_X: \mathbb{R}^n \rightarrow \mathbb{R}$  is nonnegative and integrable with respect to the Lebesgue measure  $\lambda_n$  on  $(\mathbb{R}^n, \mathcal{B}_n)$ , and  $P_X = f_X \circ \lambda_n$ , then  $X$  is continuous with probability density  $f_X$ , and

$$F_X(x_1, \dots, x_n) = \int_B f_X d\lambda_n, \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad (5.67)$$

where  $B := \times_{i=1}^n ]-\infty, x_i]$ . If  $f_X$  is Riemann integrable, then

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f_X(t_1, \dots, t_n) dt_1 \dots dt_n, \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n. \quad (5.68)$$

More generally, for any  $B \in \mathcal{B}_n$ ,

$$P_X(B) = P(X \in B) = \int 1_B \cdot f_X d\lambda_n. \quad (5.69)$$

The probability  $P_X(B)$  can be represented as the  $(n+1)$ -dimensional *volume* between the joint density and the  $(x_1, \dots, x_n)$ -hyperplane above  $B$  (see Fig. 8.8).  $\triangleleft$

In analogy to Corollary 5.93, independence of continuous real-valued random variables can also be formulated in terms of probability densities, using the marginal densities  $f_{X_1}, \dots, f_{X_n}$  (see Lemma 5.79).

#### Corollary 5.100 (Independence and Probability Densities)

Let  $(X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$ ,  $n > 1$ , be a random variable and suppose that all random variables  $X_i$ ,  $i = 1, \dots, n$ , have a density  $f_{X_i}$  with respect to the Lebesgue measure  $\lambda$  on  $(\mathbb{R}, \mathcal{B})$ . Then  $X_1, \dots, X_n$  are independent if and only if

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) := f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad (5.70)$$

is a joint density of  $(X_1, \dots, X_n)$  with respect to  $\lambda_n$ .

(Proof p. 195)

**Remark 5.101 (Independence, Densities, and Riemann Integrals)** If all densities  $f_{X_i}$ ,  $i = 1, \dots, n$ , are Riemann integrable, then  $f_{X_1, \dots, X_n}$  in (5.70) is Riemann integrable as well (see, e. g., Ellis & Gulick, 2006).  $\triangleleft$

**Remark 5.102 (Other Random Variables)** In this section we considered univariate and multivariate real-valued random variables  $X$ . All these random variables have a distribution  $P_X$  and also a (cumulative) distribution function  $F_X$ . If  $X$  is discrete, then its distribution can be described by the probability function, and the distribution function  $F_X$  is a step function (see Example 5.90 and Fig. 8.1). If  $X$  is continuous, then its distribution can be specified by a density with respect to the Lebesgue measure. The distribution functions of continuous random variables do not have discontinuities or jumps (see Rem. 5.97 and Fig. 8.7 for an example). Note, however, that there are random variables that are neither discrete nor continuous. Their distribution functions have jumps but are not step functions.  $\triangleleft$

## 5.8 Proofs

### *Proof of Lemma 5.7*

According to Theorem 2.49, the composition  $g(X)$  of  $X$  and  $g$  is  $(\mathcal{A}, \mathcal{A}')$ -measurable. Therefore, it is a random variable on  $(\Omega, \mathcal{A}, P)$ . For all  $A \in \mathcal{A}$ ,

$$\begin{aligned}
 P_{g(X)}(A) &= P\{[g(X)]^{-1}(A)\} && [(5.2)] \\
 &= P\{\{\omega \in \Omega: g[X(\omega)] \in A\}\} && [\text{Def. 2.1}] \\
 &= P\{\{\omega \in \Omega: X(\omega) \in g^{-1}(A)\}\} && [\text{Def. 2.1}] \\
 &= P\{X^{-1}[g^{-1}(A)]\} && [\text{Def. 2.1}] \\
 &= P_X[g^{-1}(A)] && [(5.2)] \\
 &= (P_X)_g(A). && [(5.2)]
 \end{aligned}$$

### *Proof of Lemma 5.21*

$$\begin{aligned}
 X \stackrel{P}{=} Y &\Rightarrow \exists A \in \mathcal{A}: (\forall \omega \in \Omega \setminus A: X(\omega) = Y(\omega) \text{ and } P(A) = 0) && [(5.9)] \\
 &\Rightarrow \exists A \in \mathcal{A}: (\forall \omega \in \Omega \setminus A: X(\omega) = Y(\omega) \text{ and } Q(A) = 0) && [Q \ll_{\mathcal{A}} P] \\
 &\Rightarrow X \stackrel{Q}{=} Y. && [(5.9)]
 \end{aligned}$$

### *Proof of Corollary 5.25*

Proposition (i) is a special case of Theorem 2.86. Analogously to the proof of Theorem 2.86, we can prove Propositions (ii) and (iii) using Remark 2.67.

**Proof of Lemma 5.27**

Let  $A := \{\omega \in \Omega: X(\omega) \neq Y(\omega)\} \in \mathcal{C}$ . Hence, if  $X \stackrel{P^B}{=} Y$ , then the conjunction of  $P^B(A) = 0$  and  $P \ll_{\mathcal{C}} P^B$  implies  $P(A) = 0$ .

**Proof of Lemma 5.29**

$P \ll_{\sigma(X)} P^B \Rightarrow P_X \ll_{\mathcal{A}'_X} P^B_X$ : If  $P \ll_{\sigma(X)} P^B$ , then, for all  $C' \in \mathcal{A}'_X$ :

$$\begin{aligned} P^B_X(C') = 0 &\Rightarrow P^B[X^{-1}(C')] = 0 && [(5.2)] \\ &\Rightarrow P[X^{-1}(C')] = 0 && [P \ll_{\sigma(X)} P^B] \\ &\Rightarrow P_X(C') = 0. && [(5.2)] \end{aligned}$$

$P_X \ll_{\mathcal{A}'_X} P^B_X \Rightarrow P \ll_{\sigma(X)} P^B$ : If  $P_X \ll_{\mathcal{A}'_X} P^B_X$  and  $C \in \sigma(X)$ , then there is a  $C' \in \mathcal{A}'_X$  such that

$$C = X^{-1}(C').$$

Hence,

$$\begin{aligned} P^B(C) = 0 &\Rightarrow P^B[X^{-1}(C')] = 0 \\ &\Rightarrow P^B_X(C') = 0 && [(5.2)] \\ &\Rightarrow P_X(C') = 0 && [P_X \ll_{\mathcal{A}'_X} P^B_X] \\ &\Rightarrow P[X^{-1}(C')] = 0 && [(5.2)] \\ &\Rightarrow P(C) = 0. \end{aligned}$$

**Proof of Lemma 5.47**

If  $X_1 \stackrel{P}{=} X_2$ , then

$$\begin{aligned} X_1 &\perp\!\!\!\perp_P Y \\ &\Rightarrow \forall (A', B') \in \mathcal{A}' \times \mathcal{A}'_Y: \\ &\quad P[X_1^{-1}(A') \cap Y^{-1}(B')] = P[X_1^{-1}(A')] \cdot P[Y^{-1}(B')] && [(5.24)] \\ &\Rightarrow \forall (A', B') \in \mathcal{A}' \times \mathcal{A}'_Y: \\ &\quad P[X_2^{-1}(A') \cap Y^{-1}(B')] = P[X_2^{-1}(A')] \cdot P[Y^{-1}(B')] \quad [X_1 \stackrel{P}{=} X_2, \text{Def. 5.3, Cor. 5.24}] \\ &\Rightarrow X_2 \perp\!\!\!\perp_P Y. && [(5.24)] \end{aligned}$$

**Proof of Lemma 5.48**

According to Remark 5.46,  $1_B \perp\!\!\!\perp_P \mathcal{C} \Leftrightarrow B \perp\!\!\!\perp_P \mathcal{C}$ . Hence, for all  $C \in \mathcal{C}$ ,

$$\begin{aligned}
P^B(C) = 0 &\Rightarrow P(B \cap C) = 0 \\
&\Rightarrow P(B) \cdot P(C) = 0 && [B \perp\!\!\!\perp_p C] \\
&\Rightarrow P(C) = 0. && [P(B) > 0]
\end{aligned}$$

### **Proof of Lemma 5.49**

$\Rightarrow$  If  $X_1, \dots, X_n$  are independent, then, for all  $A'_i \in \mathcal{A}'_i$ ,  $i = 1, \dots, n$ ,

$$\begin{aligned}
P_{X_1, \dots, X_n}(A'_1 \times \dots \times A'_n) &= P[X_1^{-1}(A'_1) \cap \dots \cap X_n^{-1}(A'_n)] \\
&= P[X_1^{-1}(A'_1)] \cdot \dots \cdot P[X_n^{-1}(A'_n)] \\
&= P_{X_1}(A'_1) \cdot \dots \cdot P_{X_n}(A'_n) \\
&= (P_{X_1} \otimes \dots \otimes P_{X_n})(A'_1 \times \dots \times A'_n).
\end{aligned}$$

Hence, according to the definition of the product measure (see Def. 1.67),

$$P_{X_1, \dots, X_n} = P_{X_1} \otimes \dots \otimes P_{X_n}.$$

$\Leftarrow$  If the right-hand side of (5.30) holds, then, for all  $A'_i \in \mathcal{A}'_i$ ,  $i = 1, \dots, n$ ,

$$\begin{aligned}
P[X_1^{-1}(A'_1) \cap \dots \cap X_n^{-1}(A'_n)] &= P_{X_1, \dots, X_n}(A'_1 \times \dots \times A'_n) \\
&= P_{X_1}(A'_1) \cdot \dots \cdot P_{X_n}(A'_n) \\
&= P[X_1^{-1}(A'_1)] \cdot \dots \cdot P[X_n^{-1}(A'_n)].
\end{aligned}$$

According to Equation (5.28), this implies independence of  $X_1, \dots, X_n$ .

### **Proof of Lemma 5.51**

Assume that  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  is a random variable and that there is an  $\alpha \in \Omega'_X$  and  $\{\alpha\} \in \mathcal{A}'_X$  such that  $X \stackrel{p}{=} \alpha$ . If  $A := \{\omega \in \Omega: X(\omega) \neq \alpha\}$ , then  $A = \Omega \setminus X^{-1}(\{\alpha\}) \in \mathcal{A}$ ,  $P(A) = 0$  and  $P(A^c) = 1$ . This implies, for all  $A' \in \mathcal{A}'_X$ ,

$$\begin{aligned}
P[X^{-1}(A')] &= P[X^{-1}(A') \cap A^c] && [\text{Box 4.1 (viii)}] \\
&= P[\{\omega \in \Omega: X(\omega) \in A', X(\omega) = \alpha\}] \\
&= \begin{cases} P(\emptyset), & \text{if } \alpha \notin A' \\ P(A^c), & \text{if } \alpha \in A' \end{cases} \\
&= \begin{cases} 0, & \text{if } \alpha \notin A' \\ 1, & \text{if } \alpha \in A'. \end{cases} && (5.71)
\end{aligned}$$

This implies, for all  $A' \in \mathcal{A}'_X$  and all  $C \in \mathcal{C}$ ,

$$\begin{aligned}
P[X^{-1}(A') \cap C] &= \begin{cases} 0, & \text{if } \alpha \notin A', \\ P(C), & \text{if } \alpha \in A', \end{cases} && [\text{Box 4.1 (v), (viii)}] \\
&= P[X^{-1}(A')] \cdot P(C). && [(5.71)]
\end{aligned}$$

**Proof of Lemma 5.61**

(i) ( $\Rightarrow$ ) If  $X$  is discrete, then there is a finite or countable  $\Omega'_0 \subset \Omega'_X$  with  $\{x\} \in \mathcal{A}'_X$  for all  $x \in \Omega'_0$  and  $P_X(\Omega'_0) = 1$ . This implies

$$P\left(1 \neq \sum_{x \in \Omega'_0} 1_{X=x}\right) = P(X \notin \Omega'_0) = 0,$$

and therefore

$$1 \stackrel{P}{=} \sum_{x \in \Omega'_0} 1_{X=x}.$$

( $\Leftarrow$ ) If Equation (5.33) holds and  $\Omega'_0 \subset \Omega'_X$  is finite or countable with  $\{x\} \in \mathcal{A}'_X$  for all  $x \in \Omega'_0$ , then  $1 \stackrel{P}{=} \sum_{x \in \Omega'_0} 1_{X=x}$  and therefore

$$1 = P\left(1 = \sum_{x \in \Omega'_0} 1_{X=x}\right) = P(X \in \Omega'_0) = P_X(\Omega'_0).$$

(ii) If  $X$  is real-valued, then  $X$  is discrete if and only if there is a finite or countable  $\Omega' \subset \mathbb{R}$  such that

$$\begin{aligned} X &= X \cdot 1 \\ &\stackrel{P}{=} X \cdot \sum_{x \in \Omega'} 1_{X=x} && \text{[(i)]} \\ &= \sum_{x \in \Omega'} X \cdot 1_{X=x} \\ &= \sum_{x \in \Omega'} x \cdot 1_{X=x}. && [1_{X=x}(\omega) = 0 \text{ if } X(\omega) \neq x, \text{ (5.6)}] \end{aligned}$$

**Proof of Corollary 5.63**

$\Rightarrow$ . If  $X$  is discrete and  $\Omega'_0 \subset \Omega'_X$  (see Def. 5.56), i. e., if  $\Omega'_0$  is finite or countable and  $P_X(\Omega'_0) = 1$ , then  $\Omega'_> \subset \Omega'_0$ , because  $P_X(\Omega'_X \setminus \Omega'_0) = 0$ . Hence,  $\Omega'_>$  is finite or countable as well, which proves (a). Furthermore, for finite or countable sets  $\Omega'_0, \Omega'_>$ :

$$P_X(\Omega'_0 \setminus \Omega'_>) = \sum_{x \in \Omega'_0 \setminus \Omega'_>} P(X=x) = 0. \quad \text{[Def. of } \Omega'_> \text{]}$$

Hence,

$$P_X(\Omega'_X \setminus \Omega'_>) = P_X(\Omega'_X \setminus \Omega'_0) + P_X(\Omega'_0 \setminus \Omega'_>) = 0$$

and, according to (5.11),

$$1_{X \in (\Omega'_X \setminus \Omega'_>)} \stackrel{P}{=} 0. \quad (5.72)$$

Therefore,

$$X = 1_{X \in \Omega'_>} \cdot X + 1_{X \in (\Omega'_X \setminus \Omega'_>)} \cdot X \quad [1 = 1_{X \in \Omega'_>} + 1_{X \in (\Omega'_X \setminus \Omega'_>)}]$$

$$\begin{aligned}
& \stackrel{p}{=} X \cdot \sum_{x \in \Omega'_>} 1_{X=x} && [(5.72), \Omega'_> \text{ is finite or countable}] \\
& = \sum_{x \in \Omega'_>} X \cdot 1_{X=x} \\
& = \sum_{x \in \Omega'_>} x \cdot 1_{X=x}. && [X \cdot 1_{X=x} = x \cdot 1_{X=x}]
\end{aligned}$$

←. This is an immediate implication of Lemma 5.61 (ii).

### **Proof of Lemma 5.66**

If  $x_i \in \Omega'_i$  and  $\Omega'_{0,x_i} = \emptyset$ , then  $p_{X_i}(x_i) = P(X_i = x_i) = 0$ . If  $\Omega'_{0,x_i} \neq \emptyset$ , then it is finite or countable, because  $\Omega'_0$  is finite or countable. Then, using  $P_X(\Omega'_0) = 1$ ,

$$\begin{aligned}
& p_{X_i}(x_i) \\
& = P(X_i = x_i) \\
& = P(X_1 \in \Omega'_1, \dots, X_{i-1} \in \Omega'_{i-1}, X_i = x_i, X_{i+1} \in \Omega'_{i+1}, \dots, X_n \in \Omega'_n) && [\text{Box 4.1 (viii)}] \\
& = P((X_1, \dots, X_n) \in (\Omega'_1 \times \dots \times \Omega'_{i-1} \times \{x_i\} \times \Omega'_{i+1} \times \dots \times \Omega'_n)) \\
& = P((X_1, \dots, X_n) \in ((\Omega'_1 \times \dots \times \Omega'_{i-1} \times \{x_i\} \times \Omega'_{i+1} \times \dots \times \Omega'_n) \cap \Omega'_0)) && [\text{Box 4.1 (viii)}] \\
& = \sum_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \Omega'_{0,x_i}} P(X_1 = x_1, \dots, X_{i-1} = x_{i-1}, X_i = x_i, X_{i+1} = x_{i+1}, \dots, X_n = x_n) && [\text{Box 4.1 (x)}] \\
& = \sum_{(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in \Omega'_{0,x_i}} p_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n). && [(5.31)]
\end{aligned}$$

### **Proof of Lemma 5.68**

If  $X = (X_1, \dots, X_n)$  and  $X_1, \dots, X_n$  are independent, then, for all  $(x_1, \dots, x_n) \in \prod_{i=1}^n \Omega'_{0,i}$ ,

$$p_X(x_1, \dots, x_n) = P((X_1, \dots, X_n) = (x_1, \dots, x_n)) \quad [(5.31)]$$

$$= P(X_1 = x_1, \dots, X_n = x_n)$$

$$= P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n) \quad [(5.28)]$$

$$= p_{X_1}(x_1) \cdot \dots \cdot p_{X_n}(x_n). \quad [(5.31)]$$

For all  $A'_1 \in \mathcal{A}'_1, \dots, A'_n \in \mathcal{A}'_n$ ,

$$(A'_1 \times \dots \times A'_n) \cap (\Omega'_{0,1} \times \dots \times \Omega'_{0,n}) = (A'_1 \cap \Omega'_{0,1}) \times \dots \times (A'_n \cap \Omega'_{0,n}). \quad (5.73)$$

Now assume that Equation (5.38) holds and define  $\Omega'_{sn} := \prod_{i=1}^n \Omega'_{0,i}$ . Then,

$$\begin{aligned}
& P(X_1 \in A'_1, \dots, X_n \in A'_n) \\
& = P[(X_1, \dots, X_n) \in (A'_1 \times \dots \times A'_n)] \\
& = P[(X_1, \dots, X_n) \in (A'_1 \times \dots \times A'_n) \cap \Omega'_{sn}] && [P_X(\Omega'_{sn}) = 1, \text{Box 4.1, (viii)}]
\end{aligned}$$

$$= \sum_{(x_1, \dots, x_n) \in (A'_1 \times \dots \times A'_n) \cap \Omega'_{sn}} P[(X_1, \dots, X_n) = (x_1, \dots, x_n)] \quad [\text{Box 4.1 (x)}]$$

$$= \sum_{(x_1, \dots, x_n) \in (A'_1 \times \dots \times A'_n) \cap \Omega'_{sn}} p_X(x_1, \dots, x_n) \quad [(5.36)]$$

$$= \sum_{(x_1, \dots, x_n) \in (A'_1 \times \dots \times A'_n) \cap \Omega'_{sn}} p_{X_1}(x_1) \cdots p_{X_n}(x_n) \quad [(5.38)]$$

$$= \left( \sum_{x_1 \in A'_1 \cap \Omega'_{0,1}} p_{X_1}(x_1) \right) \cdots \left( \sum_{x_n \in A'_n \cap \Omega'_{0,n}} p_{X_n}(x_n) \right) \quad [(5.73)]$$

$$= P(X_1 \in A'_1) \cdots P(X_n \in A'_n), \quad [P_{X_i}(\Omega'_{0,i}) = 1, \text{Box 4.1 (viii), (x)}]$$

which implies independence of  $X_1, \dots, X_n$  [see (5.28)].

### **Proof of Theorem 5.77**

For  $A' \in \mathcal{A}'_X$ ,

$$P_X(A') = \sum_{x \in A' \cap \Omega'_0} P_X(\{x\}) \quad [P_X(\Omega'_0) = 1, \text{Box 4.1 (viii), (x)}]$$

$$= \sum_{x \in A' \cap \Omega'_0} p_X(x) \quad [(5.31)]$$

$$= \sum_{x \in \Omega'_0} p_X(x) \cdot 1_{A'}(x)$$

$$= \sum_{x \in \Omega'_0} \int p_X \cdot 1_{A'} d\delta_x \quad [(3.23)]$$

$$= \int 1_{A'} \cdot p_X d\left( \sum_{x \in \Omega'_0} \delta_x \right) \quad [(3.53)]$$

$$= \int_{A'} p_X d\left( \sum_{x \in \Omega'_0} \delta_x \right) \quad [(3.30)]$$

$$= \int_{A'} p_X d\mu. \quad [(5.42)]$$

Hence, according to Definition 5.70, the probability function  $p_X$  is a density of  $P_X$  with respect to the measure  $\mu$  on the measurable space  $(\Omega'_X, \mathcal{A}'_X)$  and  $p_X = \frac{dP}{d\mu}$  (see Th. 3.72).

### **Proof of Lemma 5.78**

Let  $A' \in \mathcal{A}'_X$  with  $\mu(A') = 0$ . Then

$$\mu \otimes \nu(A' \times \Omega'_Y) = \mu(A') \cdot \nu(\Omega'_Y) = 0 \cdot \nu(\Omega'_Y) = 0,$$

and this holds even if  $\nu(\Omega'_Y) = \infty$ . Together with  $P_{X,Y} \ll_{\mathcal{A}'_X \otimes \mathcal{A}'_Y} \mu \otimes \nu$  this implies

$$P_X(A') = P(X \in A', Y \in \Omega'_Y) = P_{X,Y}(A' \times \Omega'_Y) = 0.$$

Therefore,  $P_X \ll_{\mathcal{A}'_X} \mu$ . The proof for  $P_Y$  is analogous.

**Proof of Lemma 5.79**

For all  $A' \in \mathcal{A}'_X$ ,

$$\begin{aligned}
& \int 1_{A'}(x) \left( \int f_{X,Y}(x,y) \nu(dy) \right) \mu(dx) \\
&= \int 1_{A'}(x) f_{X,Y}(x,y) \mu \otimes \nu [d(x,y)] \quad [\text{Th. 3.76}] \\
&= \int 1_{A'}(x) P_{X,Y}[d(x,y)] \quad [\text{Th. 3.67}] \\
&= P_{X,Y}(A' \times \Omega'_Y) \quad [(3.30), (3.8)] \\
&= P_X(A'). \quad [(5.21)]
\end{aligned}$$

Theorem 3.65 then implies the lemma. The proof for  $f_Y$  is analogous.

**Proof of Corollary 5.92**

Note that

$$\mathbb{R}^{n-1} = \bigcup_{m=1}^{\infty} ]-\infty, m]^{n-1}. \quad (5.74)$$

For all  $X_i \in \mathbb{R}$ ,

$$\begin{aligned}
F_{X_i}(x_i) &= P(X_i \leq x_i) \\
&= P_{X_1, \dots, X_n}(\mathbb{R} \times \dots \times \mathbb{R} \times ]-\infty, x_i] \times \mathbb{R} \times \dots \times \mathbb{R}) \quad [(5.21)] \\
&= P(X_1 \in \mathbb{R}, \dots, X_{i-1} \in \mathbb{R}, X_i \in ]-\infty, x_i], X_{i+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \quad [(5.2)] \\
&= P\left((X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) \in \mathbb{R}^{n-1}, X_i \in ]-\infty, x_i]\right) \\
&= P\left(X_j \in \bigcup_{m_j=1}^{\infty} ]-\infty, m_j], j \in \{1, \dots, i-1, i+1, \dots, n\}, X_i \in ]-\infty, x_i]\right) \quad [(5.74)] \\
&= \lim_{m_j \rightarrow \infty, j \neq i} P\left(X_j \in ]-\infty, m_j], j \in \{1, \dots, i-1, i+1, \dots, n\}, X_i \in ]-\infty, x_i]\right) \\
&\quad [\text{Th. 1.68 (i)}] \\
&= \lim_{m_j \rightarrow \infty, j \neq i} F_{X_1, \dots, X_n}(m_1, \dots, m_{i-1}, x_i, m_{i+1}, \dots, m_n) \quad [(5.56)] \\
&= \lim_{x_j \rightarrow \infty, j \neq i} F_{X_1, \dots, X_n}(x_1, \dots, x_n).
\end{aligned}$$

The limits exist, because  $F_{X_1, \dots, X_n}$  is monotone in all coordinates.

**Proof of Corollary 5.93**

(a)  $\Rightarrow$  (b) For all  $(x_1, \dots, x_n) \in \mathbb{R}^n$ ,

$$\begin{aligned}
F_{X_1, \dots, X_n}(x_1, \dots, x_n) &= P_{X_1, \dots, X_n}([-\infty, x_1] \times \dots \times ]-\infty, x_n]) \quad [(5.56)] \\
&= P_{X_1} \otimes \dots \otimes P_{X_n}([-\infty, x_1] \times \dots \times ]-\infty, x_n]) \quad [\text{Lemma 5.49 and (a)}] \\
&= P_{X_1}([-\infty, x_1]) \cdot \dots \cdot P_{X_n}([-\infty, x_n]) \quad [(1.50)] \\
&= F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n). \quad [\text{Def. 5.81}]
\end{aligned}$$

(b)  $\Rightarrow$  (a) For all  $(x_1, \dots, x_n) \in \mathbb{R}^n$ ,

$$\begin{aligned}
& P_{X_1, \dots, X_n}([-\infty, x_1] \times \dots \times ]-\infty, x_n]) \\
&= F_{X_1, \dots, X_n}(x_1, \dots, x_n) && [(5.56)] \\
&= F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n) && [(b)] \\
&= P_{X_1}([-\infty, x_1]) \cdot \dots \cdot P_{X_n}([-\infty, x_n]) && [\text{Def. 5.81}] \\
&= P_{X_1} \otimes \dots \otimes P_{X_n}([-\infty, x_1] \times \dots \times ]-\infty, x_n]). && [(1.50)]
\end{aligned}$$

Because  $\{]-\infty, x_1] \times \dots \times ]-\infty, x_n] : (x_1, \dots, x_n) \in \mathbb{R}^n\}$  is a  $\cap$ -stable generating system of  $\mathcal{B}_n$  (see (1.21)), Theorem 1.72 yields  $P_{X_1, \dots, X_n} = P_{X_1} \otimes \dots \otimes P_{X_n}$ . Applying Lemma 5.49 then completes the proof.

### **Proof of Corollary 5.100**

(i)  $(X_1, \dots, X_n)$  are independent  $\Rightarrow f_{X_1, \dots, X_n}$  in (5.70) is a joint density.

Because  $f_{X_1, \dots, X_n}$  defined in (5.70) is nonnegative and integrable with respect to  $\lambda_n$ , we can conclude that  $f_{X_1, \dots, X_n} \odot \lambda_n$  defines a finite measure on  $(\mathbb{R}^n, \mathcal{B}_n)$  (see Th. 3.65). Furthermore, if  $X_1, \dots, X_n$  are independent and  $f_{X_i}$  is a density of  $X_i$  for all  $i = 1, \dots, n$ , then, for all  $(x_1, \dots, x_n) \in \mathbb{R}^n$  and  $B = \prod_{i=1}^n ]-\infty, x_i]$ ,

$$\begin{aligned}
& \int_B f_{X_1, \dots, X_n}(t_1, \dots, t_n) \lambda_n[d(t_1, \dots, t_n)] \\
&= \int_B f_{X_1}(t_1) \cdot \dots \cdot f_{X_n}(t_n) \lambda_n[d(t_1, \dots, t_n)] && [(5.70)] \\
&= \prod_{i=1}^n \int_{]-\infty, x_i]} f_{X_i}(t_i) \lambda(dt_i) && [\text{Th. 3.76}] \\
&= \prod_{i=1}^n F_{X_i}(x_i) && [(5.62)] \\
&= F_{X_1, \dots, X_n}(x_1, \dots, x_n) && [\text{Cor. 5.93}] \\
&= P_{X_1, \dots, X_n}(B).
\end{aligned}$$

This shows that  $f_{X_1, \dots, X_n} \odot \lambda_n(B) = P_{X_1, \dots, X_n}(B)$ , and this implies that  $f_{X_1, \dots, X_n}$  defined in (5.70) is a density of  $(X_1, \dots, X_n)$  with respect to  $\lambda_n$  (see Def. 5.70).

(ii)  $f_{X_1, \dots, X_n}$  in (5.70) is a density of  $(X_1, \dots, X_n) \Rightarrow X_1, \dots, X_n$  are independent. If Equation (5.70) holds and  $B = \prod_{i=1}^n ]-\infty, x_i]$ , then for all  $(x_1, \dots, x_n) \in \mathbb{R}^n$

$$\begin{aligned}
& F_{X_1, \dots, X_n}(x_1, \dots, x_n) \\
&= \int_B f_{X_1, \dots, X_n}(t_1, \dots, t_n) \lambda_n[d(t_1, \dots, t_n)] && [(5.67)] \\
&= \int_{]-\infty, x_n]} \dots \int_{]-\infty, x_1]} f_{X_1}(t_1) \cdot \dots \cdot f_{X_n}(t_n) \lambda(dt_1) \dots \lambda(dt_n) && [\text{Th. 3.76}] \\
&= \prod_{i=1}^n \int_{]-\infty, x_i]} f_{X_i}(t_i) \lambda(dt_i) \\
&= \prod_{i=1}^n F_{X_i}(x_i). && [(5.62)]
\end{aligned}$$

Now, Corollary 5.93 implies that  $X_1, \dots, X_n$  are independent.

### 5.9 Exercises

▷ **Exercise 5-1** Consider the random variable  $X$  defined in Example 2.34. Which are the elements  $\omega$  in the inverse image  $X^{-1}(\{1\})$  and which are the probabilities of the events  $\{\omega\}$ ?

▷ **Exercise 5-2** Consider again the random variable  $X$  defined in Example 2.34. What are the values of the distribution of  $X$  and the distribution of  $U$ ?

▷ **Exercise 5-3** Consider Example 5.9. Show that  $1_{X \in A'}$  is a random variable on  $(\Omega, \mathcal{A}, P)$  and that  $1_{X \in A'} = 1_{A'}(X) = 1_{A'} \circ X$ .

▷ **Exercise 5-4** Consider Example 5.15 and show that  $X_1$  is prior to  $X_1 \cdot X_2$  in  $(\mathcal{F}_1, \mathcal{F}_2)$ .

▷ **Exercise 5-5** Show that  $X \stackrel{p_B}{=} Y$  for  $B = \{\omega_1, \omega_2, \omega_3\}$  and that  $X \stackrel{p}{=} Y$  does *not* hold in Example 5.23.

▷ **Exercise 5-6** Consider the random variable  $(U, X): \Omega \rightarrow \Omega_U \times \{0, 1\}$  defined in Example 2.34 (p. 53). Which are the elements  $\omega$  in the inverse image  $(U, X)^{-1}(\{(Joe, 1)\})$  and which are the probabilities to the events  $\{\omega\}$ ?

▷ **Exercise 5-7** Show that the random variables  $U$  and  $X$  presented in Table 2.2 (p. 53) are independent.

▷ **Exercise 5-8** Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $B \in \mathcal{B}$ , and  $\mathcal{C} \subset \mathcal{A}$  a  $\sigma$ -algebra. Prove that the following propositions are equivalent to each other:  $1_B \perp\!\!\!\perp_P \mathcal{C}$ ,  $\sigma(\{B\}) \perp\!\!\!\perp_P \mathcal{C}$ ,  $\{B\} \perp\!\!\!\perp_P \mathcal{C}$ ,  $B \perp\!\!\!\perp_P \mathcal{C}$ .

▷ **Exercise 5-9** In Example 5.50 we showed that  $X$  and  $U$  are independent and in Example 5.54 we defined the mappings  $f: \{0, 1, 2\} \rightarrow \{0, 1\}$  and  $g: \Omega_U \rightarrow \{male, female\}$ . Use Equation (5.24) to show that the mappings  $f(X)$  (control vs. any of the two treatments) and  $g(U)$  (sex) are independent as well.

▷ **Exercise 5-10** In Example 5.85 we specified the values of the distribution function  $F_X$  for the random variable *number of flipping heads*. Use these values to compute all eight values of the distribution  $P_X$ .

### Solutions

▷ **Solution 5-1** The inverse image of the set  $\{1\}$  under  $X$  is

$$X^{-1}(\{1\}) = \{(Joe, yes, -), (Joe, yes, +), (Ann, yes, -), (Ann, yes, +)\}.$$

The events  $\{\omega\}$ ,  $\omega \in X^{-1}(\{1\})$  have the probabilities  $P[\{(Joe, yes, -)\}] = .04$ ,  $P[\{(Joe, yes, +)\}] = .16$ ,  $P[\{(Ann, yes, -)\}] = .12$ , and  $P[\{(Ann, yes, +)\}] = .08$ .

▷ **Solution 5-2** The random variable  $X$  (the treatment variable) has the following distribution:

$$\begin{aligned} P_X(\{1\}) &= P[X^{-1}(\{1\})] = .40, & P_X(\{0\}) &= P[X^{-1}(\{0\})] = .60, \\ P_X(\Omega') &= P[X^{-1}(\Omega')] = 1, & P_X(\emptyset) &= P[X^{-1}(\emptyset)] = 0, \end{aligned}$$

whereas the distribution of the random variable  $U$  (the observational-unit variable), is:

$$P_U(\{Joe\}) = P[U^{-1}(\{Joe\})] = .50 \quad P_U(\{Ann\}) = P[U^{-1}(\{Ann\})] = .50$$

$$P_U(\Omega_U) = P[U^{-1}(\Omega_U)] = 1 \quad P_U(\emptyset) = P[U^{-1}(\emptyset)] = 0.$$

▷ **Solution 5-3** Let  $A' \in \mathcal{A}'_X$  and consider the indicator function  $1_{X \in A'}: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ .  
Measurability: For all  $B \in \mathcal{B}$ ,

$$\begin{aligned} (1_{X \in A'})^{-1}(B) &= (1_{X^{-1}(A')})^{-1}(B) && [(5.6)] \\ &= \{\omega \in \Omega: 1_{X^{-1}(A')}(\omega) \in B\} && [(2.2)] \\ &= \begin{cases} \emptyset, & \text{if } 0 \notin B, 1 \notin B, \\ X^{-1}(A'), & \text{if } 0 \notin B, 1 \in B \\ \Omega \setminus X^{-1}(A'), & \text{if } 0 \in B, 1 \notin B \\ \Omega, & \text{if } \{0, 1\} \subset B, \end{cases} \end{aligned}$$

and all these sets are elements of  $\mathcal{A}$ , because  $X$  is assumed to be a random variable.

Furthermore, for all  $\omega \in \Omega$ ,

$$\begin{aligned} 1_{X \in A'}(\omega) &= 1_{X^{-1}(A')}(\omega) && [(5.6)] \\ &= \begin{cases} 1, & \text{if } \omega \in X^{-1}(A') \\ 0, & \text{if } \omega \notin X^{-1}(A') \end{cases} \\ &= \begin{cases} 1, & \text{if } X(\omega) \in A' \\ 0, & \text{if } X(\omega) \notin A' \end{cases} && [(2.2)] \\ &= 1_{A'}[X(\omega)] \\ &= 1_{A'} \circ X(\omega). && [(2.25)] \end{aligned}$$

▷ **Solution 5-4**

The filtration  $(\mathcal{F}_1, \mathcal{F}_2)$  is defined by  $\mathcal{F}_1 := \sigma(X_1)$  and  $\mathcal{F}_2 := \sigma(X_1, X_2)$ . Hence,

$$\sigma(X_1) \subset \mathcal{F}_1 = \{\Omega, \emptyset, \{(h, h), (h, t)\}, \{(t, h), (t, t)\}\}$$

In contrast,

$$\sigma(X_1 \cdot X_2) = \{\Omega, \emptyset, \{(h, h)\}, \{(h, t), (t, h), (t, t)\}\} \not\subset \mathcal{F}_1.$$

However,  $\sigma(X_1 \cdot X_2) \subset \mathcal{F}_2$ , because  $\mathcal{F}_2 = \mathcal{P}(\Omega)$ . Hence, according to Definition 5.13 (i),  $X_1$  is prior to  $X_1 \cdot X_2$  in the filtration  $(\mathcal{F}_1, \mathcal{F}_2)$ .

▷ **Solution 5-5** In order to prove  $X \stackrel{P^B}{=} Y$  for  $B = \{\omega_1, \omega_2, \omega_3\}$ , it is sufficient to show  $P^B(A) = 0$ , where  $A := \{\omega \in \Omega: X(\omega) \neq Y(\omega)\} \in \mathcal{A} = \mathcal{P}(\Omega)$ . In this example,  $A = \{\omega_3, \omega_4\}$ . Now  $P(\{\omega_3\}) = 0$ , which implies

$$P^B(\{\omega_3\}) = \frac{P(B \cap \{\omega_3\})}{P(B)} = \frac{P(\{\omega_3\})}{P(B)} = \frac{0}{P(B)} = 0.$$

Furthermore,

$$P^B(\{\omega_4\}) = \frac{P(B \cap \{\omega_4\})}{P(B)} = \frac{P(\emptyset)}{P(B)} = \frac{0}{P(B)} = 0.$$

Additivity of the measure  $P^B$  then implies  $P^B(A) = 0$ .

In order to prove that  $X \stackrel{P}{=} Y$  does not hold, it suffices to show  $P(A) \neq 0$ . Now  $P(\{\omega_3\}) = 0$  and  $P(\{\omega_4\}) = .50$ . Therefore,  $P(A) = P(\{\omega_4\}) = .50 > 0$ .

▷ **Solution 5-6** The inverse image of the set  $\{1\}$  under  $X$  is

$$X^{-1}(\{1\}) = \{(Joe, yes, -), (Joe, yes, +), (Ann, yes, -), (Ann, yes, +)\}.$$

The events  $\{\omega\}$ ,  $\omega \in X^{-1}(\{1\})$  have the probabilities  $P[\{(Joe, yes, -)\}] = .04$ ,  $P[\{(Joe, yes, +)\}] = .16$ ,  $P[\{(Ann, yes, -)\}] = .12$ , and  $P[\{(Ann, yes, +)\}] = .08$ .

▷ **Solution 5-7** In Example 4.43, we already showed that the events  $A := \{Joe\} \times \Omega_X \times \Omega_Y$  (that Joe is sampled) and  $B := \Omega_U \times \{yes\} \times \Omega_Y$  (that the person sampled is treated) are independent. According to Box 4.2 (iii), this implies that the  $\sigma$ -algebras  $\{A, A^c, \Omega, \emptyset\}$  and  $\{B, B^c, \Omega, \emptyset\}$  are independent as well. If  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  and  $U: (\Omega, \mathcal{A}, P) \rightarrow (\Omega_U, \mathcal{A}_U)$  are the random variables defined Table 2.2 (p. 53), then  $X^{-1}(\mathcal{A}') = \{A, A^c, \Omega, \emptyset\}$  and  $U^{-1}(\mathcal{A}_U) = \{B, B^c, \Omega, \emptyset\}$  are the  $\sigma$ -algebras generated by  $U$  and  $X$ , respectively. Hence, in this example,  $X$  and  $U$  are independent.

▷ **Solution 5-8** First of all,

$$\{B\} \perp\!\!\!\perp_P \mathcal{C} \Leftrightarrow B \perp\!\!\!\perp_P \mathcal{C} \quad [\text{Rem. 4.41}].$$

Furthermore,  $\sigma(1_B) = \{\Omega, \emptyset, B, B^c\}$  (see Example 2.31) and  $\sigma(\{B\}) = \{\Omega, \emptyset, B, B^c\}$ . Hence,  $\sigma(1_B) = \sigma(\{B\})$ , and this implies

$$1_B \perp\!\!\!\perp_P \mathcal{C} \Leftrightarrow \sigma(\{B\}) \perp\!\!\!\perp_P \mathcal{C} \quad [\text{Rem. 5.44}].$$

Finally,

$$\sigma(\{B\}) \perp\!\!\!\perp_P \mathcal{C} \Leftrightarrow \{B\} \perp\!\!\!\perp_P \mathcal{C} \quad [\text{Box 4.2 (ii)}].$$

▷ **Solution 5-9** Consider the two events  $A := f(X)^{-1}(\{0\})$  and  $B := g(U)^{-1}(\{male\})$ .

$$P(A \cap B) = \frac{40}{99}$$

and

$$P(A) \cdot P(B) = \frac{60}{99} \cdot \frac{2}{3} = \frac{40}{99}.$$

Because  $f(X)^{-1}(\mathcal{P}(\{0, 1\})) = \{A, A^c, \Omega, \emptyset\}$  and  $g(U)^{-1}(\mathcal{P}(\{male, female\})) = \{B, B^c, \Omega, \emptyset\}$ , this proves that  $f(X)$  and  $g(U)$  are independent (see Exercise 5-8).

▷ **Solution 5-10** We consider all elements  $B \in \mathcal{B}$  of the Borel  $\sigma$ -algebra on  $\mathbb{R}$  and assign the following values:

$$P_X(B) = \begin{cases} F_X(-1) = 0, & \text{if } 0, 1, 2 \notin B \\ F_X(0) = \frac{1}{4}, & \text{if } 0 \in B, 1, 2 \notin B \\ F_X(1) = \frac{3}{4}, & \text{if } 0, 1 \in B, 2 \notin B \\ F_X(2) = 1, & \text{if } 0, 1, 2 \in B \\ F_X(1) - F_X(0) = \frac{2}{4}, & \text{if } 1 \in B, 0, 2 \notin B \\ F_X(2) - F_X(0) = \frac{3}{4}, & \text{if } 1, 2 \in B, 0 \notin B \\ F_X(2) - F_X(1) = \frac{1}{4}, & \text{if } 2 \in B, 0, 1 \notin B \\ F_X(2) - F_X(1) + F_X(0) = \frac{1}{2}, & \text{if } 0, 2 \in B, 1 \notin B. \end{cases}$$

## Chapter 6

# Expectation, Variance, and Other Moments

In chapter 4 we introduced a probability measure as a special finite measure and in chapter 5 we defined a random variable as a measurable mapping on a probability space. In this chapter we will translate integration theory (see ch. 3) to probability theory introducing *expectations* of numerical random variables and other important concepts that are special expectations: *central* and *noncentral moments*, and *variances*. Even *covariances* and *correlations* are special expectations (see ch. 7). All these quantities describe important properties of random variables, although, in general, they do not determine the complete distribution.

### 6.1 Expectation

#### 6.1.1 Definition

Reading the following definition remember that a random variable  $Y$  is called *quasi-integrable* with respect to  $P$  if  $\int Y^+ dP$  or  $\int Y^- dP$  are finite, where  $Y^+$  and  $Y^-$  denote the positive and negative parts of  $Y$ , respectively (see Rem. 2.62 and Def. 3.28).

**Definition 6.1 (Expectation)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a numerical random variable that is quasi-integrable with respect to  $P$ . Then we define

$$E(Y) := \int Y dP, \tag{6.1}$$

call it the *expectation* of  $Y$  (w.r.t.  $P$ ), and say that it *exists*. Instead of *expectation* w.r.t.  $P$  we also use the term  *$P$ -expectation*.

**Remark 6.2 (Existence of the Expectation)** Note that  $E(Y)$  can be infinite. Furthermore, if  $E(Y)$  exists 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*.  $\triangleleft$

**Remark 6.3 (Notation and Synonymous Terms)** A synonym for expectation is *expectation value*. The reference to the measure  $P$  is usually omitted if the context is unambiguous. If we consider the expectation with respect to another probability measure on  $(\Omega, \mathcal{A})$ , e. g., the conditional-probability measure  $P^B$  (see Def. 4.29), then we adapt the notation correspondingly:

$$E^B(Y) := \int Y dP^B. \quad (6.2)$$

*Expectation with respect to  $P^B$*  is used synonymously to  *$P^B$ -expectation*.  $\triangleleft$

**Remark 6.4 (Random Variables With Finite Expectations)** Without substantial loss of generality, we will often assume that a random variable  $Y$  is *real-valued* if its expectation is finite. According to Remark 3.42, if the random variable  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  has a finite expectation, then there is random variable  $Y^*: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  such that  $Y \stackrel{P}{=} Y^*$ .  $\triangleleft$

### 6.1.2 Expectation of a Discrete Random Variable

In this section we use the notation  $P(Y=y_i) := P[Y^{-1}(\{y_i\})]$  introduced in Remark 5.4.

**Remark 6.5 (Random Variable With a Finite Number of Real Values)** Suppose  $y_1, \dots, y_n \in \mathbb{R}$ ,  $n \in \mathbb{N}$ , denote all (negative, 0, or positive) values of a real-valued random variable  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ , let  $P_Y$  denote the distribution, and  $p_Y$  the probability function of  $Y$ . Then the expectation  $E(Y)$  exists, and

$$E(Y) = \sum_{i=1}^n y_i \cdot P(Y=y_i) = \sum_{i=1}^n y_i \cdot P_Y(\{y_i\}) = \sum_{i=1}^n y_i \cdot p_Y(y_i) \quad (6.3)$$

[see Cor. 3.59, (3.45), Eq. (5.34), and Def. 5.56]. Hence, if  $Y$  has only a finite number  $n$  of values, then its expectation is simply the sum of its values, each one weighted by its probability  $P(Y=y_i) = P_Y(\{y_i\}) = p_Y(y_i)$ .  $\triangleleft$

**Example 6.6 (Expectation of an Indicator)** If  $(\Omega, \mathcal{A}, P)$  is a probability space and  $1_A$  is the indicator of  $A \in \mathcal{A}$ , then

$$E(1_A) = 0 \cdot P(1_A=0) + 1 \cdot P(1_A=1) = P(1_A=1) = P(A) \quad (6.4)$$

is the expectation of  $1_A$  (see also Example 3.14). Considering the event  $\{Y=y\}$  and using the notation  $1_{Y=y} := 1_{\{Y=y\}}$ , this yields

$$E(1_{Y=y}) = P(Y=y). \quad (6.5)$$

$\triangleleft$

**Example 6.7 (Joe and Ann With Random Assignment – continued)** In Example 1.9 we defined the set

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

the event that the drawn person is treated, and the set

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

the event that  $\{+\}$  (success) occurs, irrespective of which person is drawn and whether or not the person is treated. In Table 2.2 (p. 53) we assigned probabilities to each elementary event  $\{\omega\}$ ,  $\omega \in \Omega$  and defined  $X := 1_B$ , the treatment variable, as well as  $Y := 1_C$  the outcome variable. Applying Equation (6.4) to the indicator  $1_B$  yields:

$$\begin{aligned} E(X) &= E(1_B) = P(B) \\ &= P[\{(Joe, yes, -)\}] + P[\{(Joe, yes, +)\}] + P[\{(Ann, yes, -)\}] + P[\{(Ann, yes, +)\}] \\ &= .04 + .16 + .12 + .08 = .40. \end{aligned}$$

Similarly, for the indicator  $1_C$ , we obtain

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

&lt;

**Example 6.8 (Tossing a Dice – continued)** In Example 3.6 we considered the random variable  $X = \text{number of dots}$ . In this example we specified the probability space  $(\Omega, \mathcal{A}, P)$  by  $\Omega := \{\omega_1, \dots, \omega_6\}$ ,  $\mathcal{A} := \mathcal{P}(\Omega)$ , and  $P(\{\omega_1\}) = \dots = P(\{\omega_6\}) = 1/6$ . Because  $P(X=x_i) = P_X(\{i\}) = P(\{\omega_i\})$ ,  $i = 1, \dots, 6$ , Equation (6.3) yields

$$E(X) = \sum_{i=1}^6 i \cdot P(X=i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5.$$

&lt;

**Remark 6.9 (Random Variable With a Countable Number of Real Values)** Let  $y_1, y_2, \dots \in \mathbb{R}$  denote the values of a real-valued random variable  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  and suppose that the expectation of  $Y$  exists. Then

$$E(Y) = \sum_{i=1}^{\infty} y_i \cdot P(Y=y_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n y_i \cdot P(Y=y_i) \quad (6.6)$$

[see (3.45), Eq. (5.34), and Th. 3.60]. Examples in which the expectation of a random variable is the ‘infinite sum’ of its values weighted by their probabilities are random variables with a Poisson distribution and with a geometric distribution (see Theorems 8.16 and 8.22). <

**Example 6.10 (A Discrete Random Variable Without Expectation)** Suppose that  $\Omega = \mathbb{N}_0 = \{0, 1, 2, \dots\}$  and consider the random variable  $Y: (\Omega, \mathcal{P}(\Omega), P) \rightarrow (\mathbb{R}, \mathcal{B})$  defined by

$$Y(i) = y_i = (-1)^i i! e, \quad \forall i \in \mathbb{N}_0, \quad (6.7)$$

with

$$P(Y=y_i) = \frac{1}{e} \cdot \frac{1}{i!}, \quad \forall i \in \mathbb{N}_0. \quad (6.8)$$

Note that  $e = \sum_{i=0}^{\infty} \frac{1}{i!}$ . Hence, dividing both sides by  $e$  yields  $\sum_{i=0}^{\infty} \frac{1}{i!e} = 1$ . Therefore, Equation (6.8) specifies a probability distribution. Now consider

$$\sum_{i=0}^n y_i \cdot P(Y=y_i) = \sum_{i=0}^n (-1)^i = \begin{cases} 1, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd.} \end{cases} \quad (6.9)$$

Obviously, in this example, the limit

$$\lim_{n \rightarrow \infty} \sum_{i=0}^n y_i \cdot P(Y=y_i)$$

occurring in Equation (6.6) does not exist. Therefore, according to Definitions 6.1 and 3.28  $E(Y)$  is not defined.  $\triangleleft$

### 6.1.3 Computing the Expectation Using a Density

According to the following theorem, the expectation of a continuous real-valued random variable can also be computed using its density  $f_Y$  with respect to the Lebesgue measure (see Def. 5.94) and the *Riemann integral*.

#### **Theorem 6.11 (Expectation of a Continuous Random Variable)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a continuous random variable with expectation  $E(Y)$  and a density  $f_Y$  that is Riemann integrable. Then

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy. \quad (6.10)$$

(Proof p. 212)

Examples of continuous random variables and their expectations are treated in chapter 8, section 8.2.

**Example 6.12 (A Continuous Random Variable Without Expectation)** Consider the continuous random variable  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  with density

$$f_Y(y) = \frac{1}{\pi} \cdot \frac{1}{1+y^2}, \quad \forall y \in \mathbb{R}, \quad (6.11)$$

and distribution function

$$F_Y(y) = \frac{1}{2} + \frac{1}{\pi} \cdot \arctan y, \quad \forall y \in \mathbb{R}. \quad (6.12)$$

A distribution  $P_Y$  with density (6.11) is called a *standard Cauchy distribution* or *central  $t$ -distribution* with 1 degree of freedom.

The integral of the positive part of  $y \cdot f_Y(y)$  is

$$\int_{-\infty}^{\infty} [y \cdot f_Y(y)]^+ dy = \int_0^{\infty} y \cdot \frac{1}{\pi} \cdot \frac{1}{1+y^2} dy = \frac{1}{2\pi} \ln(1+y^2) \Big|_0^{-\infty} = \infty,$$

and the integral of the negative part is

$$\int_{-\infty}^{\infty} [y \cdot f_Y(y)]^- dy = \int_{-\infty}^0 -y \cdot \frac{1}{\pi} \cdot \frac{1}{1+y^2} dy = -\frac{1}{2\pi} \ln(1+y^2) \Big|_{-\infty}^0 = \infty$$

[for the notation cf. Eq. (3.69)]. Hence,  $y \cdot f_Y(y)$  is not quasi- $P$ -integrable on  $\mathbb{R}$  (see Def. 3.28) and  $E(Y)$  does not exist.  $\triangleleft$

#### 6.1.4 Transformation Theorem

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_Y, \mathcal{A}'_Y)$  and  $g: (\Omega'_Y, \mathcal{A}'_Y, P_Y) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  be random variables. Then we denote the expectation of  $g$  with respect to the distribution  $P_Y$  by  $E_Y(g)$ . The following corollary immediately follows from Theorem 3.57, translating the measure theory terms to probability theory. This corollary is relevant whenever we consider the expectation of a composition  $g \circ Y = g(Y)$  of a numerical function  $g$  [see Eq. (2.25)] and a mapping  $Y$  or the expectation of  $g$  with respect to the distribution  $P_Y$  [see Eq. (5.2)].

##### Theorem 6.13 (Transformation Theorem)

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_Y, \mathcal{A}'_Y)$  be a random variable and  $g: (\Omega'_Y, \mathcal{A}'_Y) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  be measurable.

(i) If  $g$  is nonnegative or has a finite expectation  $E_Y(g)$ , then

$$E_Y(g) = \int g dP_Y = \int g(y) P_Y(dy) = \int g(Y) dP = E[g(Y)]. \quad (6.13)$$

(ii)  $E_Y(g)$  is finite if and only if  $E[g(Y)]$  is finite.

The virtue of Equation (6.13) is that we do not have to know the distribution of  $g(Y)$ . Instead, the distribution of  $Y$  suffices.

**Remark 6.14 (A Special Case)** If we consider the special case in which  $g$  is the identity function  $id: \bar{\mathbb{R}} \rightarrow \bar{\mathbb{R}}$ , defined by  $id(y) = y$ , for all  $y \in \bar{\mathbb{R}}$ , then  $id(Y) = Y$  and Equations (6.13) yield

$$E(Y) = \int Y dP = \int y P_Y(dy). \quad (6.14)$$

$\triangleleft$

**Remark 6.15 (Finite Number of Values)** If  $Y$  takes on only a finite number of different values  $y_1, \dots, y_n \in \mathbb{R}$ , then Equation (6.13) simplifies to

$$E[g(Y)] = E_Y(g) = \int g \, dP_Y = \sum_{i=1}^n g(y_i) \cdot P(Y=y_i), \quad (6.15)$$

where  $P(Y=y_i) = P_Y(\{y_i\}) = p_Y(y_i)$ ,  $i = 1, \dots, n$ , and  $p_Y$  denotes the probability function of  $Y$  (see Def. 5.56).  $\triangleleft$

**Remark 6.16 (Countable Number of Values)** If  $Y$  takes on a countable number of different values  $y_1, y_2, \dots \in \mathbb{R}$  and  $\sum_{i=1}^{\infty} g^+(y_i) \cdot P(Y=y_i) < \infty$  or  $\sum_{i=1}^{\infty} g^-(y_i) \cdot P(Y=y_i) < \infty$ , then

$$E[g(Y)] = E_Y(g) = \int g \, dP_Y = \sum_{i=1}^{\infty} g(y_i) \cdot P(Y=y_i). \quad (6.16)$$

Note that (6.16) also applies if  $g$  is nonnegative, because in this case  $g^- = 0$  holds for the negative part of  $g$ , which implies  $\sum_{i=1}^{\infty} g^-(y_i) = 0 < \infty$ .  $\triangleleft$

Equation (6.13) immediately implies the following corollary according to which the expectations of two random variables  $X$  and  $Y$  are identical if they have identical distributions, provided that the expectations exist (see also Remark 6.27).

**Corollary 6.17 (Identical Distributions Imply Identical Expectations)**

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$  be random variables and  $g: (\Omega', \mathcal{A}') \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  a measurable function that is nonnegative or with expectation  $E_X(g)$ . If  $P_X = P_Y$ , then  $E[g(X)] = E[g(Y)]$ .

This property allows us to use the term *expectation of a distribution* instead of expectation of a random variable.

In the following lemma we consider a bivariate random variable  $(X, Y)$  and a numerical function  $g(X)$ , i. e., a function that only depends on  $X$ . According to this lemma, the expectation of  $g$  with respect to the joint distribution  $P_{X,Y}$  is identical to the expectation of  $g$  with respect to the marginal distribution  $P_X$ .

**Lemma 6.18 (Expectation w.r.t. Joint and Marginal Distributions)**

Let  $(X, Y): (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X \times \Omega'_Y, \mathcal{A}'_X \otimes \mathcal{A}'_Y)$  be a bivariate random variable with joint distribution  $P_{X,Y}$  and let  $g: (\Omega'_X, \mathcal{A}'_X) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  be a measurable function that is nonnegative or with expectation  $E_X(g)$ . Then

$$E_{X,Y}(g) = E_X(g), \quad (6.17)$$

which is equivalent to

$$\int g(x) P_{X,Y} [d(x, y)] = \int g(x) P_X (dx). \quad (6.18)$$

(Proof p. 212)

**Example 6.19 (Flipping two Coins – continued)** Consider again the random variable  $X = \text{number of flipping heads}$  and the indicator  $1_H: \Omega \rightarrow \mathbb{R}$  of the event that *at least one heads is tossed*. In Example 2.47 we showed that  $1_H = g \circ X$ , where  $g: \mathbb{R} \rightarrow \mathbb{R}$  is defined by

$$g(x) = \begin{cases} 1 & \text{if } x \in \{1, 2\}, \\ 0, & \text{otherwise} \end{cases} \quad \text{for all } x \in \mathbb{R}.$$

According to (6.15),

$$E(1_H) = E[g(X)] = E_X(g) = 0 \cdot P(\{X=0\}) + 1 \cdot P(X \in \{1, 2\}) = P(X \in \{1, 2\}) = \frac{3}{4}.$$

&lt;

**Example 6.20 (Expectation of  $Y^2$ )** Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a real-valued random variable, let  $g: (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$  be measurable and let  $g(Y) := Y^2$ . Then, according to Equation (6.13),

$$E(Y^2) = E[g(Y)] = E_Y(g) = \int g dP_Y = \int y^2 P_Y(dy).$$

[Note that  $Y^2$  is nonnegative and  $E(Y^2)$  can be infinite.] This equation shows that the expectation of  $Y^2$  solely depends on the distribution  $P_Y$  of  $Y$ , which illustrates Corollary 6.17. Using the integral  $\int g dP_Y$  is often the most convenient way of computing the expectation  $E(Y^2)$ . If  $Y$  takes on only a finite number of values  $y_1, \dots, y_n \in \mathbb{R}$ , then this equation simplifies to

$$E(Y^2) = \sum_{i=1}^n y_i^2 \cdot P_Y(\{y_i\}) = \sum_{i=1}^n y_i^2 \cdot P(Y=y_i). \quad (6.19)$$

These equations only involve the probabilities  $P(Y=y_i) = P_Y(\{y_i\})$ , not the probabilities  $P(Y^2=y^2)$ . <

**Example 6.21 (Multiplication With Indicators)** Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a numerical random variable with expectation  $E(Y)$ . If  $A \in \mathcal{A}$  and  $P(A) = 0$ , then  $(1_A Y) \stackrel{p}{=} 0$  and Rule (i) of Box 6.1 implies

$$E(1_A Y) = 0. \quad (6.20)$$

If  $C = A \cup B$ ,  $A \cap B = \emptyset$ , and  $A, B \in \mathcal{A}$ , then  $1_C Y = 1_A Y + 1_B Y$  and Rule (vi) of Box 6.1 implies

$$E(1_C Y) = E(1_A Y) + E(1_B Y). \quad (6.21)$$

&lt;

The following corollary shows how to compute the expectation of the composition  $g(Y)$  using the density of  $Y$ . As mentioned before, the virtue of Equation (6.22) is that we do not have to know the density of  $g(Y)$ , the density of  $Y$  suffices.

**Corollary 6.22 (Transformation Theorem, Continuous Random Variable)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a continuous random variable with a Riemann integrable density  $f_Y$ . If  $g: (\mathbb{R}, \mathcal{B}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is a measurable function that is nonnegative or numerical with finite expectation  $E_Y(g) = \int g dP_Y$ , then

$$E[g(Y)] = E_Y(g) = \int_{-\infty}^{\infty} g(y) f_Y(y) dy. \quad (6.22)$$

(Proof p. 212)

**6.1.5 Rules of Computation**

Some rules of computation for expectations are gathered in Box 6.1 (for proofs see Exercise 6-1).

**Example 6.23 (Expectation of a Sample Mean)** Let  $Y_1, \dots, Y_n$  be a random sample. This means that the random variables  $Y_1, \dots, Y_n$  are i.i.d. (see Rem. 5.42). Furthermore, if the  $Y_1, \dots, Y_n$  are real-valued, define

$$\bar{Y} = \frac{1}{n} \cdot \sum_{i=1}^n Y_i, \quad (6.23)$$

the *arithmetic mean*, which in statistics is also called the *sample mean*. If  $Y_1$  is nonnegative or with finite expectation and

$$\mu_Y := E(Y_1) \quad (6.24)$$

denotes the identical expectations of the variables  $Y_1, \dots, Y_n$ , then

$$E(\bar{Y}) = \mu_Y \quad (6.25)$$

(see Exercise 6-4). ◁

Now we turn to a generalization of Rule (x) of Box 6.1.

**Theorem 6.24 (Expectation of the Product of Random Variables)**

Let  $Y_i: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ ,  $i = 1, \dots, n$ , be real-valued random variables that are nonnegative or with finite expectations and assume that the  $Y_1, \dots, Y_n$  are independent. Then

$$E\left(\prod_{i=1}^n Y_i\right) = \prod_{i=1}^n E(Y_i). \quad (6.26)$$

If the expectations  $E(Y_i)$ ,  $i = 1, \dots, n$ , are finite, then  $E\left(\prod_{i=1}^n Y_i\right)$  is finite, too.

For a proof see, e. g., Bauer (1996, Theorem 8.1, p. 49). Later we will weaken the independence assumption [see Rem. 7.10 and Box 7.1 (i)]. However, if the variables  $Y_i$  are *not independent*, then Equation (6.26) does *not necessarily hold* (see Remark 7.10).

**Box 6.1 Rules of Computation for Expectations**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a random variable with expectation  $E(Y)$  and let  $\alpha \in \mathbb{R}$ . Then:

$$Y \stackrel{P}{=} \alpha \Rightarrow E(Y) = \alpha. \quad (\text{i})$$

$$E(\alpha + Y) = \alpha + E(Y). \quad (\text{ii})$$

$$E(\alpha \cdot Y) = \alpha \cdot E(Y). \quad (\text{iii})$$

Let  $A, B \in \mathcal{A}$ . Then

$$E(\mathbf{1}_A \cdot \mathbf{1}_B) = P(A \cap B). \quad (\text{iv})$$

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

If  $Y_1, Y_2$  are nonnegative or real-valued with finite expectation, then

$$E(Y_1 + Y_2) = E(Y_1) + E(Y_2). \quad (\text{vi})$$

For  $i = 1, \dots, n$ , let  $Y_i: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be random variables with finite expectations  $E(Y_i)$  and  $\alpha_i \in \mathbb{R}$ . Then

$$E\left(\sum_{i=1}^n \alpha_i \cdot Y_i\right) = \sum_{i=1}^n \alpha_i \cdot E(Y_i). \quad (\text{vii})$$

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be random variables that are nonnegative or with finite expectations. Then:

$$X \stackrel{P}{=} Y \Rightarrow E(X) = E(Y). \quad (\text{viii})$$

$$X \stackrel{P}{=} Y \Leftrightarrow \forall A \in \mathcal{A}: E(\mathbf{1}_A X) = E(\mathbf{1}_A Y). \quad (\text{ix})$$

$$X \perp\!\!\!\perp_P Y \Rightarrow E(X \cdot Y) = E(X) \cdot E(Y). \quad (\text{x})$$

**6.2 Moments, Variance, and Standard Deviation**

The expectation  $E(Y)$  of a numerical random variable  $Y$  is also called the *first moment of  $Y$* , provided that this expectation exists, whereas the expectation  $E(Y^2)$  is called the *second moment of  $Y$*  (see Example 6.20). For second (and higher) moments we distinguish between *moments* and *central moments*.

**Remark 6.25 (Higher Moments of  $Y$ )** Analogously to Example 6.20, we may consider  $g(Y) = |Y^n|$  or  $g(Y) = Y^n$  with  $n \geq 1$ . Note that

- (i)  $E(|Y^n|) = E(|Y|^n)$  always exists, because  $|Y^n|$  is nonnegative (see Def. 6.1).
- (ii) If, for  $n \in \mathbb{N}$ , the expectation  $E(|Y|^n)$  is finite, then  $E(Y^n)$  exists and is finite (see Cor. 3.38).
- (iii) If, for  $n \in \mathbb{N}$ , the expectation  $E(|Y|^n)$  is finite, then  $E(Y^m)$  exists and is finite as well for all  $m$  with  $1 \leq m \leq n$  (see Exercise 6-2).
- (iv) If  $n$  is even, then the random variable  $Y^n$  is nonnegative and  $E(Y^n)$  exists. In contrast, if  $n$  is odd, then the expectation  $E(Y^n)$  does not necessarily exist.

◁

**Definition 6.26 (Moments)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a numerical random variable and let  $n \in \mathbb{N}$ .

- (i) The expectation  $E(|Y|^n)$  is called the *nth absolute moment* of  $Y$ .
- (ii) If  $Y$  is nonnegative or such that  $E(Y^n)$  is finite, then  $E(Y^n)$  is called the *nth moment* of  $Y$ .
- (iii) If  $E(Y^n)$  is finite or if  $Y$  is nonnegative and the expectation  $E(Y)$  is finite, then we call  $E([Y - E(Y)]^n)$  the *nth central moment* of  $Y$ .

**Remark 6.27 (Moments Under  $P$ -Equivalence)** If the expectations of  $Y^n$  and therefore also of  $[Y - E(Y)]^n$  are finite, then they can be represented as expectations of functions  $g(Y)$  of  $Y$ , where  $g: (\overline{\mathbb{R}}, \overline{\mathcal{B}}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is a measurable function with finite expectation  $E_Y(g)$  [see Proposition (ii) of Th. 6.13]. Therefore, according to Corollary 6.17, all moments (central or noncentral) of a numerical random variable  $Y$  solely depend on its distribution  $P_Y$ . Hence, if two random variables  $Y_1$  and  $Y_2$  have the same distribution  $P_{Y_1} = P_{Y_2}$ , then they have the same moments. For instance, if  $Y_1 \stackrel{P}{=} Y_2$  and the expectations  $E(Y_1^2)$  and  $E(Y_2^2)$  are finite, then  $E(Y_1) = E(Y_2)$  and  $E(Y_1^2) = E(Y_2^2)$ . This allows us to use the terms (central) moments of a distribution instead of (central) moments of a random variable. ◁

Variance and standard deviation are the most important parameters describing the *variability* of a random variable. They are defined as follows:

**Definition 6.28 (Variance and Standard Deviation)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a numerical random variable and assume that  $E(Y^2) < \infty$ . Then the *variance* of  $Y$  is defined by

$$\text{Var}(Y) := E([Y - E(Y)]^2), \quad (6.27)$$

and the *standard deviation* of  $Y$  by the positive square root of the variance, i. e.,

$$\text{SD}(Y) := \sqrt{\text{Var}(Y)}. \quad (6.28)$$

**Box 6.2 Rules of Computation for Variances**

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be random variables with finite second moments and let  $\alpha \in \mathbb{R}$ . Then:

$$\text{Var}(Y) = E(Y^2) - E(Y)^2. \quad (\text{i})$$

$$\text{Var}(\alpha + Y) = \text{Var}(Y). \quad (\text{ii})$$

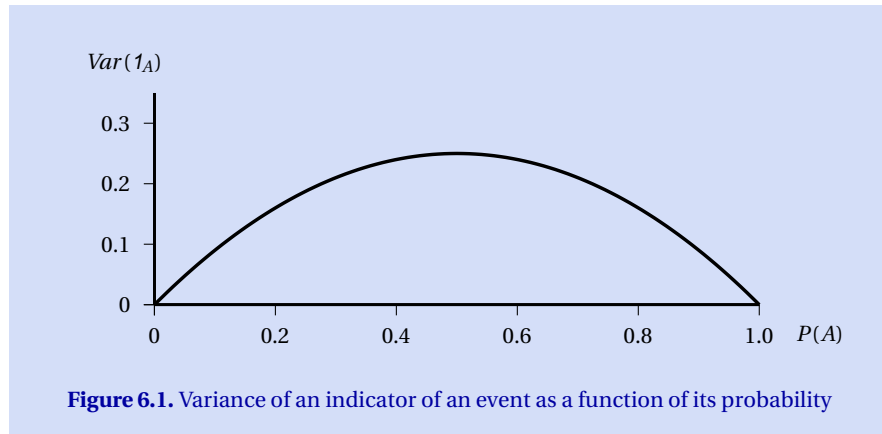
$$\text{Var}(\alpha \cdot Y) = \alpha^2 \cdot \text{Var}(Y). \quad (\text{iii})$$

$$\exists \alpha \in \mathbb{R} : Y \stackrel{P}{=} \alpha \Leftrightarrow \text{Var}(Y) = 0. \quad (\text{iv})$$

$$X \stackrel{P}{=} Y \Rightarrow \text{Var}(X) = \text{Var}(Y). \quad (\text{v})$$

For  $i = 1, \dots, n$ , let the random variables  $Y_i: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be independent with finite second moments and  $\alpha_i \in \mathbb{R}$ . Then

$$\text{Var}\left(\sum_{i=1}^n \alpha_i \cdot Y_i\right) = \sum_{i=1}^n \alpha_i^2 \cdot \text{Var}(Y_i). \quad (\text{vi})$$



**Figure 6.1.** Variance of an indicator of an event as a function of its probability

According to this definition,  $\text{Var}(Y)$  is the expectation of the squared *mean centered* random variable  $Y - E(Y)$ . Hence, the variance of  $Y$  is the second central moment of  $Y$ . Note that variances and standard deviations are nonnegative. The variance of  $Y$  is also denoted by  $\sigma_Y^2$  and the standard deviation by  $\sigma_Y$ . Box 6.2 summarizes some important properties of variances (see Exercise 6-3).

**Example 6.29 (Location vs. Variability)** Consider two real-valued random variables  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  with  $P(X=-1) = P(X=1) = 0.5$  and  $P(Y=-10) = P(Y=10) = 0.5$ . Then  $E(X) = E(Y) = 0$  but  $\text{Var}(X) = 1 \neq \text{Var}(Y) = 100$ . In contrast, if  $P(X=-1) = P(X=1) = 0.5$  and  $P(Y=9) = P(Y=11) = 0.5$ , then  $E(X) = 0 \neq E(Y) = 10$  but  $\text{Var}(X) = \text{Var}(Y) = 1$ . This illustrates that the expectation describes

the ‘location’ of a random variable [see Rule (ii) of Box 6.1], while the variance is invariant with respect to translations [see Rule (ii) of Box 6.2]. In contrast, the variance describes the ‘variability’ of a random variable, whereas, in general, the expectation does not.  $\triangleleft$

**Example 6.30 (Variance of an Indicator)** Let  $(\Omega, \mathcal{A}, P)$  be a probability space and let  $1_A$  denote the indicator of  $A \in \mathcal{A}$ . Then

$$\begin{aligned} \text{Var}(1_A) &= E(1_A^2) - [E(1_A)]^2 && \text{[Box 6.2, (i)]} \\ &= E(1_A) - [E(1_A)]^2 && [1_A^2 = 1_A] \\ &= E(1_A) \cdot [1 - E(1_A)] && \\ &= P(A) \cdot [1 - P(A)]. && \text{[(6.4)]} \end{aligned} \quad (6.29)$$

According to Equation (6.4), the expectation of the indicator  $1_A$  is  $P(A)$ , and Equation (6.29) shows that its variance is  $P(A) \cdot [1 - P(A)]$ . Obviously, the variance of an indicator variable does not contain any information additional to the expectation  $E(1_A) = P(A)$ . In fact, in this case,  $E(1_A)$  contains the full information about the distribution of  $1_A$ . This is not surprising because the distribution of  $1_A$  is completely determined by the single parameter  $P(A)$ . Unlike the expectation, the variance of  $1_A$  *does not* comprise the full information about the distribution of  $1_A$ . For example,  $\text{Var}(1_A) = .1 \cdot .9 = .09$  if  $P(A) = .1$  or  $P(A) = .9$ . The variance of  $1_A$  has its maximum for  $P(A) = 1/2$  and goes to 0 if  $P(A)$  approaches 0 or 1 (see Fig. 6.1).  $\triangleleft$

**Example 6.31 (Joe and Ann With Random Assignment— continued)** In the example presented in Table 2.2 (p. 53),  $X$  is an indicator variable. Hence, its variance is most easily computed as follows:

$$\text{Var}(X) = P(X=1) \cdot [1 - P(X=1)] = .40 \cdot .60 = .24.$$

Similarly, the variance of  $Y$  is obtained by

$$\text{Var}(Y) = P(Y=1) \cdot [1 - P(Y=1)] = .51 \cdot (1 - .51) = 0.2499.$$

$\triangleleft$

**Example 6.32 (Variance and Standard Error of the Sample Mean)** Let  $Y_1, \dots, Y_n$  be a sample (see Example 6.23), and  $\bar{Y}$  the sample mean [see Eq. (6.23)]. If  $E(Y_1^2) < \infty$ , and

$$\sigma_Y^2 := \text{Var}(Y_1) \quad (6.30)$$

denotes the identical variances of the  $Y_1, \dots, Y_n$ , then

$$\sigma_{\bar{Y}}^2 := \text{Var}(\bar{Y}) = \frac{\sigma_Y^2}{n} \quad (6.31)$$

(see Exercise 6-4). Hence,

$$\sigma_{\bar{Y}} := \sqrt{\sigma_{\bar{Y}}^2} = \text{SD}(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}}. \quad (6.32)$$

In statistics  $\text{SD}(\bar{Y})$  is also denoted  $\text{SE}(\bar{Y})$  and called the *standard error of the sample mean*.  $\triangleleft$

**Remark 6.33 (Z-Transformation)** Every real-valued random variable  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  with finite and positive variance  $\text{Var}(Y)$  can be transformed by

$$Z_Y := \frac{Y - E(Y)}{SD(Y)}. \quad (6.33)$$

Note that  $E(Z_Y) = 0$  and  $\text{Var}(Z_Y) = 1$  (see Exercise 6-5). This transformation is called the *Z-transformation*.

As a special case consider the mean  $\bar{Y}$  of a sample  $Y_1, \dots, Y_n$  with expectation  $\mu_Y := E(Y_1)$  and finite standard deviation  $\sigma_Y := SD(Y_1) > 0$  (see Example 6.32). Then the Z-transformation of  $\bar{Y}$  is

$$Z_{\bar{Y}} = \sqrt{n} \cdot \frac{\bar{Y} - \mu_Y}{\sigma_Y} \quad (6.34)$$

(see Exercise 6-6). The random variable  $Z_{\bar{Y}}$  will be used in the Central Limit Theorem (see Th. 8.34).  $\triangleleft$

**Remark 6.34 (The Expectation Minimizes the Mean Squared Error)** If  $E(Y^2) < \infty$ , then  $\alpha \in \mathbb{R}$  minimizes  $MSE(a) := E[(Y - a)^2]$  if and only if  $\alpha = E(Y)$ . This means that the expectation  $E(Y)$  that the expectation minimizes the mean squared error (see Exercise 6-7).  $\triangleleft$

**Definition 6.35 (Coefficient of Variation)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a nonnegative numerical random variable. If  $E(Y^2)$  is finite and  $E(Y) \neq 0$ , then

$$CV(Y) := \frac{SD(Y)}{|E(Y)|} \quad (6.35)$$

is called the *coefficient of variation* of  $Y$ .

**Remark 6.36 (Properties of the Coefficient of Variation)** This coefficient is a non-negative dimensionless quantity that expresses the variability of  $Y$  in units of its expectation. Note that sometimes the coefficient of variation is defined as  $SD(Y)/E(Y)$ . If  $\alpha \neq 0$ , then  $CV(\alpha Y) = CV(Y)$  [see Box 6.2 (iii), (6.28), and 6.1 (iii)], i. e.,  $CV(Y)$  is invariant under multiplication with (nonzero) constant. In contrast, if  $\beta \neq -E(Y)$ , then  $CV(Y + \beta) = SD(Y)/|E(Y) + \beta|$  [see Box 6.2 (ii), (6.28), and 6.1 (ii)], i. e.,  $CV(Y)$  is not invariant under translations.  $\triangleleft$

The following parameter quantifies, in a sense, the deviation of a distribution from symmetry to its expectation.

**Definition 6.37 (Skewness)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a numerical random variable. If  $E(Y^3)$  is finite and  $\text{Var}(Y) > 0$ , then

$$E\left(\frac{[Y - E(Y)]^3}{SD(Y)^3}\right) \quad (6.36)$$

is called the skewness of  $Y$ .

### 6.3 Proofs

#### **Proof of Theorem 6.11**

$$E(Y) = \int Y dP \quad [(6.1)]$$

$$= \int y P_Y(dy) \quad [(3.62)]$$

$$= \int y f_Y(y) \lambda(dy) \quad [(3.72), (5.61)]$$

$$= \int_{-\infty}^{\infty} y f_Y(y) dy. \quad [\text{Riemann integrability of } f_Y, (3.68)]$$

#### **Proof of Lemma 6.18**

For all  $A' \in \mathcal{A}'_X$  and  $g = 1_{A'}$ ,

$$\int 1_{A'}(x) P_X(dx) = P_X(A') \quad [(3.9)]$$

$$= P_{X,Y}(A' \times \Omega'_Y) \quad [(5.21)]$$

$$= \int 1_{A'}(x) \cdot 1_{\Omega'_Y}(y) P_{X,Y}[d(x, y)] \quad [(3.9), (1.38)]$$

$$= \int 1_{A'}(x) P_{X,Y}[d(x, y)]. \quad [1_{\Omega'_Y}(y) = 1]$$

Now the proposition follows applying the standard methods of proofs described in Remark 3.30.

#### **Proof of Corollary 6.22**

$$E[g(Y)] = E_Y(g) = \int g(y) P_Y(dy) \quad [(6.13)]$$

$$= \int g(y) f_Y(y) \lambda(dy) \quad [(3.72), (5.61)]$$

$$= \int_{-\infty}^{\infty} y f_Y(y) dy. \quad [\text{Riemann integrability of } y \cdot f_Y(y), (3.68)]$$

## 6.4 Exercises

▷ **Exercise 6-1** Prove the rules of computation of Box 6.1.

▷ **Exercise 6-2** Show: If, for  $n \in \mathbb{N}$ , the expectation  $E(Y^n)$  exists and is finite, then  $E(Y^m)$  exists and is finite as well for all  $1 \leq m \leq n$ .

▷ **Exercise 6-3** Prove the rules of Box 6.2.

▷ **Exercise 6-4** Prove Equations (6.25) and (6.31).

▷ **Exercise 6-5** Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a numerical random variable with finite second moment  $E(Y^2)$  and  $\text{Var}(Y) > 0$ . Show that the expectation of  $Z := [Y - E(Y)]/SD(Y)$  is 0 and its variance is 1.

▷ **Exercise 6-6** Let  $Y_i: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ ,  $i = 1, 2, \dots, n$ , be a sample with expectations  $E(Y_i) = \mu$  and finite and positive variances  $\text{Var}(Y_i) = \sigma^2$ ,  $i = 1, \dots, n$ . Furthermore, let  $\bar{Y} := S/n$  be the arithmetic mean, where  $S := \sum_{i=1}^n Y_i$ , and

$$\bar{Z} := \frac{(\bar{Y} - \mu) \cdot \sqrt{n}}{\sigma}.$$

Show that  $E(\bar{Z}) = 0$  and  $\text{Var}(\bar{Z}) = 1$ .

▷ **Exercise 6-7** Assume  $E(Y^2) < \infty$  and show that  $\alpha \in \mathbb{R}$  minimizes  $MSE(a) := E[(Y - a)^2]$  if and only if  $\alpha = E(Y)$ .

## Solutions

▷ **Solution 6-1** Because the expectation of a numerical random variable is defined as an integral, we simply can refer to the corresponding propositions of chapter 3.

(i) This is Equation (3.8) with  $A = \Omega$  and  $\mu(A) = P(\Omega) = 1$ .

(ii) This is a special case of Equation (3.34).

(iii) This is a special case of Equation (3.32).

(iv) Note that  $1_A \cdot 1_B = 1_{A \cap B}$ . Hence, this rule follows from Equation (3.8), with  $\alpha = 1$ .

(v) This is a special case of Lemma 3.45.

(vi) If  $Y_1, Y_2$  are nonnegative, then this equation is a special case of Equation (3.65). If  $Y_1$  or  $Y_2$  has a finite expectation, then this equation is a special case of Equation (3.33).

(vii) This rule follows from Equation (3.34) and complete induction.

(viii) This is a special case of Lemma 3.47.

(ix) This is a special case of Theorem 3.48.

(x) This is a special case of Theorem 6.24.

▷ **Solution 6-2** Assume that, for  $n \in \mathbb{N}$ , the expectation  $E(Y^n)$  exists and is finite. Furthermore, let  $A := \{\omega \in \Omega: |Y(\omega)| > 1\}$ . Now, for all  $m$  with  $1 \leq m \leq n$ ,

$$|Y(\omega)^m| \leq |Y(\omega)^n|, \quad \forall \omega \in A \quad \text{and} \quad |Y(\omega)^m| \leq 1, \quad \forall \omega \in A^c.$$

Therefore, applying Corollary 3.38, for  $1 \leq m \leq n$ :

$$E(|Y^m|) = \int |Y^m| dP \quad [(6.1)]$$

$$= \int_A |Y^m| dP + \int_{A^c} |Y^m| dP \quad [(3.36)]$$

$$\leq \int_A |Y^n| dP + \int_{A^c} 1 dP \quad [\text{Lemma 3.26}]$$

$$\leq \int_A |Y^n| dP + 1 \quad [(3.30), 1_{A^c} \leq 1, \text{Lemma 3.26}]$$

$$< \infty.$$

▷ **Solution 6-3** Rules (i), (ii), and (iii) are special cases of rules (i), (ii), and (iii) of Box 7.1, which are proved in Exercise 7-3.

(iv) ⇒: This proposition follows from (vii) of Box 7.1.

⇐:

$$\text{Var}(Y) = 0 \Rightarrow [Y - E(Y)]^2 \stackrel{p}{=} 0 \quad [(6.27), \text{Th. 3.43}]$$

$$\Rightarrow Y - E(Y) \stackrel{p}{=} 0 \quad [[Y - E(Y)]^2 = 0 \Leftrightarrow Y - E(Y) = 0]$$

$$\Rightarrow Y \stackrel{p}{=} E(Y) \in \mathbf{R}. \quad [E(Y)^2 < \infty \Rightarrow E(Y) \in \mathbf{R}, \text{Rem. 6.25 (iii)}]$$

(v)  $X \stackrel{p}{=} Y \Rightarrow X^2 \stackrel{p}{=} Y^2$  [see (2.34)]. Hence,  $X \stackrel{p}{=} Y$  implies  $E(X) = E(Y)$  and  $E(X^2) = E(Y^2)$  (see Cor. 6.17). Therefore, rule (i) of Box 6.2 yields  $\text{Var}(X) = \text{Var}(Y)$ .

(vi) This proposition follows from rules (ix) and (vi) of Box 7.1.

▷ **Solution 6-4**

Equation (6.25) can be derived as follows:

$$E(\bar{Y}) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n Y_i\right) \quad [(6.23)]$$

$$= \frac{1}{n} \cdot \sum_{i=1}^n E(Y_i) \quad [\text{Box 6.1 (vii)}]$$

$$= \frac{1}{n} \cdot n \cdot \mu_Y = \mu_Y. \quad [Y_1, \dots, Y_n \text{ are identically distributed, (6.24)}]$$

Equation (6.31) can be derived as follows:

$$\text{Var}(\bar{Y}) = \text{Var}\left(\frac{1}{n} \cdot \sum_{i=1}^n Y_i\right) \quad [(6.23)]$$

$$= \frac{1}{n^2} \cdot \sum_{i=1}^n \text{Var}(Y_i) \quad [\stackrel{p}{=} Y_1, \dots, Y_n, \text{Box 6.2 (vi)}]$$

$$= \frac{1}{n^2} \cdot n \cdot \sigma_Y^2 = \frac{\sigma_Y^2}{n}. \quad [Y_1, \dots, Y_n \text{ are identically distributed, (6.30)}]$$

▷ **Solution 6-5** Let  $\mu := E(Y)$  and  $\sigma := SD(Y)$ . Then

$$E(Z) = E\left(\frac{Y - \mu}{\sigma}\right) = E\left(\frac{1}{\sigma} \cdot (Y - \mu)\right) = \frac{1}{\sigma} \cdot E(Y - \mu) \quad [\text{Box 6.1 (iii)}]$$

$$= \frac{1}{\sigma} \cdot [E(Y) - E(\mu)] = \frac{1}{\sigma} \cdot (\mu - \mu) = 0 \quad [\text{Box 6.1 (vi), (i)}].$$

$$\begin{aligned} \text{Var}(Z) &= \text{Var}\left(\frac{Y-\mu}{\sigma}\right) = \text{Var}\left(\frac{1}{\sigma} \cdot (Y-\mu)\right) \\ &= \frac{1}{\sigma^2} \cdot \text{Var}(Y) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1 \quad [\text{Box 6.2 (iii), (ii)}]. \end{aligned}$$

▷ **Solution 6-6** Using Equations (6.25), (6.32),  $E(Z_Y) = 0$ , and  $\text{Var}(Z_Y) = 1$  (see Exercise 6-5) yields

$$E(\bar{Z}) = E\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right) = 0 \quad \text{and} \quad \text{Var}(\bar{Z}) = \text{Var}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right) = 1.$$

▷ **Solution 6-7** For all  $\alpha \in \mathbb{R}$ , using Box 6.1 (iii), (ii),

$$\begin{aligned} E[(Y-\alpha)^2] &= E[(Y-E(Y)+E(Y)-\alpha)^2] \\ &= E[(Y-E(Y))^2] + [E(Y)-\alpha]^2 + 2 \cdot E[(Y-E(Y)) \cdot (E(Y)-\alpha)] \\ &= E[(Y-E(Y))^2] + [E(Y)-\alpha]^2 \\ &\geq E[(Y-E(Y))^2], \end{aligned}$$

and '=' holds if and only if  $E(Y) = \alpha$ .



## Chapter 7

# Linear Quasi-Regression, Covariance, and Correlation

Expectation and variance are parameters that describe important properties of a univariate numerical random variable and its distribution. Now we consider *two* numerical random variables, say  $X$  and  $Y$ , and their joint distribution. In other words, we consider the distribution of the bivariate real-valued random variable  $(X, Y): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^2, \mathcal{B}_2)$ . We also introduce a new random variable that can be used to describe a specific kind of dependence of  $Y$  on  $X$ . It is the kind of dependence of  $Y$  on  $X$  that is represented by the best fitting linear function  $\alpha_0 + \alpha_1 X$ , ‘best fitting’ in terms of the minimal mean squared error. This function is the composition of  $X$  and the *linear quasi-regression* or the *linear least-squares regression*. *Covariance* and *correlation* are important parameters quantifying the strength of the kind of dependence that can be described by a linear quasi-regression.

### 7.1 Linear Quasi-Regression

**Remark 7.1 (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 Klenke, 2013, Remark 5.2, p. 102). Hence, according to Remark 3.42, there is no substantial loss of generality if we additionally assume that  $X$  and  $Y$  are real-valued.  $\triangleleft$

#### Definition 7.2 (Linear Quasi-Regression)

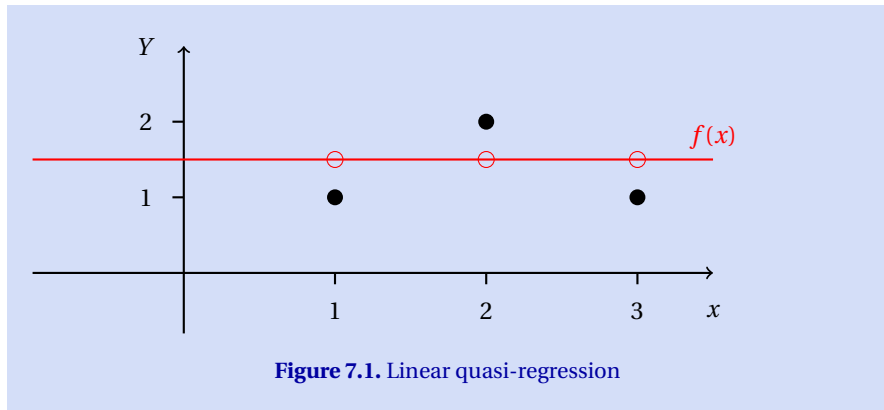
Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be two real-valued random variables, assume  $E(X^2), E(Y^2) < \infty$ , and  $\text{Var}(X) > 0$ . Then the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$f(x) = \alpha_0 + \alpha_1 x, \quad \forall x \in \mathbb{R}, \quad (7.1)$$

where the pair  $(\alpha_0, \alpha_1)$  minimizes the function  $\text{MSE}: \mathbb{R}^2 \rightarrow \mathbb{R}$  with

$$\text{MSE}(a_0, a_1) = E([Y - (a_0 + a_1 X)]^2), \quad \forall a_0, a_1 \in \mathbb{R}, \quad (7.2)$$

is called the *linear quasi-regression* of  $Y$  on  $X$ . The composition of  $X$  and  $f$  is denoted by  $Q_{\text{lin}}(Y|X)$ , i. e.,



$$Q_{lin}(Y|X) = f(X) = \alpha_0 + \alpha_1 X. \quad (7.3)$$

**Remark 7.3 (Distinguishing Between  $f$  and  $f(X)$ )** To emphasize, the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  is called a linear quasi-regression. In this context the random variable  $X$  is called the *regressor* and  $Y$  the *regressand*. Note that  $f$  is a function assigning a real number to all real numbers. This applies even if  $X$  only takes on two different real values. In contrast, the number of different values of the composition  $Q_{lin}(Y|X) = f(X): \Omega \rightarrow \mathbb{R}$  is smaller or equal to the number of values of  $X$ , provided that  $X$  takes on a finite number of values only.  $\triangleleft$

**Remark 7.4 (Coefficient of Determination)** Under the assumptions of Definition 7.2 and  $\text{Var}(Y) > 0$ , we define

$$Q_{Y|X}^2 := \frac{\text{Var}[Q_{lin}(Y|X)]}{\text{Var}(Y)} \quad (7.4)$$

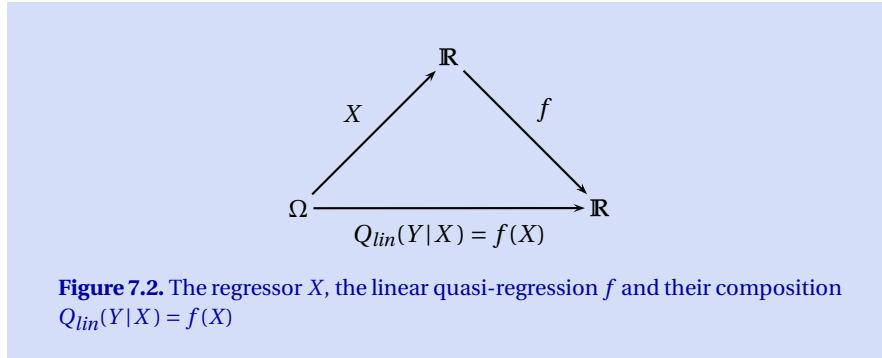
and it call the *coefficient of determination* of the linear quasi-regression  $Q_{lin}(Y|X)$ . In Remark 7.29, this definition is extended to the case in which  $X$  is an  $n$ -dimensional real-valued random variable.  $\triangleleft$

**Example 7.5 (Discrete Regressor With Three Different Values)** Let  $X$  and  $Y$  be real-valued random variables on  $(\Omega, \mathcal{A}, P)$  with values 1, 2, 3 and 1, 2, respectively. Furthermore, assume that their distribution is specified by

$$P(X=1, Y=1) = .25, \quad P(X=2, Y=2) = .5, \quad P(X=3, Y=1) = .25.$$

Then the linear quasi-regression  $f: \mathbb{R} \rightarrow \mathbb{R}$  is specified by

$$f(x) = \alpha_0 + \alpha_1 \cdot x = 1.5 + 0 \cdot x = 1.5, \quad \forall x \in \mathbb{R},$$



and the composition of  $X$  and  $f$  is

$$Q_{lin}(Y|X) = \alpha_0 + \alpha_1 \cdot X = 1.5 + 0 \cdot X = 1.5$$

(see Exercise 7-1). The black points in Figure 7.1 represent the three pairs of values of  $X$  and  $Y$ . All values of the the linear quasi-regression are on the horizontal quasi-regression line, which, in this example, is parallel to  $x$ -axis because its slope is 0. The circles on this line represent the values  $f(x)$  for  $x = 1, 2, 3$ , i. e., for those values of  $X$  with a nonzero probability  $P_X(\{x\}) > 0$ . In contrast, in this example,  $P_X(\{x\}) = 0$ , for all  $x \in \mathbb{R} \setminus \{1, 2, 3\}$ . Nevertheless, as mentioned before, a linear quasi-regression  $f$  is a function assigning a real number *to all real numbers*.  $\triangleleft$

**Remark 7.6 (Linear Quasi-Regression vs. Regression)** As the term ‘linear quasi-regression’ suggests, there is also a ‘genuine’ regression of  $Y$  on  $X$  (see Def. 10.25) and the two terms are not necessarily identical. As will be explained in more detail in Remark 10.27, the ‘genuine’ regression is a function  $g: \mathbb{R} \rightarrow \mathbb{R}$  such that the composition  $g(X)$  is  $X$ -measurable and minimizes the mean squared error  $E([Y - g(X)]^2)$ . In contrast to the conditional expectation  $E(Y|X)$  [see Eq. (10.1)], the composition  $Q_{lin}(Y|X) = f(X)$  minimizes the function  $MSE$  specified in Equation (7.2). Hence,  $f$  has to be a *linear* function even in those cases in which there are no  $a_0, a_1 \in \mathbb{R}$ , such that  $E(Y|X) \stackrel{p}{=} a_0 + a_1 X$  (see Example 7.5).  $\triangleleft$

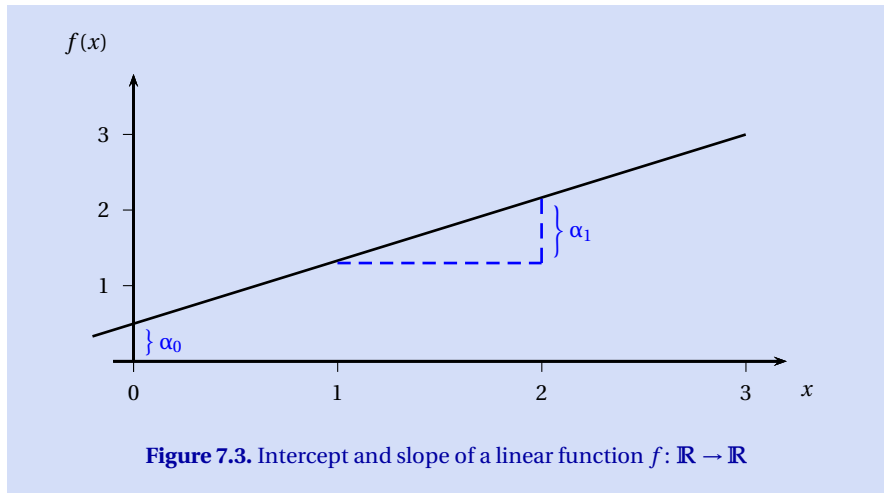
**Remark 7.7 (Intercept and Slope)** Note that the composition  $Q_{lin}(Y|X) = f(X) = \alpha_0 + \alpha_1 X$  is a random variable on  $(\Omega, \mathcal{A}, P)$  (see Fig. 7.2). The coefficient  $\alpha_0$  is called the *intercept* and  $\alpha_1$  the *slope* of (the linear quasi-regression)  $f$  (see Fig. 7.3). Obviously,

$$f(0) = \alpha_0 + \alpha_1 \cdot 0 = \alpha_0. \quad (7.5)$$

Furthermore, if  $x_1, x_2 \in \mathbb{R}$  and  $x_2 > x_1$ , then

$$\alpha_1 = \frac{1}{x_2 - x_1} \cdot [f(x_2) - f(x_1)] \quad (7.6)$$

(see Exercise 7-2). Equation (7.6) yields



$$\alpha_1 = f(x_2) - f(x_1), \quad \text{if } x_2 - x_1 = 1. \quad (7.7)$$

These equations justify calling  $\alpha_0$  the *intercept* and  $\alpha_1$  the *slope* of the linear quasi-regression  $f$  (see Fig. 7.3). Note that these equations also apply if  $P(X=0) = P(X=x_1) = P(X=x_2) = 0$ . They even apply if  $0, x_1, x_2 \notin X(\Omega)$ , because, by definition,  $f$  is a function on  $\mathbb{R}$ .

Figure 7.3 illustrates the intercept and the slope of a linear function such as the linear quasi-regression  $f$ . In this figure,  $\alpha_0 = .5$  and  $\alpha_1 = .85$ . If  $X$  is discrete, then  $Q_{lin}(Y|X) = f(X)$  is discrete as well. More precisely, the number of different values of  $Q_{lin}(Y|X)$  is always smaller than or equal to the number of different values of  $X$ . In contrast, the linear quasi-regression  $f: \mathbb{R} \rightarrow \mathbb{R}$  takes on uncountably many values unless its slope is 0. In this case its sole value is  $\alpha_0$ .  $\triangleleft$

## 7.2 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, i. e., the degree to which the two variables vary together in the following sense: If one variable takes on a large value (i. e., large positive deviation from its expectation), then the other one tends to take on a large value as well. Furthermore, if one variable takes on a small value (i. e., 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.

**Definition 7.8 (Covariance)**

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \mathcal{B})$  be two numerical random variables with  $E(X^2), E(Y^2) < \infty$ . Then the covariance of  $X$  and  $Y$  is defined by

$$\text{Cov}(X, Y) := E([X - E(X)] \cdot [Y - E(Y)]). \quad (7.8)$$

Comparing Equations (7.8) and (6.27) to each other shows that the variance is the covariance of a numerical random variable with itself.

**Remark 7.9 (Correlated Numerical Random Variables)** According to this definition, the *covariance* of  $X$  and  $Y$  is the expectation of the product of the 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*; otherwise, we say that they are *uncorrelated*.  $\triangleleft$

**Remark 7.10 (Rules of Computation)** The most important rules of computation for covariances are summarized in Box 7.1. Proofs are provided in Exercise 7-3. Rule (i) immediately implies

$$E(X \cdot Y) = E(X) \cdot E(Y) + \text{Cov}(X, Y). \quad (7.9)$$

Hence,  $X$  and  $Y$  are uncorrelated if and only if  $E(X \cdot Y) = E(X) \cdot E(Y)$  [see Rule (vi)]. Furthermore, this equation and Theorem 6.24 imply that  $X$  and  $Y$  are uncorrelated if  $X$  and  $Y$  are independent.

Symmetry of the covariance [see Box 7.1 (v)] yields an alternative way to write Rule (viii) of Box 7.1:

$$\text{Var}\left(\sum_{i=1}^n \alpha_i Y_i\right) = \sum_{i=1}^n \alpha_i^2 \text{Var}(Y_i) + 2 \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^n \alpha_i \alpha_j \text{Cov}(Y_i, Y_j). \quad (7.10)$$

This equation simplifies to

$$\text{Var}\left(\sum_{i=1}^n \alpha_i Y_i\right) = \sum_{i=1}^n \alpha_i^2 \text{Var}(Y_i), \quad (7.11)$$

if  $Y_1, \dots, Y_n$  is a sequence of pairwise uncorrelated random variables. Note that independence of  $Y_1, \dots, Y_n$  implies  $\text{Cov}(Y_i, Y_j) = 0$ ,  $i \neq j$  [see Rule (vi)].

For  $n = 2$ , Rule (viii) simplifies to

$$\begin{aligned} & \text{Var}(\alpha_1 Y_1 + \alpha_2 Y_2) \\ &= \alpha_1^2 \text{Var}(Y_1) + \alpha_2^2 \text{Var}(Y_2) + 2\alpha_1 \alpha_2 \cdot \text{Cov}(Y_1, Y_2). \end{aligned} \quad (7.12)$$

Similarly, for  $n = m = 2$ , Rule (ix) simplifies to

**Box 7.1 Rules of Computation for Covariances**

Let  $X, Y$  be numerical random variables on the probability space  $(\Omega, \mathcal{A}, P)$  with  $E(X^2), E(Y^2) < \infty$ . Furthermore, let  $\alpha, \beta \in \mathbb{R}$ . Then:

$$\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y). \quad (\text{i})$$

$$\text{Cov}(\alpha + X, \beta + Y) = \text{Cov}(X, Y). \quad (\text{ii})$$

$$\text{Cov}(\alpha X, \beta Y) = \alpha \beta \text{Cov}(X, Y). \quad (\text{iii})$$

$$\text{Cov}(X, X) = \text{Var}(X). \quad (\text{iv})$$

$$\text{Cov}(X, Y) = \text{Cov}(Y, X). \quad (\text{v})$$

$$X \perp\!\!\!\perp Y \Rightarrow \text{Cov}(X, Y) = 0. \quad (\text{vi})$$

$$\exists \alpha \in \mathbb{R}: X \stackrel{p}{=} \alpha \Rightarrow \text{Cov}(X, Y) = 0. \quad (\text{vii})$$

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

$$\text{Var}\left(\sum_{i=1}^n \alpha_i Y_i\right) = \sum_{i=1}^n \alpha_i^2 \text{Var}(Y_i) + \sum_{i=1}^n \sum_{j=1, i \neq j}^n \alpha_i \alpha_j \text{Cov}(Y_i, Y_j). \quad (\text{viii})$$

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

$$\text{Cov}\left(\sum_{i=1}^n \alpha_i X_i, \sum_{j=1}^m \beta_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j \text{Cov}(X_i, Y_j). \quad (\text{ix})$$

If  $X_1 \stackrel{p}{=} X_2$  and  $E(Y^2), E(X_1^2), E(X_2^2) < \infty$ , then

$$\text{Cov}(Y, X_1) = \text{Cov}(Y, X_2). \quad (\text{x})$$

$$\begin{aligned} & \text{Cov}(\alpha_1 X_1 + \alpha_2 X_2, \beta_1 Y_1 + \beta_2 Y_2) \\ &= \alpha_1 \beta_1 \text{Cov}(X_1, Y_1) + \alpha_1 \beta_2 \text{Cov}(X_1, Y_2) \\ & \quad + \alpha_2 \beta_1 \text{Cov}(X_2, Y_1) + \alpha_2 \beta_2 \text{Cov}(X_2, Y_2). \end{aligned} \quad (7.13)$$

&lt;

**Remark 7.11 (Covariance of Indicators)** For  $A, B \in \mathcal{A}$ , Rule (i) of Box 7.1 and Equations (1.33) and (6.4) yield

$$\text{Cov}(1_A, 1_B) = E(1_A \cdot 1_B) - E(1_A) \cdot E(1_B) \quad (7.14)$$

$$= P(A \cap B) - P(A) \cdot P(B). \quad (7.15)$$

&lt;

The following theorem helps to understand the relationship between the covariance and the variances of two numerical random variables  $X$  and  $Y$ .

**Theorem 7.12 (Cauchy-Schwarz Inequality)**

If  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  are random variables with  $E(X^2), E(Y^2) < \infty$ , then

$$\text{Cov}(X, Y)^2 \leq \text{Var}(X) \cdot \text{Var}(Y). \quad (7.16)$$

Furthermore, if  $\text{Cov}(X, Y) \neq 0$ , then

$$\text{Cov}(X, Y)^2 = \text{Var}(X) \cdot \text{Var}(Y) \Leftrightarrow \exists a, b \in \mathbb{R}: Y \stackrel{P}{=} a + b \cdot X. \quad (7.17)$$

(Proof p. 233)

**Remark 7.13 (Squared Weighted Sum of Random Variables)** If  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  are random variables with  $E(X^2), E(Y^2) < \infty$  and  $\alpha, \beta \in \mathbb{R}$ , then  $E[(\alpha X + \beta Y)^2] < \infty$  (see Exercise 7-4).  $\triangleleft$

In the following theorem we revisit the linear quasi-regression, studying three equivalent propositions. The first of these propositions deals with the residual variable  $\epsilon := Y - f(X)$ , where  $f$  is the linear quasi-regression of  $Y$  on  $X$ . Note that this residual is not necessarily identical to the residual with respect to a conditional expectation that will be treated in chapters 9 to 11.

**Theorem 7.14 (Three Characterizations of the Linear Quasi-Regression)**

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be two real-valued random variables with  $E(X^2), E(Y^2) < \infty$ , and  $\text{Var}(X) > 0$ . Furthermore, let  $\alpha_0, \alpha_1 \in \mathbb{R}$ ,  $f(X) = \alpha_0 + \alpha_1 X$  be the composition of  $X$  and  $f: \mathbb{R} \rightarrow \mathbb{R}$ , and define  $\epsilon := Y - f(X)$ . Then the following three propositions are equivalent to each other:

(i)  $E(\epsilon) = \text{Cov}(X, \epsilon) = 0$ .

(ii)  $\alpha_0 = E(Y) - \alpha_1 E(X)$  and  $\alpha_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$ .

(iii)  $f(X) = Q_{lin}(Y|X)$ , i. e.,  $\alpha_0, \alpha_1$  minimize the function  $MSE(a_0, a_1)$  defined by Equation (7.2).

(Proof p. 234)

**Remark 7.15 (Uniqueness)** Suppose that the assumptions of Theorem 7.14 hold and  $f(X) = Q_{lin}(Y|X)$ . Then proposition (ii) of this theorem implies that the coefficients  $\alpha_0$  and  $\alpha_1$  are uniquely defined. Because  $Q_{lin}(Y|X) = \alpha_0 + \alpha_1 X$ , the linear quasi-regression  $f: \mathbb{R} \rightarrow \mathbb{R}$  is uniquely defined as well.  $\triangleleft$

**Remark 7.16 (Relationship Between Slope and Covariance)** According to proposition (ii), a *zero covariance* between  $X$  and  $Y$  implies that the slopes of the linear quasi-regressions of  $Y$  on  $X$  and of  $X$  on  $Y$  are zero, provided that  $\text{Var}(Y) >$

0. A *negative covariance* implies that the slopes of the linear quasi-regressions of  $Y$  on  $X$  and of  $X$  on  $Y$  are negative, and a *positive covariance* implies that the slopes of the linear quasi-regressions of  $Y$  on  $X$  and of  $X$  on  $Y$  are positive.  $\triangleleft$

**Example 7.17 (Discrete Regressor With Three Different Values – continued)** In Example 7.5 we specified the distribution of  $(X, Y)$ . Now we use the equations in Theorem 7.14 (ii) in order to compute the coefficients  $\alpha_0$  and  $\alpha_1$ . For this purpose we have to compute the expectations of  $X$  and  $Y$ , the variance of  $X$ , and the covariance of  $X$  and  $Y$ . Hence, with  $n = 3$ ,  $m = 2$ ,  $x_1 = 1, x_2 = 2, x_3 = 3$ , and  $y_1 = 1, y_2 = 2$ ,

$$E(X) = \sum_{i=1}^n x_i \cdot P(X=x_i) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} = 2,$$

$$E(Y) = \sum_{i=1}^m y_i \cdot P(Y=y_i) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2},$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 = \sum_{i=1}^n x_i^2 \cdot P(X=x_i) - E(X)^2 \quad [\text{Box 6.2 (i), (6.19)}] \\ &= 1^2 \cdot \frac{1}{4} + 2^2 \cdot \frac{1}{2} + 3^2 \cdot \frac{1}{4} - 2^2 = \frac{1}{2}, \end{aligned}$$

$$\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) \quad [\text{Box 7.1 (i)}]$$

$$= \sum_{i=1}^n \sum_{j=1}^m x_i \cdot y_j \cdot P((X, Y)=(x_i, y_j)) - E(X) \cdot E(Y) \quad [(6.3)]$$

$$= 1 \cdot 1 \cdot \frac{1}{4} + 2 \cdot 2 \cdot \frac{1}{2} + 3 \cdot 1 \cdot \frac{1}{4} - 2 \cdot \frac{3}{2}$$

$$= \frac{1}{4} + \frac{8}{4} + \frac{3}{4} - \frac{12}{4} = 0.$$

Using the equations in Theorem 7.14 (ii) yields  $\alpha_1 = \text{Cov}(X, Y) / \text{Var}(X) = \frac{0}{1/2} = 0$  and  $\alpha_0 = E(Y) - \alpha_1 E(X) = \frac{3}{2} - 0 \cdot 2 = 1.5$ , the same result as obtained in Exercise 7-1, in which we minimize the function  $MSE(a_0, a_1)$ .  $\triangleleft$

### 7.3 Correlation

As mentioned before, the covariance between two numerical random variables quantifies the strength of the dependence that can be described by a linear quasi-regression. However, the covariance is not invariant under multiplication with constants [scale transformations; see Box 7.1 (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 (see Rem. 7.22).

#### Definition 7.18 (Correlation)

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be two numerical random variables with  $E(X^2)$ ,

$E(Y^2) < \infty$ . Then the correlation of  $X$  and  $Y$  is defined by

$$\text{Corr}(X, Y) := \begin{cases} \frac{\text{Cov}(X, Y)}{SD(X) \cdot SD(Y)}, & \text{if } SD(X), SD(Y) > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (7.18)$$

**Remark 7.19 (Correlation of a Random Variable With Itself)** Assume  $\text{Var}(X) > 0$ . Because  $\text{Cov}(X, X) = \text{Var}(X) = SD(X) \cdot SD(X)$ , Equation (7.18) implies that  $\text{Corr}(X, X) = 1$ . Similarly, because  $\text{Cov}(X, -X) = -\text{Var}(X) = -SD(X) \cdot SD(X)$ , Equation (7.18) implies that  $\text{Corr}(X, -X) = -1$ .  $\triangleleft$

**Remark 7.20 (Range of the Correlation)** An implication of (7.16) is

$$-1 \leq \text{Corr}(X, Y) \leq 1, \quad (7.19)$$

provided that  $\text{Corr}(X, Y)$  exists, i. e., provided that the assumptions hold under which the correlation is defined (see Exercise 7-5).  $\triangleleft$

**Remark 7.21 (Correlation and  $Z$ -Transformed Variables)** If the standard deviations of  $X$  and  $Y$  are positive, then the correlation is also the expectation of the product of the  $Z$ -transformed variables [see Eq. (6.33)], i. e.,

$$\text{Corr}(X, Y) = E\left(\frac{X - E(X)}{SD(X)} \cdot \frac{Y - E(Y)}{SD(Y)}\right) \quad (7.20)$$

(see Exercise 7-6).  $\triangleleft$

**Remark 7.22 (An Invariance Property of the Correlation)** The correlation of linear transformations of  $X$  and  $Y$  is

$$\text{Corr}(a_0 + a_1 X, b_0 + b_1 Y) = \begin{cases} \text{Corr}(X, Y), & \text{if } a_1 \cdot b_1 > 0 \\ -\text{Corr}(X, Y), & \text{if } a_1 \cdot b_1 < 0, \\ 0, & \text{if } a_1 \cdot b_1 = 0 \end{cases} \quad (7.21)$$

where  $a_0, a_1, b_0, b_1 \in \mathbb{R}$  (see Exercise 7-7). This equation implies that the correlation is invariant (up to change of signs) under linear transformations, which include *translations* ( $a_1 = 1$  and  $b_1 = 1$ ) and *scale transformations* ( $a_0 = b_0 = 0$  and  $a_1, b_1 \neq 0$ ).  $\triangleleft$

Theorem 7.12 implies the following corollary about the cases  $\text{Corr}(X, Y) = 1$  or  $\text{Corr}(X, Y) = -1$ , i. e., the ‘perfect’ correlation.

**Corollary 7.23 (Perfect Correlation of Two Random Variables)**

Let the assumptions of Definition 7.8 be satisfied and suppose that  $\text{Var}(X), \text{Var}(Y) > 0$ . Then  $|\text{Corr}(X, Y)| = 1$  if and only if there are  $a_0, a_1 \in \mathbb{R}$ ,  $a_1 \neq 0$ , such that  $Y \stackrel{p}{=} a_0 + a_1 X$ . In this case

$$\text{Corr}(X, Y) = \begin{cases} 1, & \text{if } a_1 > 0 \\ -1, & \text{if } a_1 < 0. \end{cases} \quad (7.22)$$

**Remark 7.24 (Covariance and Standard Deviations)** Let  $Y \stackrel{p}{=} a_0 + a_1 X$ ,  $a_0, a_1 \in \mathbb{R}$ . If  $a_1 \geq 0$ , then  $\text{Cov}(X, Y) = \text{SD}(X) \cdot \text{SD}(Y)$ . If  $a_1 < 0$  then  $\text{Cov}(X, Y) = -\text{SD}(X) \cdot \text{SD}(Y)$  (see Exercise 7-8).  $\triangleleft$

**Remark 7.25 (Slope of a Linear Quasi-Regression and Correlation)** If  $\alpha_1$  is the slope of  $Q_{lin}(Y|X)$  (see Def. 7.2), then

$$\alpha_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \text{Corr}(X, Y) \cdot \frac{\text{SD}(Y)}{\text{SD}(X)} \quad (7.23)$$

[see proposition (ii) of Th. 7.14]. This equation shows that the slope  $\alpha_1$  of the linear quasi-regression has the same sign as the covariance and the correlation. The size of the absolute value of  $\alpha_1$  depends on the ratio  $\text{SD}(Y)/\text{SD}(X)$  of the standard deviations. The smaller the standard deviation of  $X$  compared to the standard deviation of  $Y$ , the larger the absolute value of  $\alpha_1$ . Furthermore, given a fixed variance  $\text{Var}(X)$ , this equation also shows that  $\alpha_1$  is proportional to  $\text{Cov}(X, Y)$  and  $\text{Corr}(X, Y)$ . In this sense, all three parameters  $\alpha_1$ ,  $\text{Cov}(X, Y)$ , and  $\text{Corr}(X, Y)$  quantify the strength of the dependence of  $Y$  on  $X$  described by a linear quasi-regression. Note, however, that  $\alpha_1$  and  $\text{Cov}(X, Y)$  are not invariant under scale transformations of  $X$  and  $Y$ . This can be seen in the following equation for the slope  $\alpha_1^*$  of the linear quasi-regression of  $bY$  on  $aX$ ,  $a, b \in \mathbb{R}$ ,  $a, b \neq 0$ :

$$\alpha_1^* = \frac{\text{Cov}(aX, bY)}{\text{Var}(aX)} = \text{Corr}(aX, bY) \cdot \frac{\text{SD}(bY)}{\text{SD}(aX)} = \frac{b}{a} \cdot \alpha_1. \quad (7.24)$$

Hence, the slope of the linear quasi-regression of  $bY$  on  $aX$  is identical to the slope of the linear quasi-regression of  $Y$  on  $X$  multiplied by  $\frac{b}{a}$ . In contrast, the slope of the linear quasi-regression is invariant under translations  $c + X$ ,  $d + Y$ ,  $c, d \in \mathbb{R}$  (see Exercise 7-9).  $\triangleleft$

**Remark 7.26 (Correlation and Coefficient of Determination)** Under the assumptions of Definition 7.18 and  $\text{Var}(X), \text{Var}(Y) > 0$ ,

$$\begin{aligned} Q_{Y|X}^2 &= \frac{\text{Var}[Q_{lin}(Y|X)]}{\text{Var}(Y)} = \frac{\text{Var}[\alpha_0 + \alpha_1 \text{Var}(X)]}{\text{Var}(Y)} = \frac{\alpha_1^2 \text{Var}(X)}{\text{Var}(Y)} \\ &= \text{Corr}(X, Y)^2. \end{aligned} \quad (7.25)$$

 $\triangleleft$

**Example 7.27 (Joe and Ann With Random Assignment – continued)** Consider the example presented in Table 2.2 (p. 53). In this example, the covariance of  $X$  and  $Y$  is most easily computed using

$$\begin{aligned} \text{Cov}(X, Y) &= E(X \cdot Y) - E(X) \cdot E(Y) && \text{[Box 7.1 (i)]} \\ &= \sum_{(x,y)} (x \cdot y) \cdot P(X=x, Y=y) - P(X=1) \cdot P(Y=1) && \text{[(6.15)]} \\ &= P(X=1, Y=1) - P(X=1) \cdot P(Y=1) \\ &= (.16 + .08) - .40 \cdot .51 = .036, \end{aligned}$$

where  $P(X=1) = E(X) = .40$  and  $P(Y=1) = E(Y) = .51$  have been computed in Example 6.7. Note that  $\sum_{(x,y)}$  is the sum over all pairs  $(x, y)$  of values of  $X$  and  $Y$ . In this example, there are four such pairs, only one of which, namely  $(1, 1)$ , yields a product  $x \cdot y \neq 0$ . Using the results of Example 6.31 on the variances of  $X$  and  $Y$  yields the correlation

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X) \cdot \text{SD}(Y)} = \frac{.036}{\sqrt{.24} \cdot \sqrt{.2499}} \approx .147$$

Hence, treatment and outcome variables have a positive correlation. This is in accordance with comparing the conditional probability of success given treatment,  $P(C|B) = .60$ , to the conditional probability of success given no treatment,  $P(C|B^c) = .45$  (see Example 4.16).

In this example,

$$\begin{aligned} Q_{lin}(Y|X) &= \alpha_0 + \alpha_1 \cdot X \\ &= \left( E(Y) - \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \cdot E(X) \right) + \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \cdot X && \text{[Th. 7.14 (ii)]} \\ &= \left( .51 - \frac{.036}{.24} \cdot .40 \right) + \frac{.036}{.24} \cdot X \\ &= .45 + .15 \cdot X, \end{aligned}$$

and the linear quasi-regression  $f: \mathbb{R} \rightarrow \mathbb{R}$  of  $Y$  on  $X$  is specified by

$$f(x) = .45 + .15 \cdot x, \quad \forall x \in \mathbb{R}.$$

◁

## 7.4 Expectation Vector and Covariance Matrix

### 7.4.1 Random Vector and Random Matrix

Let  $X = (X_1, \dots, X_n)$  be an  $n$ -variate numerical random variable on a probability space  $(\Omega, \mathcal{A}, P)$ . In order to utilize matrix algebra, we consider the *column random vector*

$$\mathbf{x} := \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix},$$

i. e., the column vector of the random variables  $X_1, \dots, X_n$ . Correspondingly, we consider the *row random vector*  $\mathbf{x}' := [X_1, \dots, X_n]$ , the *transpose* of  $\mathbf{x}$ .

In this section, we also consider a *random matrix*, i. e., a matrix

$$\mathbf{X} := \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix} \quad (7.26)$$

of type  $n \times m$  of numerical random variables  $X_{ij}$  on a probability space  $(\Omega, \mathcal{A}, P)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ . Such a random matrix is an  $n \times m$ -array of an  $(n \cdot m)$ -variate random variable (see section 5.3).

#### 7.4.2 Expectation of a Random Vector and a Random Matrix

The *expectation of a (row or column) random vector* is defined as the (row or column) vector of the expectations of its components, i. e.,

$$E(\mathbf{x}') := [E(X_1), \dots, E(X_n)] \quad (7.27)$$

and  $E(\mathbf{x}) := [E(X_1), \dots, E(X_n)]'$ , provided that the expectations exist. Hence,

$$E(\mathbf{x}') = (E(\mathbf{x}))'. \quad (7.28)$$

Analogously to Equation (7.27), the *expectation of an  $n \times m$ -random matrix* is defined as the  $n \times m$ -matrix of the expectations of its components, i. e.,

$$E \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix} := \begin{bmatrix} E(X_{11}) & E(X_{12}) & \dots & E(X_{1m}) \\ E(X_{21}) & E(X_{22}) & \dots & E(X_{2m}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{n1}) & E(X_{n2}) & \dots & E(X_{nm}) \end{bmatrix}, \quad (7.29)$$

provided that the expectations exist. Obviously, if  $\mathbf{X}'$  denotes the transpose of the matrix  $\mathbf{X}$ , then

$$E(\mathbf{X}') = (E(\mathbf{X}))'. \quad (7.30)$$

In Box 7.2 we present some rules of computation for the expectations of random vectors and random matrices (for proofs see Exercise 7-10). In this box we use

$$\mathbf{a}' \mathbf{x} := \sum_{i=1}^n a_i \cdot X_i, \quad (7.31)$$

**Box 7.2 Rules of Computation for Expectations of Random Matrices**

Let  $X = (X_1, \dots, X_n)$  be an  $n$ -variate and  $Y = (Y_1, \dots, Y_m)$  an  $m$ -variate real-valued random variable on a probability space  $(\Omega, \mathcal{A}, P)$  such that the expectations of  $X_i$  and  $Y_j$  are finite, for all  $i = 1, \dots, n, j = 1, \dots, m$ . Furthermore, let  $\mathbf{x} = [X_1, \dots, X_n]'$  and  $\mathbf{y} = [Y_1, \dots, Y_m]'$  denote column vectors,  $\mathbf{a} = [a_1, \dots, a_n]'$  a column vector of real numbers, and let  $\mathbf{A}$  and  $\mathbf{B}$  be matrices of types  $k \times n$  and  $k \times m$ , respectively, each of their components being a real number. Furthermore, let  $\mathbf{C}$  and  $\mathbf{D}$  be matrices of real numbers of types  $l \times n$  and  $r \times m$ , respectively. Then

$$\mathbf{x} \stackrel{p}{=} \mathbf{a} \Rightarrow E(\mathbf{x}) = \mathbf{a}. \quad (\text{i})$$

$$E(\mathbf{a} + \mathbf{x}) = \mathbf{a} + E(\mathbf{x}). \quad (\text{ii})$$

$$E(\mathbf{a}'\mathbf{x}) = \mathbf{a}'E(\mathbf{x}) = E(\mathbf{x})'\mathbf{a} = E(\mathbf{x}'\mathbf{a}). \quad (\text{iii})$$

$$E(\mathbf{A}\mathbf{x}) = \mathbf{A}E(\mathbf{x}). \quad (\text{iv})$$

$$E(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}) = \mathbf{A}E(\mathbf{x}) + \mathbf{B}E(\mathbf{y}). \quad (\text{v})$$

Let  $\mathbf{X}$  be an  $(n \times k)$ -matrix and  $\mathbf{Y}$  an  $(m \times k)$ -matrix of real-valued random variables on  $(\Omega, \mathcal{A}, P)$ , all with finite second moments. Then

$$\mathbf{X} \stackrel{p}{=} \mathbf{A}' \Rightarrow E(\mathbf{X}) = \mathbf{A}'. \quad (\text{vi})$$

$$E(\mathbf{A}' + \mathbf{X}) = \mathbf{A}' + E(\mathbf{X}). \quad (\text{vii})$$

$$E(\mathbf{C}\mathbf{X}) = \mathbf{C}E(\mathbf{X}). \quad (\text{viii})$$

$$E(\mathbf{C}\mathbf{X}\mathbf{Y}'\mathbf{D}') = \mathbf{C}E(\mathbf{X}\mathbf{Y}')\mathbf{D}'. \quad (\text{ix})$$

the *inner product* of an  $n$ -vector  $\mathbf{a} = [a_1, \dots, a_n]'$  of real numbers and the random vector  $\mathbf{x}$ . Correspondingly,

$$\mathbf{A}\mathbf{x} := \begin{bmatrix} \mathbf{a}'_1 \mathbf{x} \\ \vdots \\ \mathbf{a}'_k \mathbf{x} \end{bmatrix}, \quad (7.32)$$

where  $\mathbf{a}'_l$ ,  $l = 1, \dots, k$ , denotes the  $l$ th row of the  $(k \times n)$ -matrix  $\mathbf{A}$  of real numbers.

**7.4.3 Covariance Matrix of two Multivariate Random Variables**

Now we consider two multivariate numerical random variables  $X = (X_1, \dots, X_n)$  and  $Y = (Y_1, \dots, Y_m)$  on a probability space  $(\Omega, \mathcal{A}, P)$ . In particular, we assume that the second moments of all these random variables are finite and focus on their

covariance matrix, again utilizing the representation of  $X$  and  $Y$  as row or column vectors that has been introduced at the beginning of section 7.4.1.

Note that  $[\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]'$  is an  $n \times m$ -matrix of the random variables

$$[X_i - E(X_i)] \cdot [Y_j - E(Y_j)], \quad i = 1, \dots, n, \quad j = 1, \dots, m.$$

Therefore, using (7.29), the *covariance matrix*  $\Sigma_{xy}$  is defined by

$$\Sigma_{xy} := E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]'). \quad (7.33)$$

In other words, the covariance matrix of  $\mathbf{x}$  and  $\mathbf{y}$  is the matrix of covariances, i. e.,

$$\Sigma_{xy} = \begin{bmatrix} \sigma_{X_1 Y_1} & \sigma_{X_1 Y_2} & \cdots & \sigma_{X_1 Y_m} \\ \sigma_{X_2 Y_1} & \sigma_{X_2 Y_2} & \cdots & \sigma_{X_2 Y_m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n Y_1} & \sigma_{X_n Y_2} & \cdots & \sigma_{X_n Y_m} \end{bmatrix}, \quad (7.34)$$

where  $\sigma_{X_i Y_j} := Cov(X_i, Y_j) = E([X_i - E(X_i)] \cdot [Y_j - E(Y_j)])$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ . If we assume that the second moments of the random variables  $X_i$  and  $Y_j$  are finite, then all covariances  $Cov(X_i, Y_j)$  are finite as well, and we say that  $\Sigma_{xy}$  exists.

If we consider a univariate random variable  $Y$ , then  $\mathbf{y} = [Y]$  is also a vector consisting of a single component, the random variable  $Y$ . In this special case,  $\Sigma_{xy}$  is a matrix of type  $n \times 1$ , the column vector

$$\Sigma_{xy} = \begin{bmatrix} \sigma_{X_1 Y} \\ \vdots \\ \sigma_{X_n Y} \end{bmatrix}.$$

Another special case is  $\mathbf{x} = \mathbf{y}$ . The covariance matrix  $\Sigma_{xx}$  of  $\mathbf{x}$  and  $\mathbf{x}$  is called the *variance-covariance matrix of  $\mathbf{x}$*  (or of  $X$ ). Hence,

$$\Sigma_{xx} := E([\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]') \quad (7.35)$$

and

$$\Sigma_{xx} = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2 X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix}. \quad (7.36)$$

The diagonal components of the matrix  $\Sigma_{xx}$  are the variances of the variables  $X_1, \dots, X_n$ , because  $\sigma_{X_i X_i} := Cov(X_i, X_i) = Var(X_i) = \sigma_{X_i}^2$ ,  $i = 1, \dots, n$ .

In Box 7.3 we present some rules of computation for covariance matrices. They are proved in Exercise 7-11.

**Box 7.3 Rules of Computation for Covariance Matrices**

Let  $X = (X_1, \dots, X_n)$  be an  $n$ -variate and  $Y = (Y_1, \dots, Y_m)$  an  $m$ -variate real-valued random variable on a probability space  $(\Omega, \mathcal{A}, P)$  such that the second moments of  $X_i$  and  $Y_j$  are finite, for all  $i = 1, \dots, n, j = 1, \dots, m$ . Furthermore, let  $\mathbf{x} = [X_1, \dots, X_n]'$  and  $\mathbf{y} = [Y_1, \dots, Y_m]'$  denote column vectors,  $\mathbf{a} = [a_1, \dots, a_n]'$  and  $\mathbf{b} = [b_1, \dots, b_m]'$  column vectors of real numbers, and let  $\mathbf{A}$  and  $\mathbf{B}$  be matrices of types  $k \times n$  and  $k \times m$ , respectively, each of their components being a real number. Finally, let  $\mathbf{O}$  denote the  $(n \times m)$ -null matrix. Then:

$$\Sigma_{\mathbf{x}\mathbf{y}} = E(\mathbf{x}\mathbf{y}') - E(\mathbf{x})E(\mathbf{y}'). \quad (\text{i})$$

$$\Sigma_{\mathbf{a}+\mathbf{x}, \mathbf{b}+\mathbf{y}} = \Sigma_{\mathbf{x}\mathbf{y}}. \quad (\text{ii})$$

$$\Sigma_{\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}} = \mathbf{A} \Sigma_{\mathbf{x}\mathbf{y}} \mathbf{B}'. \quad (\text{iii})$$

$$\Sigma_{\mathbf{x}\mathbf{y}} = \Sigma'_{\mathbf{y}\mathbf{x}}. \quad (\text{iv})$$

$$X \perp\!\!\!\perp Y \Rightarrow \Sigma_{\mathbf{x}\mathbf{y}} = \mathbf{O}. \quad (\text{v})$$

$$\mathbf{x} \stackrel{p}{=} \mathbf{a} \Rightarrow \Sigma_{\mathbf{x}\mathbf{y}} = \mathbf{O}. \quad (\text{vi})$$

Additionally, let  $\mathbf{w} = [W_1, \dots, W_r]'$  and  $\mathbf{z} = [Z_1, \dots, Z_s]'$  be real-valued random column vectors on  $(\Omega, \mathcal{A}, P)$  such that all their components have finite second moments. Furthermore, let  $\mathbf{C}$  and  $\mathbf{D}$  be matrices of real numbers of type  $l \times r$  and  $l \times s$ . Then:

$$\Sigma_{\mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{y}, \mathbf{C}\mathbf{w}+\mathbf{D}\mathbf{z}} = \mathbf{A} \Sigma_{\mathbf{x}\mathbf{w}} \mathbf{C}' + \mathbf{A} \Sigma_{\mathbf{x}\mathbf{z}} \mathbf{D}' + \mathbf{B} \Sigma_{\mathbf{y}\mathbf{w}} \mathbf{C}' + \mathbf{B} \Sigma_{\mathbf{y}\mathbf{z}} \mathbf{D}'. \quad (\text{vii})$$

$$n = s \text{ and } \mathbf{x} \stackrel{p}{=} \mathbf{z} \Rightarrow \Sigma_{\mathbf{x}\mathbf{y}} = \Sigma_{\mathbf{z}\mathbf{y}}. \quad (\text{viii})$$

**7.5 Multiple Linear Quasi-Regression**

In the following definition we generalize the concept of a linear quasi-regression considering a multivariate regressor  $X = (X_1, \dots, X_n)$ . We use the notation  $\mathbf{x} = [X_1, \dots, X_n]'$  to denote the column vector of  $X$ ,  $\boldsymbol{\beta}' = [\beta_1, \dots, \beta_n]$  for the row vector of the real numbers  $\beta_1, \dots, \beta_n$ , and  $\mathbf{b}' = [b_1, \dots, b_n]$  for the row vector of the real numbers  $b_1, \dots, b_n$ .

**Definition 7.28 (Multiple Linear Quasi-Regression)**

Let  $X_i: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ ,  $i = 1, \dots, n$ , and  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be real-valued random variables, define  $X := (X_1, \dots, X_n)$ , assume  $E(X_i^2), E(Y^2) < \infty$ ,

$i = 1, \dots, n$ , and that the inverse  $\Sigma_{xx}^{-1}$  exists. Define the function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  by

$$f(x) = \beta_0 + \sum_{i=1}^n \beta_i x_i, \quad \forall x = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad (7.37)$$

where  $\beta_0, \boldsymbol{\beta} = [\beta_1, \dots, \beta_n]'$  minimize the function  $MSE: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  with

$$MSE(b_0, b_1, \dots, b_n) = E([Y - (b_0 + \mathbf{b}'\mathbf{x})]^2), \quad \forall (b_0, b_1, \dots, b_n) \in \mathbb{R}^{n+1}. \quad (7.38)$$

Then  $f$  is called the *linear quasi-regression of  $Y$  on  $X_1, \dots, X_n$* . The composition of  $X$  and  $f$  is denoted by  $Q_{lin}(Y|X)$  or  $Q_{lin}(Y|X_1, \dots, X_n)$ , i. e.,

$$Q_{lin}(Y|X) := f(X) = \beta_0 + \boldsymbol{\beta}'\mathbf{x} = \beta_0 + \sum_{i=1}^n \beta_i X_i. \quad (7.39)$$

**Remark 7.29 (Coefficient of Determination)** Let the assumptions of Definition 7.28 hold and assume  $Var(Y) > 0$ . Then

$$Q_{Y|X}^2 := \frac{Var[Q_{lin}(Y|X)]}{Var(Y)} \quad (7.40)$$

is called the *coefficient of determination* of the linear quasi-regression  $Q_{lin}(Y|X)$ .  $\triangleleft$

In the following theorem, we generalize Theorem 7.14 considering a multivariate real-valued regressor  $X = (X_1, \dots, X_n)$ . In this theorem  $\Sigma_{x\epsilon}$  denotes the covariance vector of  $\mathbf{x}$  and  $\epsilon$ , which is defined by

$$\epsilon := Y - Q_{lin}(Y|X_1, \dots, X_n) \quad (7.41)$$

and called the *residual of  $Y$  with respect to its linear quasi-regression on  $X$* .

**Theorem 7.30 (Characterizations of the Multiple Linear Quasi-Regression)**

Let  $X_1, \dots, X_n, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be real-valued random variables, assume  $E(X_i^2), E(Y^2) < \infty$  for all  $i = 1, \dots, n$ , and that the inverse  $\Sigma_{xx}^{-1}$  exists. Furthermore, let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ , let

$$f(X_1, \dots, X_n) := \beta_0 + \boldsymbol{\beta}'\mathbf{x}, \quad \beta_0 \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n, \quad (7.42)$$

be the composition of  $(X_1, \dots, X_n)$  and  $f$ , and define  $\epsilon := Y - f(X_1, \dots, X_n)$ . Then the following three propositions are equivalent to each other:

- (i)  $E(\epsilon) = 0$  and  $\Sigma_{x\epsilon} = \mathbf{0}$ .

$$(ii) \beta_0 = E(Y) - \beta' E(\mathbf{x}) \quad \text{and} \quad \beta = \Sigma_{xx}^{-1} \Sigma_{xy}.$$

$$(iii) f(X_1, \dots, X_n) = Q_{lin}(Y|X_1, \dots, X_n).$$

(Proof p. 235)

**Remark 7.31 (Uniqueness)** Suppose that the assumptions of Theorem 7.30 hold and  $f(X_1, \dots, X_n) = Q_{lin}(Y|X)$ . Then proposition (ii) of this theorem implies that the coefficients  $\beta_0, \beta_1, \dots, \beta_n$  are uniquely defined. Because  $f(X_1, \dots, X_n) = \beta_0 + \beta' \mathbf{x} = Q_{lin}(Y|X)$ , the linear quasi-regression  $f$  and  $Q_{lin}(Y|X)$  are uniquely defined as well.  $\triangleleft$

**Corollary 7.32 (No Correlation Between Linear Quasi-Regression and its Residual)**

Under the assumptions of Definition 7.28,

$$\text{Cov}[Q_{lin}(Y|X), \epsilon] = 0. \quad (7.43)$$

and

$$\text{Var}[Q_{lin}(Y|X)] = \text{Cov}[Y, Q_{lin}(Y|X)]. \quad (7.44)$$

(Proof p. 237)

**Remark 7.33 (Correlation Between Linear Quasi-Regression and its Regressand)**

Under the assumptions of Definition 7.28 and  $\text{Var}(Y) > 0$ ,

$$Q_{Y|X}^2 = \text{Corr}[Y, Q_{lin}(Y|X)]^2 \quad (7.45)$$

(see Exercise 7-12).  $\triangleleft$

## 7.6 Proofs

### Proof of Theorem 7.12

Suppose  $\text{Var}(Y) = 0$ . Then rules (iv) of Box 6.2 and (vii) of Box 7.1 imply that  $\text{Cov}(X, Y) = 0$ . This shows that the Inequality (7.16) holds if  $\text{Var}(Y) = 0$ . Now suppose  $\text{Var}(Y) > 0$ . Then

$$\begin{aligned} 0 &\leq \text{Var}\left(X - \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \cdot Y\right) \cdot \text{Var}(Y) \\ &= \left(\text{Var}(X) - 2 \cdot \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \text{Cov}(X, Y) + \frac{\text{Cov}(X, Y)^2}{\text{Var}(Y)^2} \text{Var}(Y)\right) \cdot \text{Var}(Y) \quad [(7.12)] \\ &= \text{Var}(X) \cdot \text{Var}(Y) - \text{Cov}(X, Y)^2, \end{aligned}$$

which is equivalent to (7.16). According to the first part of the proof,

$$\text{Cov}(X, Y)^2 = \text{Var}(X) \cdot \text{Var}(Y) \Leftrightarrow \text{Var}(Y) = 0 \quad \text{or} \quad \text{Var}\left(X - \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \cdot Y\right) = 0.$$

Rule (iv) of Box 6.2 implies that this is equivalent to

$$\exists a \in \mathbb{R} : Y \stackrel{P}{=} a \quad \text{or} \quad \exists c \in \mathbb{R} : X - \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \cdot Y \stackrel{P}{=} c.$$

If  $\text{Cov}(X, Y) \neq 0$ , this is equivalent to

$$\exists a \in \mathbb{R} : Y \stackrel{P}{=} a + 0 \cdot X \quad \text{or} \quad \exists c \in \mathbb{R} : Y \stackrel{P}{=} \left( -\frac{c \cdot \text{Var}(Y)}{\text{Cov}(X, Y)} \right) + \left( \frac{\text{Var}(Y)}{\text{Cov}(X, Y)} \right) \cdot X.$$

Obviously, in both cases, there is a linear function of  $X$  that is  $P$ -equivalent to  $Y$ . Furthermore, if there are  $a, b \in \mathbb{R}$  with  $Y \stackrel{P}{=} a + bX$  — implying  $\text{Var}(Y) = b^2 \text{Var}(X)$  [see Box 6.2 (ii), (iii)] — then

$$\begin{aligned} \text{Cov}(X, Y)^2 &= \text{Cov}(X, a + bX)^2 && \text{[Box 7.1 (x)]} \\ &= b^2 \text{Cov}(X, X)^2 && \text{[Box 7.1 (ii), (iii)]} \\ &= b^2 \text{Var}(X)^2 && \text{[Box 7.1 (iv)]} \\ &= \text{Var}(X) \cdot \text{Var}(Y). && \text{[Box 6.2 (ii), (iii)]} \end{aligned}$$

### ***Proof of Theorem 7.14***

The proof is organized as follows: (iii)  $\Rightarrow$  (ii)  $\Rightarrow$  (i)  $\Rightarrow$  (iii), which will prove that (i), (ii), and (iii) are equivalent.

(iii)  $\Rightarrow$  (ii). The first partial derivatives of

$$\begin{aligned} &MSE(a_0, a_1) \\ &= E\{[Y - (a_0 + a_1 X)]^2\} \\ &= E(Y^2) + E[(a_0 + a_1 X)^2] - 2E[Y \cdot (a_0 + a_1 X)] && \text{[Box 6.1 (vi)]} \\ &= E(Y^2) + a_0^2 + a_1^2 E(X^2) + 2a_0 a_1 E(X) - 2a_0 E(Y) - 2a_1 E(X \cdot Y) && \text{[Box 6.1 (iii), (vi)]} \end{aligned}$$

with respect to  $a_0$  and  $a_1$  are

$$\frac{\partial MSE(a_0, a_1)}{\partial a_0} = 2a_0 + 2a_1 E(X) - 2E(Y)$$

and

$$\frac{\partial MSE(a_0, a_1)}{\partial a_1} = 2a_1 E(X^2) + 2a_0 E(X) - 2E(X \cdot Y).$$

If  $f(X) = \alpha_0 + \alpha_1 X = Q_{lin}(Y|X)$ , then

$$2\alpha_0 + 2\alpha_1 E(X) - 2E(Y) = 0$$

and

$$2\alpha_1 E(X^2) + 2\alpha_0 E(X) - 2E(X \cdot Y) = 0.$$

Solving the first equation for  $\alpha_0$  yields

$$\alpha_0 = E(Y) - \alpha_1 E(X).$$

Inserting this result into the second equation yields

$$\alpha_1 E(X^2) + E(Y) \cdot E(X) - \alpha_1 E(X)^2 - E(X \cdot Y) = 0.$$

Using  $\text{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$  [see Box 7.1 (i)] and  $\text{Var}(X) = E(X^2) - E(X)^2$  [see Box 6.2 (i)] we receive

$$\alpha_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}.$$

(ii)  $\Rightarrow$  (i).

$$\begin{aligned} E(\epsilon) &= E[Y - f(X)] \\ &= E[Y - (\alpha_0 + \alpha_1 X)] && \text{[def. of } f(X)\text{]} \\ &= E\{Y - [E(Y) - \alpha_1 E(X) + \alpha_1 X]\} && \text{[(ii)]} \\ &= E(Y) - E(Y) + \alpha_1 E(X) - \alpha_1 \cdot E(X) && \text{[Box 6.1 (vii)]} \\ &= 0. \\ \text{Cov}(X, \epsilon) &= \text{Cov}(X, [Y - (\alpha_0 + \alpha_1 X)]) && \text{[def. of } \epsilon\text{]} \\ &= \text{Cov}(X, Y) - \alpha_1 \text{Var}(X) && \text{[Box 7.1 (ii), (iii)]} \\ &= \text{Cov}(X, Y) - \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \cdot \text{Var}(X) && \text{[(ii)]} \\ &= 0. \end{aligned}$$

(i)  $\Rightarrow$  (iii). Let  $f(X) := \alpha_0 + \alpha_1 X$ ,  $\alpha_0, \alpha_1 \in \mathbb{R}$ , be a linear function of  $X$  with  $E(\epsilon) = 0$  and  $\text{Cov}(X, \epsilon) = 0$ , where  $\epsilon := Y - f(X)$ . Then, for any linear function  $h(X) := a_0 + a_1 X$ ,  $a_0, a_1 \in \mathbb{R}$ ,

$$\begin{aligned} E((Y - f(X))[f(X) - h(X)]) &= E(\epsilon \cdot [f(X) - h(X)]) && \text{[def. of } \epsilon\text{]} \\ &= E(\epsilon \cdot [(\alpha_0 + \alpha_1 X) - (a_0 + a_1 X)]) && \text{[defs. of } f(X), h(X)\text{]} \\ &= E(\epsilon \cdot [(\alpha_0 - a_0) + (\alpha_1 - a_1)X]) \\ &= (\alpha_0 - a_0) \cdot E(\epsilon) + (\alpha_1 - a_1)E(\epsilon \cdot X) && \text{[Box 6.1 (vii)]} \\ &= 0. && \text{[(i), Box 7.1 (i)]} \end{aligned}$$

Using this result and considering

$$\begin{aligned} E([Y - h(X)]^2) &= E\{([Y - f(X)] + [f(X) - h(X)])^2\} \\ &= E([Y - f(X)]^2) + E([f(X) - h(X)]^2) + 2 \cdot E([Y - f(X)][f(X) - h(X)]) \\ &= E([Y - f(X)]^2) + E([f(X) - h(X)]^2) \\ &\geq E([Y - f(X)]^2). \end{aligned}$$

In this inequality '=' holds if and only if  $f(X) \stackrel{p}{=} h(X)$ . Because  $f(X) = \alpha_0 + \alpha_1 X$  and  $h(X) = a_0 + a_1 X$ , the property  $f(X) \stackrel{p}{=} h(X)$  is equivalent to  $f(X) = h(X)$ .

### **Proof of Theorem 7.30**

The proof is organized as follows: (iii)  $\Rightarrow$  (ii)  $\Rightarrow$  (i)  $\Rightarrow$  (iii), which will prove that (i), (ii), and (iii) are equivalent.

(iii)  $\Rightarrow$  (ii). The first partial derivative of

$$\begin{aligned}
& MSE(b_0, b_1, \dots, b_n) \\
&= E\{[Y - (b_0 + \mathbf{b}'\mathbf{x})]^2\} \quad [(7.38)] \\
&= E(Y^2) + E((b_0 + \mathbf{b}'\mathbf{x})^2) - 2E(Y \cdot (b_0 + \mathbf{b}'\mathbf{x})) \quad [\text{Box 6.1 (vi), (iii)}] \\
&= E(Y^2) + b_0^2 + E((\mathbf{b}'\mathbf{x})^2) + 2b_0 \cdot E(\mathbf{b}'\mathbf{x}) - 2b_0 E(Y) - 2E(\mathbf{b}'\mathbf{x} \cdot Y) \quad [\text{Box 6.1 (vi), (i), (iii)}] \\
&= E(Y^2) + b_0^2 + E((\mathbf{b}'\mathbf{x})^2) + 2b_0 \cdot \mathbf{b}'E(\mathbf{x}) - 2b_0 E(Y) - 2\mathbf{b}'E(\mathbf{x} \cdot Y) \quad [\text{Box 7.2 (iii)}]
\end{aligned}$$

with respect to  $b_0$  is

$$\frac{\partial MSE(b_0, b_1, \dots, b_n)}{\partial b_0} = 2b_0 + 2\mathbf{b}'E(\mathbf{x}) - 2E(Y).$$

If  $f(X_1, \dots, X_n) = \beta_0 + \boldsymbol{\beta}'\mathbf{x} = Q_{lin}(Y|X_1, \dots, X_n)$ , then

$$2\beta_0 + 2\boldsymbol{\beta}'E(\mathbf{x}) - 2E(Y) = 0.$$

Dividing both sides by 2, and solving for  $\beta_0$  yields

$$\beta_0 = E(Y) - \boldsymbol{\beta}'E(\mathbf{x}).$$

Gathering the first partial derivatives of  $MSE(b_0, b_1, \dots, b_n)$  with respect to  $b_1, \dots, b_n$  in a vector yields

$$\left[ \frac{\partial MSE(b_0, b_1, \dots, b_n)}{\partial b_1}, \dots, \frac{\partial MSE(b_0, b_1, \dots, b_n)}{\partial b_n} \right]' = 2E(\mathbf{x}\mathbf{x}')\boldsymbol{\beta} + 2b_0E(\mathbf{x}) - 2E(\mathbf{x} \cdot Y).$$

If  $f(X_1, \dots, X_n) = \beta_0 + \boldsymbol{\beta}'\mathbf{x} = Q_{lin}(Y|X_1, \dots, X_n)$ , then  $2E(\mathbf{x}\mathbf{x}')\boldsymbol{\beta} + 2\beta_0E(\mathbf{x}) - 2E(\mathbf{x} \cdot Y) = \mathbf{0}$ , and dividing both sides by 2 yields

$$E(\mathbf{x}\mathbf{x}')\boldsymbol{\beta} + \beta_0E(\mathbf{x}) - E(\mathbf{x} \cdot Y) = \mathbf{0}.$$

Inserting our result  $\beta_0 = E(Y) - \boldsymbol{\beta}'E(\mathbf{x})$ , using  $\boldsymbol{\Sigma}_{xy} = E(\mathbf{x} \cdot Y) - E(\mathbf{x}) \cdot E(Y)$  and  $\boldsymbol{\Sigma}_{xx} = E(\mathbf{x}\mathbf{x}') - E(\mathbf{x})E(\mathbf{x}')$  [see Box 7.3 (i)] yields

$$\begin{aligned}
& E(\mathbf{x}\mathbf{x}')\boldsymbol{\beta} + (E(Y) - \boldsymbol{\beta}'E(\mathbf{x}))E(\mathbf{x}) - E(\mathbf{x} \cdot Y) \\
&= E(\mathbf{x}\mathbf{x}')\boldsymbol{\beta} + E(Y) \cdot E(\mathbf{x}) - E(\mathbf{x})E(\mathbf{x}')\boldsymbol{\beta} - E(\mathbf{x} \cdot Y) \\
&= (E(\mathbf{x}\mathbf{x}') - E(\mathbf{x})E(\mathbf{x}'))\boldsymbol{\beta} - (E(\mathbf{x} \cdot Y) - E(\mathbf{x}) \cdot E(Y)) \\
&= \boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} - \boldsymbol{\Sigma}_{xy} = \mathbf{0},
\end{aligned}$$

which is a necessary condition for a minimum. (Note, in this proof we do not have to check a sufficient condition.) Adding  $\boldsymbol{\Sigma}_{xy}$  on both sides yields  $\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xy}$ , and multiplying both sides from the left by  $\boldsymbol{\Sigma}_{xx}^{-1}$  we receive

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}.$$

(ii)  $\Rightarrow$  (i).

$$\begin{aligned}
E(\epsilon) &= E(Y - f(X)) \quad [\text{def. of } \epsilon] \\
&= E(Y - (\beta_0 + \boldsymbol{\beta}'\mathbf{x})) \quad [\text{def. of } f(X)] \\
&= E(Y - (E(Y) - \boldsymbol{\beta}'E(\mathbf{x}) + \boldsymbol{\beta}'\mathbf{x})) \quad [(ii)] \\
&= E(Y) - E(Y) + \boldsymbol{\beta}'E(\mathbf{x}) - \boldsymbol{\beta}'E(\mathbf{x}) \quad [\text{Box 6.1 (vi), (i), Box 7.2 (iii)}] \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
\Sigma_{x\epsilon} &= \Sigma_{x, Y - (\beta_0 + \beta'x)} && \text{[def. of } \epsilon \text{]} \\
&= \Sigma_{xy} - \Sigma_{xx}\beta && \text{[Box 7.3 (ii), (vii)]} \\
&= \Sigma_{xy} - \Sigma_{xx}\Sigma_{xx}^{-1}\Sigma_{xy} && \text{[(ii)]} \\
&= \mathbf{0}.
\end{aligned}$$

(i)  $\Rightarrow$  (iii). Let  $f(X) := \beta_0 + \beta'x$ ,  $\beta_0 \in \mathbb{R}$ ,  $\beta \in \mathbb{R}^n$  such that  $E(\epsilon) = 0$  and  $\Sigma_{x\epsilon} = \mathbf{0}$ , where  $\epsilon := Y - f(X)$ . Then, for any linear function  $h(X) := b_0 + \mathbf{b}'x$ ,  $b_0 \in \mathbb{R}$ ,  $\mathbf{b} \in \mathbb{R}^n$ ,

$$\begin{aligned}
&E((Y - f(X))(f(X) - h(X))) \\
&= E(\epsilon \cdot [f(X) - h(X)]) && \text{[def. of } \epsilon \text{]} \\
&= E(\epsilon \cdot [(\beta_0 + \beta'x) - (b_0 + \mathbf{b}'x)]) && \text{[defs. of } f(X), h(X) \text{]} \\
&= E(\epsilon \cdot [(\beta_0 - b_0) + (\beta - \mathbf{b})'x]) \\
&= (\beta_0 - b_0) \cdot E(\epsilon) + (\beta - \mathbf{b})'E(\epsilon \cdot x) && \text{[Box 6.1 (vi), Box 7.2 (iii)]} \\
&= (\beta - \mathbf{b})' \Sigma_{x\epsilon} && \text{[(i), Box 7.3 (i)]} \\
&= \mathbf{0}. && \text{[(i)]}
\end{aligned}$$

Using this result and considering

$$\begin{aligned}
E([Y - h(X)]^2) &= E([(Y - f(X)) + (f(X) - h(X))]^2) \\
&= E([Y - f(X)]^2) + E([f(X) - h(X)]^2) + 2 \cdot E([Y - f(X)][f(X) - h(X)]) \\
&= E([Y - f(X)]^2) + E([f(X) - h(X)]^2) \\
&\geq E([Y - f(X)]^2).
\end{aligned}$$

Hence,  $f(X) = Q_{lin}(Y|X)$ .

### **Proof of Corollary 7.32**

$$\begin{aligned}
Cov [Q_{lin}(Y|X), \epsilon] &= Cov(\beta_0 + \beta'x, \epsilon) && \text{[(7.39)]} \\
&= \beta' Cov(x, \epsilon) && \text{[Box 7.3 (ii), (iii)]} \\
&= \mathbf{0}. && \text{[Cov}(x, \epsilon) = \Sigma_{x\epsilon} = \mathbf{0}, \text{ Th. 7.30]}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
Var [Q_{lin}(Y|X)] &= Cov [Q_{lin}(Y|X), Q_{lin}(Y|X)] && \text{[Box 7.1 (iv)]} \\
&= Cov [Q_{lin}(Y|X), \epsilon] + Cov [Q_{lin}(Y|X), Q_{lin}(Y|X)] && \text{[(7.43)]} \\
&= Cov [Q_{lin}(Y|X) + \epsilon, Q_{lin}(Y|X)] && \text{[Box 7.1 (ix)]} \\
&= Cov [Y, Q_{lin}(Y|X)]. && \text{[(7.41)]}
\end{aligned}$$

## **7.7 Exercises**

$\triangleright$  **Exercise 7-1** Use Definition 7.2 in order to determine the coefficients  $\alpha_0$  and  $\alpha_1$  of the linear quasi-regression in Example 7.5.

▷ **Exercise 7-2** Consider the linear quasi-regression  $f$  with  $f(x) = \alpha_0 + \alpha_1 x$ ,  $x \in \mathbb{R}$ . Prove: If  $x_1, x_2 \in \mathbb{R}$  with  $x_1 \neq x_2$ , then  $\alpha_0 = f(0)$  and  $\alpha_1 = \frac{1}{x_2 - x_1} [f(x_2) - f(x_1)]$ .

▷ **Exercise 7-3** Prove the propositions of Box 7.1.

▷ **Exercise 7-4** Prove the proposition of Remark 7.13.

▷ **Exercise 7-5** Prove the proposition of Remark 7.20.

▷ **Exercise 7-6** Prove the proposition of Remark 7.21.

▷ **Exercise 7-7** Show that  $\text{Corr}(a_0 + a_1 X, b_0 + b_1 Y) = \text{Corr}(X, Y)$ , where  $a_0, a_1, b_0, b_1 \in \mathbb{R}$ .

▷ **Exercise 7-8** Show

$$\text{Cov}(X, Y) = \begin{cases} SD(X) \cdot SD(Y), & \text{if } a_1 > 0 \\ -SD(X) \cdot SD(Y), & \text{if } a_1 < 0, \end{cases} \quad (7.46)$$

provided that there are  $a_0, a_1 \in \mathbb{R}$  with  $Y \stackrel{p}{=} a_0 + a_1 X$ ,  $a_1 \neq 0$ .

▷ **Exercise 7-9** Prove Equation (7.24) and that the slope is invariant under translations  $c + X$ ,  $d + Y$ ,  $c, d \in \mathbb{R}$ .

▷ **Exercise 7-10** Prove the rules of computation of Box 7.2.

▷ **Exercise 7-11** Prove the rules of computation of Box 7.3.

▷ **Exercise 7-12**

Under the assumptions of Definition 7.28 and  $\text{Var}(Y) > 0$ , prove Equation 7.45

## Solutions

▷ **Solution 7-1**

$$MSE(a_0, a_1) = E([Y - (a_0 + a_1 X)]^2) \quad [(7.2)]$$

$$= \frac{1}{4} \cdot (1 - a_0 - a_1)^2 + \frac{1}{2} \cdot (2 - a_0 - 2a_1)^2 + \frac{1}{4} \cdot (1 - a_0 - 3a_1)^2 \quad [(6.3)]$$

$$= \frac{1}{4} \cdot (1 + a_0^2 + a_1^2 - 2a_0 - 2a_1 + 2a_0 a_1)$$

$$+ \frac{1}{2} \cdot (4 + a_0^2 + 4a_1^2 - 4a_0 - 8a_1 + 4a_0 a_1)$$

$$+ \frac{1}{4} \cdot (1 + a_0^2 + 9a_1^2 - 2a_0 - 6a_1 + 6a_0 a_1)$$

$$= \frac{5}{2} + a_0^2 + \frac{9}{2} \cdot a_1^2 - 3a_0 - 6a_1 + 4a_0 a_1.$$

The partial derivatives are

$$\frac{\partial MSE(a_0, a_1)}{\partial a_0} = 2a_0 - 3 + 4a_1 \quad \text{and} \quad \frac{\partial MSE(a_0, a_1)}{\partial a_1} = 9a_1 - 6 + 4a_0.$$

Fixing the partial derivatives to 0 and denoting the solutions by  $\alpha_0$  and  $\alpha_1$  yields

$$2\alpha_0 - 3 + 4\alpha_1 = 0 \quad \text{and} \quad 9\alpha_1 - 6 + 4\alpha_0 = 0.$$

The first equation implies  $\alpha_0 = -2\alpha_1 + \frac{3}{2}$ . Inserting this result into the second equation yields  $9\alpha_1 - 6 - 8\alpha_1 + 6 = 0$ , which implies  $\alpha_1 = 0$  and  $\alpha_0 = \frac{3}{2}$ . The values  $\alpha_1 = 0$  and  $\alpha_0 = \frac{3}{2}$  satisfy a necessary condition for a minimum; now we check if they also satisfy a sufficient condition. The second partial derivatives are

$$\frac{\partial^2 \text{MSE}(a_0, a_1)}{\partial a_0^2} = 2 > 0, \quad \frac{\partial^2 \text{MSE}(a_0, a_1)}{\partial a_1^2} = 9 > 0, \quad \text{and} \quad \frac{\partial^2 \text{MSE}(a_0, a_1)}{\partial a_0 \partial a_1} = 4,$$

i.e., in this case they are constant for all  $a_0, a_1 \in \mathbb{R}$ . Because  $2 \cdot 9 - 4^2 > 0$ , we can conclude that  $\text{MSE}(a_0, a_1)$  has its minimum for  $a_0 = \alpha_0 = \frac{3}{2}$  and  $a_1 = \alpha_1 = 0$  (see Ellis & Gulick, 2006, Th. 13.21, p. 877).

▷ **Solution 7-2** The equation  $f(x) = \alpha_0 + \alpha_1 x$ ,  $x \in \mathbb{R}$ , yields  $f(0) = \alpha_0$ ,

$$f(x_1) = \alpha_0 + \alpha_1 x_1, \quad \text{and} \quad f(x_2) = \alpha_0 + \alpha_1 x_2.$$

Hence,

$$f(x_2) - f(x_1) = \alpha_0 + \alpha_1 x_2 - (\alpha_0 + \alpha_1 x_1) = \alpha_1 (x_2 - x_1).$$

Multiplying both sides by  $x_2 - x_1$  yields  $\alpha_1 = \frac{1}{x_2 - x_1} [f(x_2) - f(x_1)]$ , provided that  $x_1 \neq x_2$ .

▷ **Solution 7-3** (i)

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X)) \cdot (Y - E(Y))] && [(7.8)] \\ &= E[X \cdot Y - X \cdot E(Y) - E(X) \cdot Y + E(X) \cdot E(Y)] \\ &= E(X \cdot Y) - E[X \cdot E(Y)] - E[E(X) \cdot Y] + E(X) \cdot E(Y) && [\text{Box 6.1 (vi), (i)}] \\ &= E(X \cdot Y) - 2E(X) \cdot E(Y) + E(X) \cdot E(Y) && [\text{Box 6.1 (iii)}] \\ &= E(X \cdot Y) - E(X) \cdot E(Y). \end{aligned}$$

(ii)

$$\begin{aligned} \text{Cov}(\alpha X, \beta Y) &= E[(\alpha X - E(\alpha X)) \cdot (\beta Y - E(\beta Y))] && [(7.8)] \\ &= E[(\alpha X - \alpha E(X)) \cdot (\beta Y - \beta E(Y))] && [\text{Box 6.1 (ii)}] \\ &= E[(X - E(X)) \cdot (Y - E(Y))] \\ &= \text{Cov}(X, Y). && [(7.8)] \end{aligned}$$

(iii)

$$\begin{aligned} \text{Cov}(\alpha X, \beta Y) &= E(\alpha X \cdot \beta Y) - E(\alpha X) \cdot E(\beta Y) && [\text{Box 7.1 (i)}] \\ &= \alpha \beta E(X \cdot Y) - \alpha \beta E(X) \cdot E(Y) && [\text{Box 6.1 (iii)}] \\ &= \alpha \beta [E(X \cdot Y) - E(X) \cdot E(Y)] \\ &= \alpha \beta \text{Cov}(X, Y). && [\text{Box 7.1 (i)}] \end{aligned}$$

(iv) This rule immediately follows from Equations (6.27) and (7.8).

(v) This rule immediately follows from Equation (7.8).

(vi) Independence of  $X$  and  $Y$  implies  $E(X \cdot Y) = E(X) \cdot E(Y)$  (see Th. 6.24) and  $\text{Cov}(X, Y) = 0$  [see Rule (i)].

(vii) According to Lemma 5.51,  $X$  and  $Y$  are independent if  $X \stackrel{P}{=} \alpha$ . This implies that  $\text{Cov}(X, Y) = 0$  [see Rule (vi) of Box 7.1].

(viii)

$$\begin{aligned}
& \text{Var}\left(\sum_{i=1}^n \alpha_i Y_i\right) \\
&= E\left[\left(\sum_{i=1}^n \alpha_i Y_i - E\left(\sum_{i=1}^n \alpha_i Y_i\right)\right)^2\right] \quad [(6.27)] \\
&= E\left[\left(\sum_{i=1}^n \alpha_i Y_i - \sum_{i=1}^n \alpha_i E(Y_i)\right)^2\right] \quad [\text{Box 6.1 (vii)}] \\
&= E\left[\left(\sum_{i=1}^n \alpha_i [Y_i - E(Y_i)]\right)^2\right] \\
&= E\left[\sum_{i=1}^n \alpha_i^2 (Y_i - E(Y_i))^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \alpha_i \alpha_j (Y_i - E(Y_i))(Y_j - E(Y_j))\right] \\
&= \sum_{i=1}^n \alpha_i^2 E(Y_i - E(Y_i))^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \alpha_i \alpha_j E[(Y_i - E(Y_i))(Y_j - E(Y_j))] \quad [\text{Box 6.1 (vii)}] \\
&= \sum_{i=1}^n \alpha_i^2 \text{Var}(Y_i) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \alpha_i \alpha_j \text{Cov}(Y_i, Y_j). \quad [(6.27), (7.8)]
\end{aligned}$$

(ix)

$$\begin{aligned}
& \text{Cov}\left(\sum_{i=1}^n \alpha_i X_i, \sum_{j=1}^m \beta_j Y_j\right) \\
&= E\left[\left(\sum_{i=1}^n \alpha_i X_i - E\left(\sum_{i=1}^n \alpha_i X_i\right)\right)\left(\sum_{j=1}^m \beta_j Y_j - E\left(\sum_{j=1}^m \beta_j Y_j\right)\right)\right] \quad [(7.8)] \\
&= E\left[\left(\sum_{i=1}^n \alpha_i [X_i - E(X_i)]\right)\left(\sum_{j=1}^m \beta_j [Y_j - E(Y_j)]\right)\right] \quad [\text{Box 6.1 (vii)}] \\
&= E\left[\sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j [X_i - E(X_i)] [Y_j - E(Y_j)]\right] \\
&= \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j \text{Cov}(X_i, Y_j). \quad [\text{Box 6.1 (vii), (7.8)}]
\end{aligned}$$

(x) If  $X_1 \stackrel{p}{=} X_2$ , then  $[Y - E(Y)] \cdot [X_1 - E(X_1)] \stackrel{p}{=} [Y - E(Y)] \cdot [X_2 - E(X_2)]$ . According to Corollary 5.24, these two product variables have the same distribution and according to Corollary 6.17 the same expectation. However, the expectations of these product variables are the covariances. Hence,  $\text{Cov}(Y, X_1) = \text{Cov}(Y, X_2)$ .

▷ **Solution 7-4** If  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  are random variables, then

$$(\alpha X + \beta Y)^2 = \alpha^2 X^2 + \beta^2 Y^2 + 2\alpha\beta X \cdot Y$$

is also a random variable on  $(\Omega, \mathcal{A}, P)$  (see Def. 5.1, Example 2.61, and Th. 2.57), and this implies

$$\begin{aligned}
E[(\alpha X + \beta Y)^2] &= E(\alpha^2 X^2 + \beta^2 Y^2 + 2\alpha\beta X \cdot Y) \\
&= \alpha^2 E(X^2) + \beta^2 E(Y^2) + 2\alpha\beta E(X \cdot Y). \quad [\text{Box 6.1 (vii)}]
\end{aligned}$$

The terms  $E(X^2)$  and  $E(Y^2)$  are finite by assumption and, according to Remark 7.1,  $E(X \cdot Y)$  is finite as well. This implies  $E[(\alpha X + \beta Y)^2] < \infty$ .

▷ **Solution 7-5** Inequality (7.16) implies

$$-\sqrt{\text{Var}(X) \cdot \text{Var}(Y)} \leq \text{Cov}(X, Y) \leq \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}.$$

If  $\sqrt{\text{Var}(X) \cdot \text{Var}(Y)} = SD(X) \cdot SD(Y) > 0$ , then these inequalities yield  $-1 \leq \text{Corr}(X, Y) \leq 1$ .

If  $\sqrt{\text{Var}(X) \cdot \text{Var}(Y)} = SD(X) \cdot SD(Y) = 0$ , then, by Definition 7.18,  $\text{Corr}(X, Y) = 0$ .

▷ **Solution 7-6** If  $SD(X), SD(Y) > 0$ , then

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{SD(X) \cdot SD(Y)} \quad [(7.18)]$$

$$= \frac{E[(X - E(X)) \cdot (Y - E(Y))]}{SD(X) \cdot SD(Y)} \quad [(7.8)]$$

$$= E\left(\frac{[X - E(X)] \cdot [Y - E(Y)]}{SD(X) \cdot SD(Y)}\right) \quad [\text{Box 6.1 (iii)}]$$

$$= E\left(\frac{X - E(X)}{SD(X)} \cdot \frac{Y - E(Y)}{SD(Y)}\right).$$

▷ **Solution 7-7** According to Rules (ii) and (iii) of Box 7.1,  $\text{Cov}(a_0 + a_1 X, b_0 + b_1 Y) = a_1 b_1 \text{Cov}(X, Y)$ . Similarly, according to Rules (ii) and (iii) of Box 6.2,  $\text{Var}(a_0 + a_1 X) = a_1^2 \text{Var}(X)$ , which implies

$$SD(a_0 + a_1 X) = \begin{cases} a_1 SD(X), & \text{if } a_1 > 0 \\ -a_1 SD(X), & \text{if } a_1 < 0 \\ 0, & \text{if } a_1 = 0. \end{cases}$$

Hence, for  $a_1, b_1 > 0$ ,

$$\begin{aligned} \text{Corr}(a_0 + a_1 X, b_0 + b_1 Y) &= \frac{\text{Cov}(a_0 + a_1 X, b_0 + b_1 Y)}{SD(a_0 + a_1 X) \cdot SD(b_0 + b_1 Y)} \\ &= \frac{a_1 b_1 \text{Cov}(X, Y)}{a_1 SD(X) \cdot b_1 SD(Y)} \\ &= \frac{\text{Cov}(X, Y)}{SD(X) \cdot SD(Y)} \\ &= \text{Corr}(X, Y). \end{aligned}$$

The proofs for the cases of  $a_1, b_1 \neq 0$  are analogous. Note that we defined  $\text{Corr}(X, Y) = 0$  if  $SD(X) = 0$  or  $SD(Y) = 0$ .

▷ **Solution 7-8** According to Box 7.1 (x),  $\text{Cov}(X, Y) = \text{Cov}(X, \alpha_0 + \alpha_1 X)$  if  $Y \stackrel{p}{=} \alpha_0 + \alpha_1 X$ . Box 6.2 (v), (ii), (iii) imply  $\text{Var}(Y) = \alpha_1^2 \text{Var}(X)$ . Therefore,

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, \alpha_0 + \alpha_1 X)}{SD(X) \cdot \sqrt{\alpha_1^2 \text{Var}(X)}} = \frac{\alpha_1 \text{Var}(X)}{SD(X) \cdot |\alpha_1| \cdot SD(X)} = \frac{\alpha_1}{|\alpha_1|}.$$

Hence,  $\text{Corr}(X, Y) = 1$  if  $\alpha_1 > 0$  and  $\text{Corr}(X, Y) = -1$  if  $\alpha_1 < 0$ . Therefore,

$$\text{Corr}(X, Y) := \frac{\text{Cov}(X, Y)}{SD(X) \cdot SD(Y)}$$

yields (7.46).

▷ **Solution 7-9**

$$\begin{aligned}
\alpha_1^* &= \frac{\text{Cov}(aX, bY)}{\text{Var}(aX)} && \text{[Th. 7.14 (ii)]} \\
&= \frac{ab \cdot \text{Cov}(X, Y)}{a^2 \cdot \text{Var}(X)} && \text{[Box 7.1 (iii), Box 6.2 (iii)]} \\
&= \frac{b}{a} \cdot \alpha_1. && \text{[Th. 7.14 (ii)]}
\end{aligned}$$

According to Theorem 7.14 (ii), the slope of the linear quasi-regression of  $c+Y$  on  $d+X$ , is

$$\begin{aligned}
\frac{\text{Cov}(c+X, d+Y)}{\text{Var}(c+X)} &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} && \text{[Box 7.1 (ii), Box 6.2 (ii)]} \\
&= \alpha_1. && \text{[Th. 7.14 (ii)]}
\end{aligned}$$

▷ **Solution 7-10** (i) Equation (7.27) and Rule (i) of Box 6.1 imply

$$E(\mathbf{x}) = [E(X_1), \dots, E(X_n)]' = [a_1, \dots, a_n]' = \mathbf{a}.$$

(ii) Equation (7.27) and Rule (ii) of Box 6.1 imply

$$E(\mathbf{a} + \mathbf{x}) = \begin{bmatrix} E(a_1 + X_1) \\ \vdots \\ E(a_n + X_n) \end{bmatrix} = \begin{bmatrix} a_1 + E(X_1) \\ \vdots \\ a_n + E(X_n) \end{bmatrix} = \mathbf{a} + E(\mathbf{x}).$$

(iii) Equation (7.31) and Rule (vii) of Box 6.1 imply

$$E(\mathbf{a}'\mathbf{x}) = E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \cdot E(X_i) = \mathbf{a}'E(\mathbf{x}).$$

The other equations summarized in (iii) follow from the fact that  $\mathbf{a}'\mathbf{x}$  is a one-dimensional random variable (see Example 2.61) and  $\mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a}$ .

(v) Let  $\mathbf{a}'_l$  and  $\mathbf{b}'_l$ ,  $l = 1, \dots, k$ , denote the row vectors of  $\mathbf{A}$  and  $\mathbf{B}$ , respectively. Applying Equation (7.27), Rule (vi) of Box 6.1, and Rule (iii) to the terms  $\mathbf{a}'_l \mathbf{x}$  and  $\mathbf{b}'_l \mathbf{y}$ ,  $l = 1, 2, \dots, k$ , respectively, yields

$$\begin{aligned}
E(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y}) &= E \begin{bmatrix} \mathbf{a}'_1 \mathbf{x} + \mathbf{b}'_1 \mathbf{y} \\ \vdots \\ \mathbf{a}'_k \mathbf{x} + \mathbf{b}'_k \mathbf{y} \end{bmatrix} = \begin{bmatrix} E(\mathbf{a}'_1 \mathbf{x} + \mathbf{b}'_1 \mathbf{y}) \\ \vdots \\ E(\mathbf{a}'_k \mathbf{x} + \mathbf{b}'_k \mathbf{y}) \end{bmatrix} = \begin{bmatrix} E(\mathbf{a}'_1 \mathbf{x}) + E(\mathbf{b}'_1 \mathbf{y}) \\ \vdots \\ E(\mathbf{a}'_k \mathbf{x}) + E(\mathbf{b}'_k \mathbf{y}) \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{a}'_1 E(\mathbf{x}) + \mathbf{b}'_1 E(\mathbf{y}) \\ \vdots \\ \mathbf{a}'_k E(\mathbf{x}) + \mathbf{b}'_k E(\mathbf{y}) \end{bmatrix} = \mathbf{A}E(\mathbf{x}) + \mathbf{B}E(\mathbf{y}).
\end{aligned}$$

(iv) This rule is a special case of Rule (v) with  $\mathbf{B} = \mathbf{0}$ .

(vi) Let  $\mathbf{a}_l$  and  $\mathbf{x}_l$ ,  $l = 1, \dots, k$ , denote the column vectors of  $\mathbf{A}'$  and  $\mathbf{X}$ , respectively. Then Equations (7.27), (7.29), and Rule (i) imply

$$E(\mathbf{X}) = [E(\mathbf{x}_1), \dots, E(\mathbf{x}_k)] = [E(\mathbf{a}_1), \dots, E(\mathbf{a}_k)] = [\mathbf{a}_1, \dots, \mathbf{a}_k] = \mathbf{A}'.$$

(vii) Let  $\mathbf{a}_l$  and  $\mathbf{x}_l$ ,  $l = 1, \dots, k$ , denote the column vectors of  $\mathbf{A}'$  and  $\mathbf{X}$ , respectively. Then Equations (7.27), (7.29) and Rule (ii) imply

$$\begin{aligned} E(\mathbf{A}' + \mathbf{X}) &= E(\mathbf{a}_1 + \mathbf{x}_1, \dots, \mathbf{a}_k + \mathbf{x}_k) = [E(\mathbf{a}_1 + \mathbf{x}_1), \dots, E(\mathbf{a}_k + \mathbf{x}_k)] \\ &= [\mathbf{a}_1 + E(\mathbf{x}_1), \dots, \mathbf{a}_k + E(\mathbf{x}_k)] = \mathbf{A}' + E(\mathbf{X}). \end{aligned}$$

(viii) Let  $\mathbf{c}'_i$ ,  $i = 1, \dots, l$ , denote the row vectors of  $\mathbf{C}$ . Then Equations (7.27) and (7.29) imply

$$E(\mathbf{C}\mathbf{X}) = E \begin{bmatrix} \mathbf{c}'_1 \mathbf{X} \\ \vdots \\ \mathbf{c}'_l \mathbf{X} \end{bmatrix} = \begin{bmatrix} E(\mathbf{c}'_1 \mathbf{X}) \\ \vdots \\ E(\mathbf{c}'_l \mathbf{X}) \end{bmatrix} = \begin{bmatrix} \mathbf{c}'_1 E(\mathbf{X}) \\ \vdots \\ \mathbf{c}'_l E(\mathbf{X}) \end{bmatrix} = \mathbf{C}E(\mathbf{X}).$$

(ix). Rule (viii), Equation (7.30), and the rules for the transpose of a matrix yield

$$\begin{aligned} E(\mathbf{C}\mathbf{X}\mathbf{Y}'\mathbf{D}') &= \mathbf{C}E(\mathbf{X}\mathbf{Y}'\mathbf{D}') && \text{[(viii)]} \\ &= \mathbf{C}(E(\mathbf{D}\mathbf{Y}\mathbf{X}'))' && \text{[(7.30)]} \\ &= \mathbf{C}(\mathbf{D}E(\mathbf{Y}\mathbf{X}'))' && \text{[(viii)]} \\ &= \mathbf{C}E(\mathbf{Y}\mathbf{X}')'\mathbf{D}' && \text{[(7.30)]} \\ &= \mathbf{C}E(\mathbf{X}\mathbf{Y}')\mathbf{D}'. && \text{[(7.30)]} \end{aligned}$$

▷ **Solution 7-11** (i)

$$\begin{aligned} \Sigma_{\mathbf{x}\mathbf{y}} &= E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]') && \text{[(7.33)]} \\ &= E(\mathbf{x}\mathbf{y}' - \mathbf{x}E(\mathbf{y})' - E(\mathbf{x})\mathbf{y}' + E(\mathbf{x})E(\mathbf{y})') \\ &= E(\mathbf{x}\mathbf{y}') - E(\mathbf{x})E(\mathbf{y})' - E(\mathbf{x})E(\mathbf{y}') + E(\mathbf{x})E(\mathbf{y})' && \text{[Box 7.2 (iii)]} \\ &= E(\mathbf{x}\mathbf{y}') - E(\mathbf{x})E(\mathbf{y}'). \end{aligned}$$

(ii)

$$\begin{aligned} \Sigma_{\mathbf{a}+\mathbf{x}, \mathbf{b}+\mathbf{y}} &= E([\mathbf{a} + \mathbf{x} - E(\mathbf{a} + \mathbf{x})][\mathbf{b} + \mathbf{y} - E(\mathbf{b} + \mathbf{y})]') && \text{[(7.33)]} \\ &= E([\mathbf{a} + \mathbf{x} - \mathbf{a} - E(\mathbf{x})][\mathbf{b} + \mathbf{y} - \mathbf{b} - E(\mathbf{y})]') && \text{[Box 7.2 (ii)]} \\ &= E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]') \\ &= \Sigma_{\mathbf{x}\mathbf{y}}. && \text{[(7.33)]} \end{aligned}$$

(iii)

$$\begin{aligned} \Sigma_{\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}} &= E([\mathbf{A}\mathbf{x} - E(\mathbf{A}\mathbf{x})][\mathbf{B}\mathbf{y} - E(\mathbf{B}\mathbf{y})]') && \text{[(7.33)]} \\ &= E([\mathbf{A}\mathbf{x} - \mathbf{A}E(\mathbf{x})][\mathbf{B}\mathbf{y} - \mathbf{B}E(\mathbf{y})]') && \text{[Box 7.2 (iv)]} \\ &= E[\mathbf{A}[\mathbf{x} - E(\mathbf{x})][\mathbf{B}[\mathbf{y} - E(\mathbf{y})]]'] \\ &= E(\mathbf{A}[\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]'\mathbf{B}') \\ &= \mathbf{A}E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]')\mathbf{B}' && \text{[Box 7.2 (ix)]} \\ &= \mathbf{A}\Sigma_{\mathbf{x}\mathbf{y}}\mathbf{B}'. && \text{[(7.33)]} \end{aligned}$$

(iv)

$$\begin{aligned} \Sigma_{\mathbf{x}\mathbf{y}} &= E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]') && \text{[(7.33)]} \\ &= \left( E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]')' \right)' && \text{[(\mathbf{A}')' = \mathbf{A}]} \\ &= E([\mathbf{x} - E(\mathbf{x})][\mathbf{y} - E(\mathbf{y})]')' && \text{[(7.30)]} \\ &= E([\mathbf{y} - E(\mathbf{y})][\mathbf{x} - E(\mathbf{x})]')' && \text{[(\mathbf{a}\mathbf{b}')' = \mathbf{b}\mathbf{a}']} \end{aligned}$$

$$= \boldsymbol{\Sigma}'_{y,x}. \quad [(7.33)]$$

(v) Independence of the multivariate random variables  $X$  and  $Y$  implies  $X_i \perp\!\!\!\perp Y_j$ , for all  $i = 1, \dots, n$  and  $j = 1, \dots, m$ . Therefore, Rule (vi) of Box 7.1 implies  $\text{Cov}(X_i, Y_j) = 0$ , for all  $i = 1, \dots, n$  and  $j = 1, \dots, m$ . Equation (7.34) then implies  $\boldsymbol{\Sigma}_{xy} = \mathbf{0}$ .

(vi) If  $\mathbf{x} \stackrel{p}{=} \mathbf{a}$ , then Rule (vii) of Box (7.1) yields  $\text{Cov}(X_i, Y_j) = 0$ , for all  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ . Equation (7.34) then implies  $\boldsymbol{\Sigma}_{xy} = \mathbf{0}$ .

(vii)

$$\begin{aligned} & \boldsymbol{\Sigma}_{\mathbf{Ax} + \mathbf{By}, \mathbf{Cw} + \mathbf{Dz}} \\ &= E\{[\mathbf{Ax} + \mathbf{By} - E(\mathbf{Ax} + \mathbf{By})] [\mathbf{Cw} + \mathbf{Dz} - E(\mathbf{Cw} + \mathbf{Dz})]'\} \quad [(7.33)] \\ &= E\{[\mathbf{Ax} - E(\mathbf{Ax}) + \mathbf{By} - E(\mathbf{By})] [\mathbf{Cw} - E(\mathbf{Cw}) + \mathbf{Dz} - E(\mathbf{Dz})]'\} \\ &= E\{[\mathbf{Ax} - E(\mathbf{Ax})] [\mathbf{Cw} - E(\mathbf{Cw})]'\} + E\{[\mathbf{Ax} - E(\mathbf{Ax})] [\mathbf{Dz} - E(\mathbf{Dz})]'\} \\ &\quad + E\{[\mathbf{By} - E(\mathbf{By})] [\mathbf{Cw} - E(\mathbf{Cw})]'\} + E\{[\mathbf{By} - E(\mathbf{By})] [\mathbf{Dz} - E(\mathbf{Dz})]'\} \quad [\text{Box 7.2 (iv)}] \\ &= E\{\mathbf{A}[\mathbf{x} - E(\mathbf{x})] [\mathbf{w} - E(\mathbf{w})]'\mathbf{C}'\} + E\{\mathbf{A}[\mathbf{x} - E(\mathbf{x})] [\mathbf{z} - E(\mathbf{z})]'\mathbf{D}'\} \\ &\quad + E\{\mathbf{B}[\mathbf{y} - E(\mathbf{y})] [\mathbf{w} - E(\mathbf{w})]'\mathbf{C}'\} + E\{\mathbf{B}[\mathbf{y} - E(\mathbf{y})] [\mathbf{z} - E(\mathbf{z})]'\mathbf{D}'\} \\ &= \mathbf{AE}\{[\mathbf{x} - E(\mathbf{x})] [\mathbf{w} - E(\mathbf{w})]'\}\mathbf{C}' + \mathbf{AE}\{[\mathbf{x} - E(\mathbf{x})] [\mathbf{z} - E(\mathbf{z})]'\}\mathbf{D}' \\ &\quad + \mathbf{BE}\{[\mathbf{y} - E(\mathbf{y})] [\mathbf{w} - E(\mathbf{w})]'\}\mathbf{C}' + \mathbf{BE}\{[\mathbf{y} - E(\mathbf{y})] [\mathbf{z} - E(\mathbf{z})]'\}\mathbf{D}' \quad [\text{Box 7.2 (ix)}] \\ &= \mathbf{A}\boldsymbol{\Sigma}_{xw}\mathbf{C}' + \mathbf{A}\boldsymbol{\Sigma}_{xz}\mathbf{D}' + \mathbf{B}\boldsymbol{\Sigma}_{yw}\mathbf{C}' + \mathbf{B}\boldsymbol{\Sigma}_{yz}\mathbf{D}'. \quad [(7.33)] \end{aligned}$$

(viii) If  $\mathbf{x} \stackrel{p}{=} \mathbf{z}$ , then Rule (x) of Box 7.1 implies that  $\text{Cov}(X_i, Y_j) = \text{Cov}(Z_i, Y_j)$ , for all  $i = 1, \dots, n$ ,  $j = 1, \dots, m$ . Equation (7.34) then implies  $\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{zy}$ .

▷ **Solution 7-12**

$$Q_{Y|X}^2 = \frac{\text{Var}[Q_{lin}(Y|X)]}{\text{Var}(Y)} \quad [(7.40)]$$

$$= \frac{\text{Cov}[Y, Q_{lin}(Y|X)]}{\text{Var}(Y)} \quad [(7.44)]$$

$$= \frac{\text{Cov}[Y, Q_{lin}(Y|X)] \cdot \text{Var}[Q_{lin}(Y|X)]}{\text{Var}[Q_{lin}(Y|X)] \cdot \text{Var}(Y)} \quad [(7.44)]$$

$$= \frac{\text{Cov}[Y, Q_{lin}(Y|X)] \cdot \text{Cov}[Y, Q_{lin}(Y|X)]}{\text{Var}[Q_{lin}(Y|X)] \cdot \text{Var}(Y)} \quad [(7.44)]$$

$$= \frac{\text{Cov}[Y, Q_{lin}(Y|X)]^2}{\text{SD}[Q_{lin}(Y|X)]^2 \cdot \text{SD}(Y)^2}$$

$$= \text{Corr}[Y, Q_{lin}(Y|X)]^2.$$

Note that  $\text{Corr}[Y, Q_{lin}(Y|X)] = 0$  if  $\text{Var}[Q_{lin}(Y|X)] = 0$  [see Eq. (7.18)].

## Chapter 8

### Some Distributions

In chapter 5 we defined *random variables* as particular measurable mappings and their *distributions* as their image measures. There we also extended the concept of independence of events and families of events to *independence of random variables* and *independence of families of random variables*. Furthermore, we introduced the concepts of a *probability function* and a *density* of real-valued random variables, which are useful for describing distributions. In chapters 6 and 7 we treated the expectation of a numerical random variable and related concepts such as variance, covariance, and correlation. In this chapter we provide some examples illustrating how probability functions and densities describe the distribution of a random variable and how they can be used to compute expectations and variances of real-valued random variables.

#### 8.1 Some Distributions of Discrete Random Variables

In this section we treat some examples of distributions that are specified by probability functions of discrete random variables: the discrete uniform, binomial, Poisson, and geometric distributions.

##### 8.1.1 Discrete Uniform Distribution

Reading the following definition, remember that  $p_X$  denotes the probability function assigning to each value  $x_i$  of a discrete random variable  $X$  its probability  $p_X(x) = P(X=x_i) = P_X(\{x_i\})$  (see Def. 5.56 and Rem. 5.57). Furthermore, according to Remark 5.59,  $p_X$  uniquely determines the distribution  $P_X$  of  $X$ , which follows from  $\sigma$ -additivity of the measure  $P_X$ .

**Definition 8.1 (Discrete Uniform Distribution on a Finite Set)**

For  $n \in \mathbb{N}$ , let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{P}(\Omega'_X))$  be a random variable, where  $\Omega'_X = \{x_1, \dots, x_n\}$ . Then  $X$  has a (discrete) uniform distribution on  $\Omega'_X$ , if

$$p_X(x_i) = P(X=x_i) = \frac{1}{n}, \quad \forall i = 1, \dots, n. \quad (8.1)$$

Let  $\Omega'_X = \{x_1, \dots, x_n\}$  and let  $\#A'$  denote the number elements of the set  $A'$ . According to Equation (5.46),

$$\left( \forall A' \subset \Omega'_X: P_X(A') = \frac{\#A'}{n} \right) \Leftrightarrow P_X \text{ is the uniform distribution on } \Omega'_X. \quad (8.2)$$

Hence, if  $X$  has a discrete uniform distribution, then the probability of an event  $\{X \in A'\}$  only depends on  $\#A'$ , the number of elements of  $A'$ , and not on the particular choice of these elements from  $\Omega'_X$ . The uniform distribution on  $\Omega'_X$  is the only distribution that has this property. A special case will be treated in Example 8.6.

### 8.1.2 Bernoulli Distribution

Now we consider  $n$  trials in which we observe whether or not an event  $A_i$  occurs at trial  $i$ , where  $i = 1, \dots, n$ . These events  $A_i$  will be indicated by independent identically distributed (i. i. d.) random variables  $X_1, \dots, X_n$  with values  $X_i(\omega) = 1$  if  $\omega \in A_i$ , and  $X_i(\omega) = 0$  otherwise. Hence,  $X_i = 1_{A_i}$ ,  $i = 1, \dots, n$  and  $P(X_i = 1) = p$  for all  $i = 1, \dots, n$ .

#### Definition 8.2 (Finite Sequence of Bernoulli Variables)

Let  $X_1, \dots, X_n: (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  be i. i. d. random variables with

$$P(X_i = 1) = 1 - P(X_i = 0) = p, \quad \forall i = 1, \dots, n, \quad (8.3)$$

where  $0 \leq p \leq 1$ . Then  $X_1, \dots, X_n$  is called a sequence of  $n$  Bernoulli variables with parameter  $p$ , and  $X = (X_1, \dots, X_n)$  is called an  $n$ -variate Bernoulli variable with parameter  $p$ .

**Remark 8.3 (Bernoulli Distribution)** Remember that  $a^1 = a$  and  $a^{1-1} = a^0 = 1$  for  $a \in \mathbb{R}$ . Hence, for all  $i = 1, \dots, n$ ,

$$p_{X_i}(x) = p^x (1-p)^{1-x}, \quad \forall x \in \{0, 1\}. \quad (8.4)$$

The distribution of such a random variable  $X_i$  is called the *Bernoulli distribution*.

◁

As we will show in Remark 8.5, a multivariate Bernoulli variable has a multivariate Bernoulli distribution that will now be defined.

#### Definition 8.4 (Multivariate Bernoulli Distribution)

Let  $X = (X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n))$  be an  $n$ -variate random variable and let  $0 \leq p \leq 1$ . If the probability function of  $X$  is

$$\forall (x_1, \dots, x_n) \in \{0, 1\}^n: p_X(x_1, \dots, x_n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i}, \quad (8.5)$$

then  $P_X$  is called the *multivariate Bernoulli distribution* with parameters  $n$  and  $p$ .

Because  $x_i \in \{0, 1\}$  for all  $i = 1, \dots, n$ , the probability  $p_X(x_1, \dots, x_n)$  in (8.5) can also be written

$$p_X(x_1, \dots, x_n) = p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n-\sum_{i=1}^n x_i}, \quad (8.6)$$

where  $\sum_{i=1}^n x_i$  is the number of ones in the sequence  $x_1, \dots, x_n$  (see Exercise 8-1).

**Remark 8.5 (Probability Function of a Multivariate Bernoulli Variable)** Let  $X = (X_1, \dots, X_n)$  be an  $n$ -variate Bernoulli variable with parameter  $p$ . Then its probability function  $p_X$  is defined by (8.5) and its distribution,  $P_X$ , is the multivariate Bernoulli distribution with parameters  $n$  and  $p$ , which follows from (8.4), independence of  $X_1, \dots, X_n$ , and Lemma 5.68.  $\triangleleft$

**Example 8.6 (A Special Case)** For  $p = .5$ , Equation (8.6) implies  $p_X(x_1, \dots, x_n) = .5^n$ , for all  $(x_1, \dots, x_n) \in \{0, 1\}^n$ . Hence, the multivariate Bernoulli distribution is the discrete uniform distribution on  $\{0, 1\}^n$ , i. e., in this case an  $n$ -variate Bernoulli variable  $X = (X_1, \dots, X_n)$  with parameter  $p = .5$  is uniformly distributed on  $\{0, 1\}^n$ .  $\triangleleft$

### 8.1.3 Binomial Distribution

In the following definition we use the binomial coefficient

$$\binom{n}{x} := \frac{n!}{x! \cdot (n-x)!}, \quad n \in \mathbb{N}_0, \quad x = 0, 1, \dots, n, \quad (8.7)$$

Furthermore,  $a! := a \cdot (a-1) \cdot \dots \cdot 1$  denotes the *factorial* of  $a \in \mathbb{N}_0$ , where by convention  $0! := 1$ .

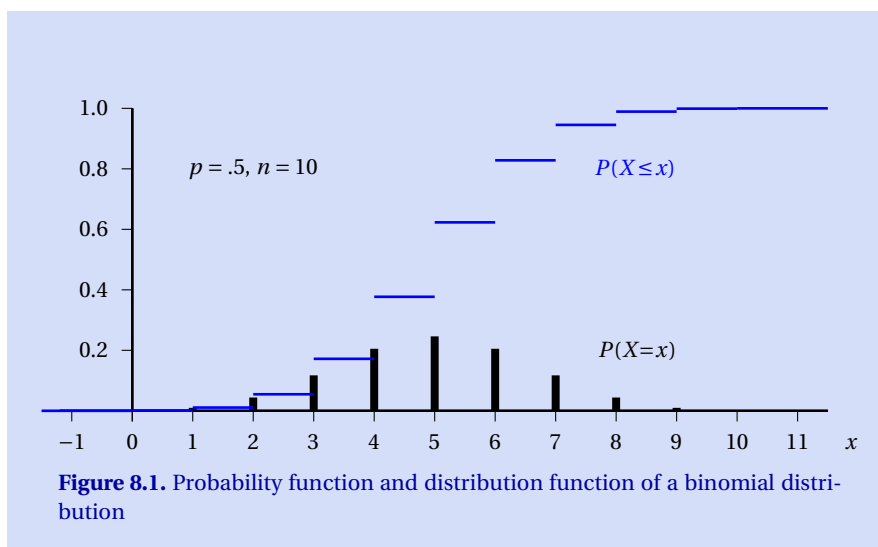
#### Definition 8.7 (Binomial Distribution)

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$  be a discrete random variable. If

$$\forall x \in \mathbb{N}_0: \quad p_X(x) = b_{n,p}(x) := \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, \dots, n, \\ 0, & \text{if } x > n, \end{cases} \quad (8.8)$$

where  $n \in \mathbb{N}$  and  $0 \leq p \leq 1$ , then we use the notation  $X \sim \mathcal{B}_{n,p}$  and say that  $X$  has a *binomial distribution* with parameters  $n$  and  $p$ .

**Remark 8.8 (Distribution Function)** If  $b_{n,p}$  is the probability function of  $X$ , i. e., if  $X \sim \mathcal{B}_{n,p}$ , then its *distribution function* is



$$\forall a \in \mathbb{R}: F_X(a) = \begin{cases} 0, & \text{if } a < 0, \\ \sum_{x=0}^k b_{n,p}(x), & \text{if } k \leq a < k+1 \text{ and } k = 0, 1, \dots, n-1, \\ 1, & \text{if } a \geq n. \end{cases} \quad (8.9)$$

Figure 8.1 shows the probability function and the distribution function of the binomial distribution for parameters  $n = 10$  and  $p = .5$ . Figure 8.3 also shows the probability function for parameters  $n = 10$ ,  $p = .2$  and  $n = 20$ ,  $p = .1$ . Note that probabilities close to zero may not be visible in this figure.  $\triangleleft$

According to the following theorem the sum of a sequence of  $n$  Bernoulli variables with parameter  $p$  has a binomial distribution with parameters  $n$  and  $p$ .

**Theorem 8.9 (Sum of Bernoulli Variables)**

Let  $X_1, \dots, X_n$  be a sequence of  $n$  Bernoulli variables with parameter  $p$ . Then

$$X := \sum_{i=1}^n X_i \sim \mathcal{B}_{n,p}. \quad (8.10)$$

(Proof p. 267)

**Remark 8.10 (A Reformulation Using Events)** Explicitly referring to events and their indicators, the theorem may also be formulated as follows: If  $(\Omega, \mathcal{A}, P)$  is a probability space and, for all  $i = 1, \dots, n$ ,  $A_i \in \mathcal{A}$  are independent with  $P(A_i) = p \geq 0$ , then  $X := \sum_{i=1}^n 1_{A_i} \sim \mathcal{B}_{n,p}$ .  $\triangleleft$

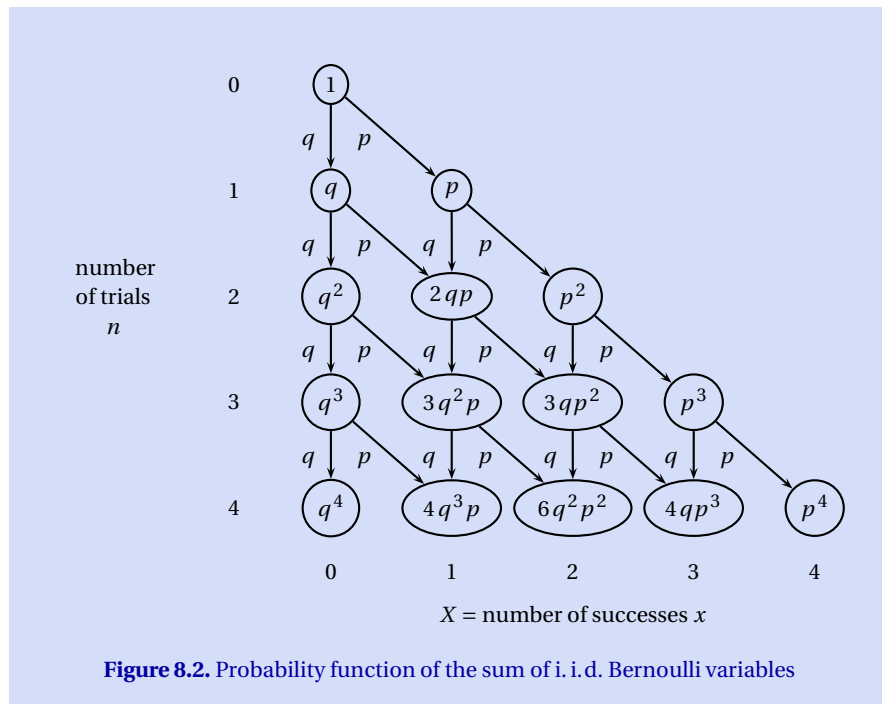


Figure 8.2 illustrates the probability function  $p_X$  of  $X = X_1 + \dots + X_n$ , where  $X_i$  indicates ‘success’ in trial  $i$  occurring with probability  $p = P(X_i=1)$  ( $q := 1 - p$ ). We consider  $n = 0, 1, \dots, 4$  trials presented in the five “rows” of the figure. The numbers in the circles and ellipses are the probabilities  $p_X(x)$  of the number  $x$  of successes, which are computed by Equation (8.8) (see the “columns” of the figure).

**Example 8.11 (Flipping a Coin  $n$  Times)** A simple example is flipping a coin  $n$  times. The event  $A_i$  is *flipping heads at trial  $i$* ,  $i = 1, \dots, n$ . If  $p = 1/2$ , then we say that the coin is fair. If  $X$  is the number of flipping heads, i. e.,  $X = \sum_{i=1}^n 1_{A_i}$ , then, for  $n = 2$ , the values of the binomial distribution are  $b_{2,1/2}(0) = P(X=0) = 1/4$ ,  $b_{2,1/2}(1) = P(X=1) = 1/2$ , and  $b_{2,1/2}(2) = P(X=2) = 1/4$ , which have already been computed in Example 5.64. For the case  $n = 4$  see Exercise 8-2.  $\triangleleft$

**Example 8.12 (Tossing  $n$  Dices)** Another example is tossing  $n$  dices. The events  $A_i$  could be *tossing a six with dice  $i$* . If the probabilities are identical for all six possible outcomes of a single trial, then we say that the dice is fair. In this case the parameter is  $p = 1/6$ . If  $X$  is the number of tossing a six, then, for  $n = 2$ , the values of the binomial distribution are  $b_{2,1/6}(0) = P(X=0) = 25/36$ ,  $b_{2,1/6}(1) = P(X=1) = 10/36$ , and  $b_{2,1/6}(2) = P(X=2) = 1/36$ .  $\triangleleft$

The following corollary shows how expectation and variance of a random variable  $X$  with a binomial distribution depend on the parameters  $n$  and  $p$ .

**Corollary 8.13 (Expectation and Variance)**

If  $X \sim \mathcal{B}_{n,p}$ , then  $E(X) = np$  and  $\text{Var}(X) = np(1-p)$ .

(Proof p. 268)

### 8.1.4 Poisson Distribution

Another discrete distribution is the Poisson distribution. As we will see in Theorem 8.17, it is “close” to the binomial distribution if  $n$  is large and  $p$  small.

**Definition 8.14 (Poisson Distribution)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$  be a discrete random variable. If

$$p_X(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}, \quad \forall x \in \mathbb{N}_0, \quad (8.11)$$

where  $\lambda > 0$ , then we use the notation  $X \sim \mathcal{P}_\lambda$  and say that  $X$  has a Poisson distribution with parameter  $\lambda$ .

Figure 8.3 displays this probability function for the parameter  $\lambda = 2$ .

**Remark 8.15 (Distribution Function)** If  $X \sim \mathcal{P}_\lambda$ , then its *distribution function* is obtained by sums of the probabilities  $p_X(x)$  specified in Equation (8.11). More precisely,

$$\forall a \in \mathbb{R}: F_X(a) = \begin{cases} 0, & \text{if } a < 0, \\ e^{-\lambda} \cdot \sum_{x=0}^n \frac{\lambda^x}{x!}, & \text{if } n \leq a < n+1 \text{ and } n \in \mathbb{N}_0. \end{cases} \quad (8.12)$$

◁

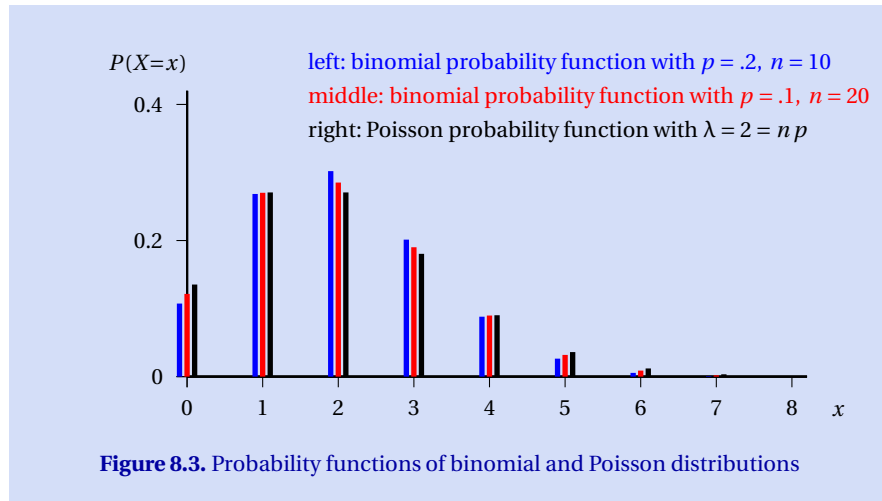
If  $X$  has a Poisson distribution, then, according to the following theorem, its expectation and variance are identical, and they are equal to the parameter  $\lambda$ .

**Theorem 8.16 (Expectation and Variance)**

If  $X \sim \mathcal{P}_\lambda$ , then  $E(X) = \text{Var}(X) = \lambda$ .

(Proof p. 268)

We can use the Poisson distribution for approximating the binomial distribution for large  $n$  and small  $p$ . The following theorem is the theoretical foundation.

**Theorem 8.17 (Poisson Limit Theorem)**

Let  $p_n$ ,  $n \in \mathbb{N}$ , be a sequence of real numbers with  $0 \leq p_n \leq 1$  for all  $n \in \mathbb{N}$ . If  $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda$  and  $0 < \lambda < \infty$ , then

$$\lim_{n \rightarrow \infty} b_{n,p_n}(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}, \quad \forall x \in \mathbb{N}_0, \quad (8.13)$$

where  $b_{n,p_n}$  denotes the probability function of a binomial distribution with parameters  $n$  and  $p_n$ .

For a proof see Bauer (1996, p. 220).

**Remark 8.18 (Poisson Distribution as an Approximation)** Note that assuming  $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda > 0$  implies that the sequence  $p_1, p_2, \dots$  converges to 0. Hence,  $p_n$  will be small for a large  $n$ . Therefore, Equation (8.13) can be applied to a binomially distributed random variable  $X$  for large  $n$  and small  $p$ . In this case,  $p$  takes the role of  $p_n$  and  $\lambda = n \cdot p$  takes the role of  $\lim_{n \rightarrow \infty} n \cdot p_n$  in Theorem 8.17. Then

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \approx \frac{\lambda^x}{x!} \cdot e^{-\lambda}, \quad \forall x = 0, 1, \dots, n,$$

i. e., the Poisson probability function approximates the Binomial probability function (see Fig. 8.3). Obviously, the approximation is better for the parameter  $n = 20, p = .1$  than for  $n = 10, p = .2$ , i. e., it is better for larger  $n$  and smaller  $p$ .

◁

### 8.1.5 Geometric Distribution

Now we turn to an infinite sequence of Bernoulli trials, which is useful if the number of trials cannot be fixed in advance. An example is the random experiment of flipping a coin until the first *heads* occurs (see Example 8.24). In contrast to the last section in which we considered a finite sequence of Bernoulli trials, now we will assume  $p \neq 0$  and  $p \neq 1$ .

**Definition 8.19 (Infinite Sequence of Bernoulli Variables)**

Let  $X_1, X_2, \dots: (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  be i. i. d. random variables with

$$P(X_i=1) = 1 - P(X_i=0) = p, \quad \forall i = 1, 2, \dots, \quad (8.14)$$

where  $0 < p < 1$ . Then  $X_1, X_2, \dots$  is called an infinite sequence of Bernoulli variables with parameter  $p$ .

The set  $\Omega$  occurring in this definition can be specified as the set of all infinite sequences of elements of  $\{0, 1\}$ . Note that it is not obvious that there is an infinite sequence of independent random variables  $X_i$  satisfying (8.14). Klenke (2013, Th. 2.19 and Example 2.18, pp. 55, 56) shows that such an infinite sequence actually exists.

In Theorem 8.23 and Example 8.24 we will see that such an infinite sequence of Bernoulli variables is closely related to the geometric distribution that is defined as follows.

**Definition 8.20 (Geometric Distribution)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{N}, \mathcal{P}(\mathbb{N}))$  be a discrete random variable. If

$$p_X(x) = (1 - p)^{x-1} \cdot p, \quad \forall x \in \mathbb{N}, \quad (8.15)$$

where  $0 < p < 1$ , then we use the notation  $X \sim \mathcal{G}_p$  and say that  $X$  has a geometric distribution with parameter  $p$ .

Note that, for  $0 < p < 1$ ,

$$\sum_{x=1}^{\infty} (1 - p)^{x-1} = \frac{1}{1 - (1 - p)} = \frac{1}{p},$$

which implies  $\sum_{x=1}^{\infty} p_X(x) = 1$ .

**Remark 8.21 (An Alternative Definition)** Sometimes the geometric distribution is alternatively defined for a random variable  $Y$  with values  $y \in \mathbb{N}_0$ . In this case, the probability function is  $p_Y(y) = (1 - p)^y \cdot p$ .  $\triangleleft$

**Theorem 8.22 (Expectation and Variance)**

If  $X \sim \mathcal{G}_p$ , then  $E(X) = 1/p$  and  $\text{Var}(X) = (1/p^2) - 1/p$ .

(Proof p. 269)

In the following theorem we consider the probability function of the random variable  $X$ , the number of the trial in which the first *one* occurs in an infinite sequence of Bernoulli trials.

**Theorem 8.23 (A Class of Random Variables With Geometric Distributions)**

Let  $X_1, X_2, \dots: (\Omega, \mathcal{A}, P) \rightarrow (\{0, 1\}, \mathcal{P}(\{0, 1\}))$  be an infinite sequence of Bernoulli variables with parameter  $p$ . If the random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{N}, \mathcal{P}(\mathbb{N}))$  is defined by

$$X(\omega) = \min\{n \in \mathbb{N}: X_n(\omega) = 1\}, \quad \forall \omega \in \Omega, \quad (8.16)$$

then  $X$  has a geometric distribution with parameter  $p$ .

(Proof p. 270)

**Example 8.24 (Flipping a Coin Until Heads Occurs)** Consider repeatedly flipping a coin and define the sequence of random variables  $X_1, X_2, \dots$  with  $X_i(\omega) = 1$  if *heads* occurs at the  $i$ th flip and  $X_i(\omega) = 0$ , otherwise. This specifies an infinite sequence of Bernoulli variables with parameter  $p = .5$ . Now, define the random variable  $X$  by Equation (8.16). Hence,  $X$  is the index of the first of the variables  $X_i$  taking on the value 1, i. e.,

$$X(\omega) = n, \quad \text{if } X_1(\omega) = \dots = X_{n-1}(\omega) = 0 \text{ and } X_n(\omega) = 1, \quad \forall \omega \in \Omega, \quad n \in \mathbb{N}.$$

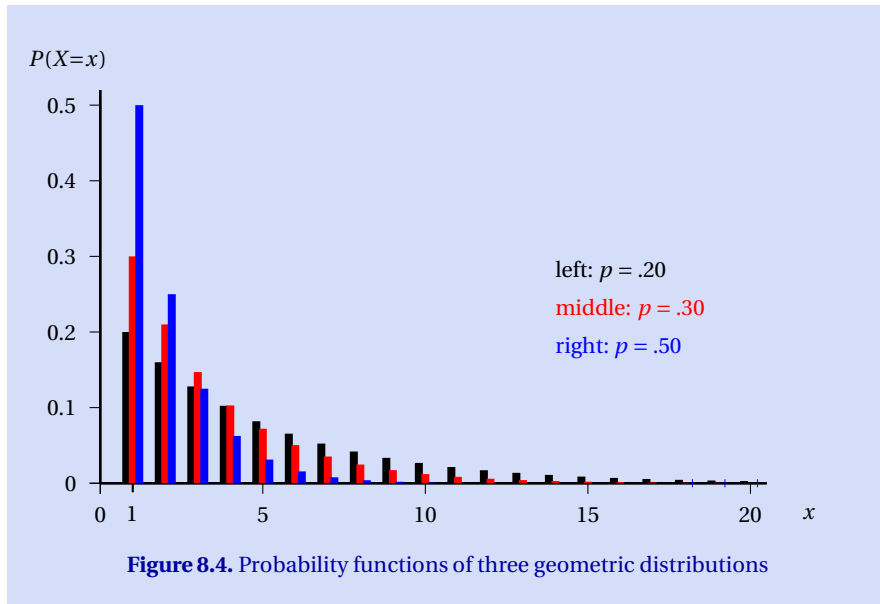
Then  $X$  has a geometric distribution with parameter  $p = .5$ . ◁

**Remark 8.25 (Distribution Function)** If  $X \sim \mathcal{G}_p$ , then

$$\forall x \in \mathbb{R}: \quad F_X(x) = P(X \leq x) = \begin{cases} 0, & \text{if } x < 1 \\ 1 - (1-p)^i, & \text{if } i \leq x < i+1, i \in \mathbb{N}, \end{cases} \quad (8.17)$$

is the distribution function of  $X$  (see Exercise 8-3). ◁

Another discrete distribution based on an infinite sequence  $X_1, X_2, \dots$  of Bernoulli variables is the negative Binomial (or Pascal) distribution (see, e. g., Johnson, Kemp, & Kotz, 2005).



## 8.2 Some Distributions of Continuous Random Variables

### 8.2.1 Continuous Uniform Distribution

We begin with some examples using the densities of various random variables with continuous uniform distributions.

**Example 8.26 (Uniform Distribution on an Interval)** Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a real-valued random variable. Then  $X$  has a *uniform distribution on the interval*  $[a, b]$ ,  $a < b$ ,  $a, b \in \mathbb{R}$ , if

$$f_X(x) = \frac{1}{b-a} \cdot \mathbf{1}_{[a,b]}(x). \quad (8.18)$$

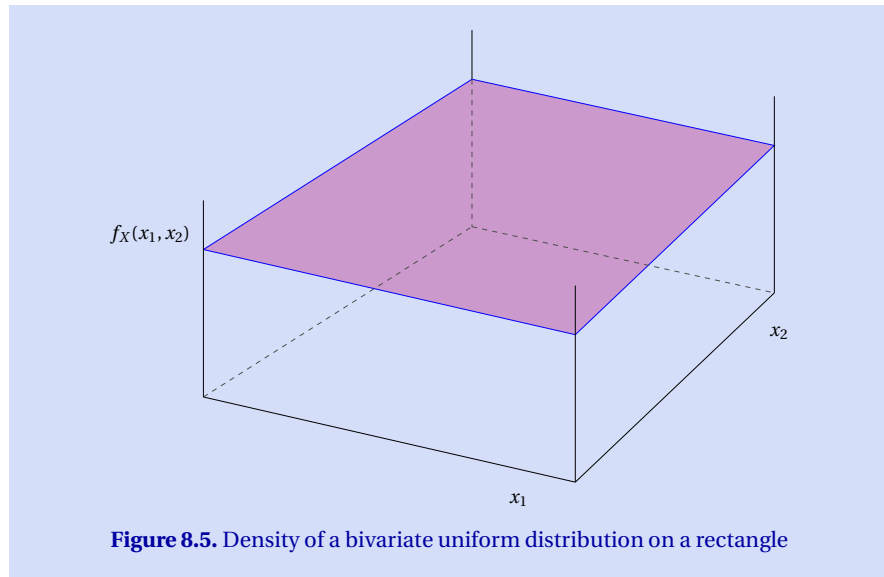
◁

**Example 8.27 (Uniform Distribution on a Rectangle)** The random variable  $X = (X_1, X_2): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^2, \mathcal{B}_2)$  has a *uniform distribution on the rectangle*  $[a, b] \times [c, d]$ ,  $a < b$ ,  $c < d$ ,  $a, b, c, d \in \mathbb{R}$ , if

$$f_X(x_1, x_2) = \frac{1}{(b-a) \cdot (d-c)} \cdot \mathbf{1}_{[a,b]}(x_1) \cdot \mathbf{1}_{[c,d]}(x_2) \quad (8.19)$$

(see Fig. 8.5). Note that  $X$  has a uniform distribution on the rectangle  $[a, b] \times [c, d]$  if and only if the following three conditions hold:

- (a)  $X_1$  and  $X_2$  are independent,
- (b)  $X_1$  has a uniform distribution on  $[a, b]$ , and



(c)  $X_2$  has a uniform distribution on  $[c, d]$

(see Cor. 5.100). ◁

**Example 8.28 (Uniform Distribution on a Circle)** Let  $X = (X_1, X_2): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^2, \mathcal{B}_2)$  be a random variable. Then  $X$  has a *uniform distribution on the circle*  $B_r = \{(x_1, x_2) \in \mathbb{R}^2: x_1^2 + x_2^2 \leq r\}$ ,  $0 < r \in \mathbb{R}$ , if

$$f_X(x_1, x_2) = \frac{1}{\pi \cdot r^2} 1_{B_r}(x_1, x_2). \quad (8.20)$$

In this case,  $X_1$  and  $X_2$  are *not* independent. ◁

**Definition 8.29 (Continuous Uniform Distribution on a Bounded Set)**

Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  be a random variable and let  $B \in \mathcal{B}_n$  such that  $0 < \lambda_n(B) < \infty$ . Then  $X$  has a (continuous) *uniform distribution on the set  $B$* , denoted  $X \sim \mathcal{U}_B$ , if  $X$  has a density satisfying

$$f_X(x) = \frac{1}{\lambda_n(B)} 1_B(x), \quad \forall x \in \mathbb{R}^n. \quad (8.21)$$

Equation (5.40) and Theorem 3.68 imply: For  $B \in \mathcal{B}_n$  with  $0 < \lambda_n(B) < \infty$ ,

$$\left( \forall A' \in \mathcal{B}_n: P(X \in A') = \frac{\lambda_n(A')}{\lambda_n(B)} \right) \Leftrightarrow X \sim \mathcal{U}_B. \quad (8.22)$$

Hence, the probability of an event  $\{X \in A'\}$ ,  $A' \in \mathcal{B}_n|_B$  only depends on the value  $\lambda_n(A')$  of the Lebesgue measure  $\lambda_n$ . This is an invariance property that characterizes a continuous uniform distribution, i. e., there is no other distribution on a bounded set  $B \in \mathcal{B}_n$  that shares this property with the uniform distribution.

In the special case  $n = 2$ , the probability  $P(X \in A')$  only depends on the area of  $A' \in \mathcal{B}_2$ ; it does *not depend* on the location or the shape (triangle, ellipse, ...) of the set  $A'$ . Figure 8.5 shows the density of a bivariate uniform distribution on a rectangle.

### 8.2.2 Normal Distribution

#### Definition 8.30 (Normal Distribution)

A continuous random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  has a normal distribution with parameters  $\mu \in \mathbb{R}$  and  $\sigma^2 > 0$ , denoted  $X \sim \mathcal{N}_{\mu, \sigma^2}$ , if it has a density satisfying

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}. \quad (8.23)$$

Figure 8.6 displays the densities of the normal distributions for three different parameter pairs of  $\mu$  and  $\sigma^2$ . Comparing the three densities to each other illustrates that  $\mu$  is a location parameter and  $\sigma^2$  a scale parameter. These parameters also have a stochastic meaning.

#### Theorem 8.31 (Expectation and Variance)

If  $X \sim \mathcal{N}_{\mu, \sigma^2}$ , then  $E(X) = \mu$  and  $\text{Var}(X) = \sigma^2$ .

For a proof see Georgii (2008, Example 4.28, p. 108).

The normal distribution with expectation  $\mu = 0$  and variance  $\sigma^2 = 1$  is called the *standard normal distribution*. In this case, Equation (8.23) simplifies to

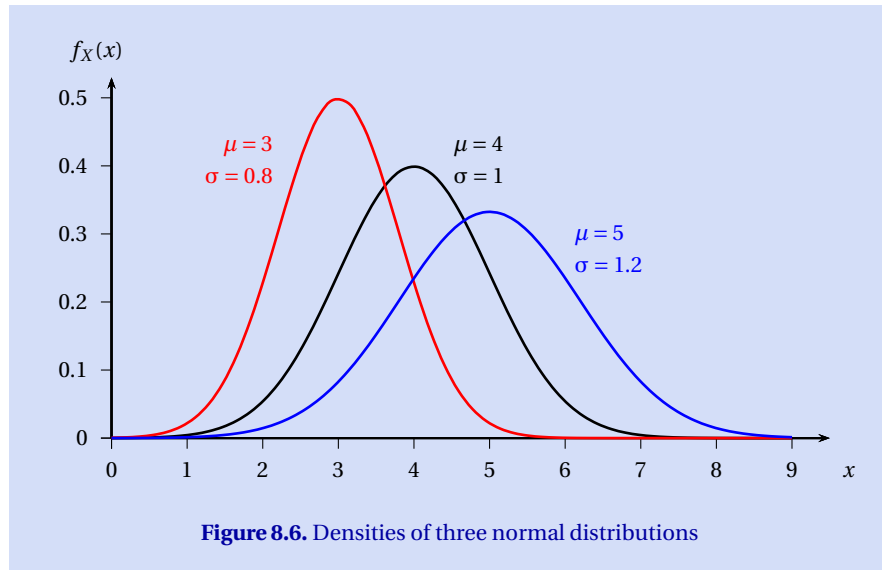
$$f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right), \quad \forall x \in \mathbb{R} \quad (8.24)$$

(see Fig. 8.7 and Exercise 8-4).

**Remark 8.32 (Distribution Function)** If  $X \sim \mathcal{N}_{\mu, \sigma^2}$ , then its *distribution function* is

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt, \quad \forall x \in \mathbb{R} \quad (8.25)$$

(see Cor. 5.95). If  $\mu = 0$  and  $\sigma^2 = 1$ , then this distribution function is often denoted by  $\Phi$  and in this case Equation (8.25) simplifies to



$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt, \quad \forall x \in \mathbb{R}. \quad (8.26)$$

Figure 8.7 displays the graphs of the density and the distribution function of the standard normal distribution. The shaded area is the Riemann integral of the density from  $-\infty$  to .60. Its value is  $\Phi(.60) \approx 0.7257$ , the value of the distribution function.  $\triangleleft$

**Remark 8.33 (Linear Functions of  $X$ )** If  $X \sim \mathcal{N}_{\mu, \sigma^2}$ , then  $\alpha + \beta X \sim \mathcal{N}_{\alpha + \beta\mu, \beta^2\sigma^2}$  (see Klenke, 2013, Exercise 1.5.3, p. 45).  $\triangleleft$

According to the central limit theorem, the limit (for  $n \rightarrow \infty$ ) of the sum of  $n$  independent identically distributed random variables  $X_1, \dots, X_n$  with positive and finite variances has a normal distribution.

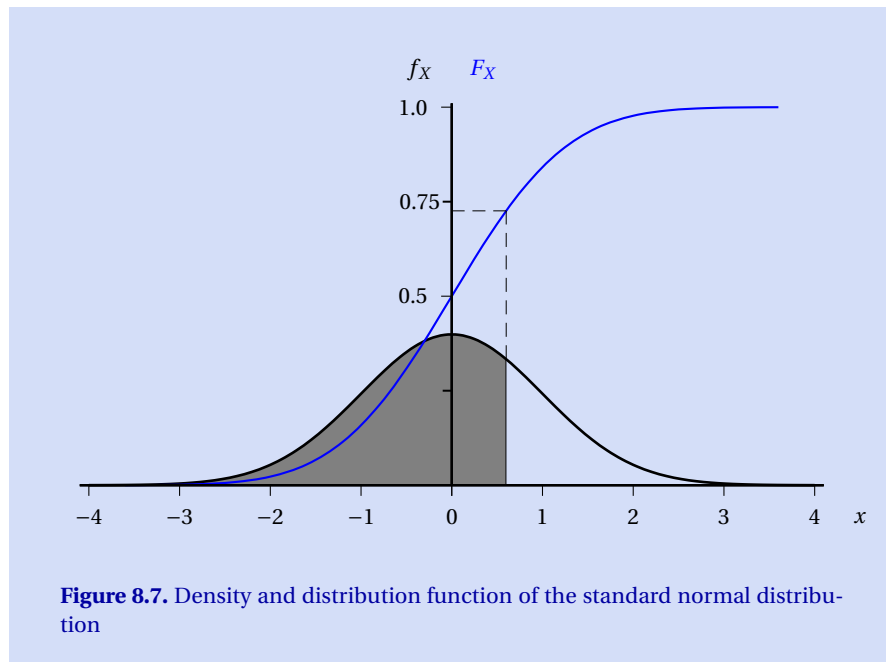
**Theorem 8.34 (Central Limit Theorem)**

Let  $X_i: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ ,  $i = 1, 2, \dots$ , be a sequence of real-valued i. i. d. random variables with finite expectations  $E(X_i) = \mu$  and finite variances  $\text{Var}(X_i) = \sigma^2 > 0$ . Furthermore, let  $\bar{X}_n := S_n/n$ , where  $S_n := \sum_{i=1}^n X_i$ , and

$$Z_n := \frac{S_n - n\mu}{\sigma \cdot \sqrt{n}} = \frac{(\bar{X}_n - \mu) \cdot \sqrt{n}}{\sigma}. \quad (8.27)$$

Then

$$\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z), \quad \forall z \in \mathbb{R}, \quad (8.28)$$



where  $\Phi(z)$  denotes the distribution function of the standard normal distribution.

For a proof see Georgii (2008, section 5.3). The second equation of (8.27) reveals that  $Z_n$  is the  $Z$ -transformation (see Rem. 6.33) of the sample mean  $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ , because  $\sigma/\sqrt{n}$  is the square root of the variance of  $\bar{X}_n$  (see Exercise 6-6).

**Remark 8.35 (Application of the Central Limit Theorem)** The central limit theorem can be applied for the approximation of the distributions of sums of independent identically distributed (i.i.d.) random variables. For large  $n$  and  $x \in \mathbb{R}$ ,

$$P\left(\sum_{i=1}^n X_i \leq x\right) \approx \Phi\left(\frac{x - n\mu}{\sigma \cdot \sqrt{n}}\right). \quad (8.29)$$

Note, if the  $X_i$  are discrete and integer-valued, then for all integers  $k$ ,

$$P\left(\sum_{i=1}^n X_i \leq k\right) = P\left(\sum_{i=1}^n X_i < k+1\right), \quad (8.30)$$

and therefore one may use the ‘correction’

$$P\left(\sum_{i=1}^n X_i \leq k\right) \approx \Phi\left(\frac{k + \frac{1}{2} - n\mu}{\sigma \cdot \sqrt{n}}\right). \quad (8.31)$$

Whether or not  $n$  is large enough for a good approximation, can be dealt with by the Berry-Esséen bound [see Georgii, 2008, Remark 5.30 (c), p. 138]. A rough summary is: The more symmetric the distribution of the  $X_i$  is, the faster the convergence of the limit in (8.28), i. e., small  $n$  already yield good approximations. For example, for  $n \geq 12$ , good approximations of the distribution function of  $Z_n$  are already obtained, for the symmetric distributions of  $X_i \sim \mathcal{U}_{[0,1]}$  and  $X_i \sim \mathcal{B}_{n,1/2}$ . In contrast, the approximation is still bad for the skewed (asymmetric) distribution of  $X_i \sim \mathcal{B}_{n,1/20}$  even if  $n = 50$ . In the latter case, the approximation of a binomial distribution by a Poisson distribution (see Remark 8.18) is better than by a normal distribution.  $\triangleleft$

### 8.2.3 Multivariate Normal Distribution

Now we present a generalization of the univariate normal distribution considering an  $n$ -dimensional real-valued random variable  $X = (X_1, \dots, X_n)$ . Reading the following definition remember that an  $n \times n$ -matrix  $\mathbf{A}$  is called *symmetric* if it is identical to its transpose, i. e., if  $\mathbf{A} = \mathbf{A}'$ , and that it is *positive definite*, if  $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$  for all column vectors  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{x} \neq \mathbf{0}$ .

#### Definition 8.36 (Multivariate Normal Distribution)

Let  $\boldsymbol{\mu}$  be an  $n$ -dimensional vector of real numbers and  $\boldsymbol{\Sigma}$  a symmetric and positive definite  $n \times n$ -matrix of real numbers. Furthermore, let  $X = (X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  be an  $n$ -dimensional random variable. If the function  $f_X: \mathbb{R}^n \rightarrow \mathbb{R}$  defined by

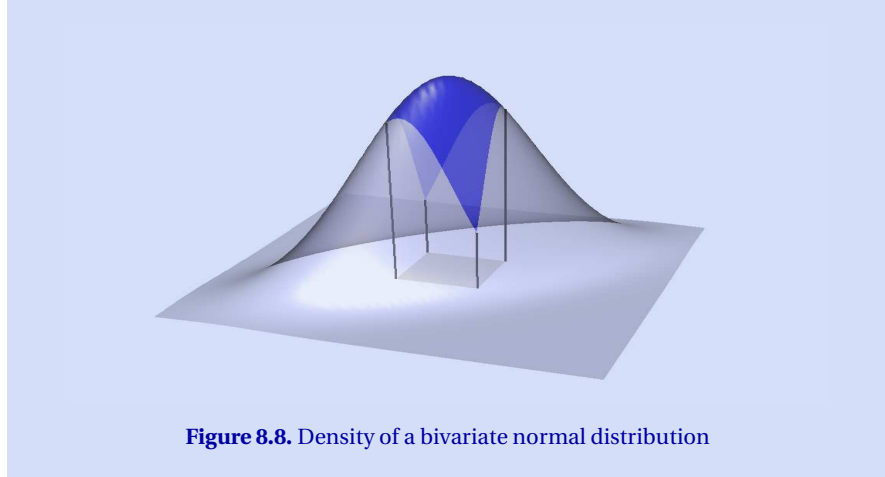
$$f_X(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \boldsymbol{\Sigma}}} \cdot \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad \forall \mathbf{x} \in \mathbb{R}^n \quad (8.32)$$

is a density of  $X$ , then we say that  $X$  has an  $n$ -variate (or  $n$ -dimensional) normal distribution with parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ , and we denote it by  $X \sim \mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ .

In section 7.4.2 we introduced the expectation of an  $n$ -dimensional random vector and in section 7.4.3 the variance-covariance matrix of an  $n$ -variate random variable  $X = (X_1, \dots, X_n)$ . These concepts are used in the following theorem showing that the parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  have a stochastic interpretation.

#### Theorem 8.37 (Expectation Vector and Covariance Matrix)

Let  $E(\mathbf{x})$  denote the expectation of the random vector  $\mathbf{x} = [X_1, \dots, X_n]'$  and  $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}$  the variance-covariance matrix of the  $n$ -variate random variable  $X = (X_1, \dots, X_n)$ . If  $X \sim \mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ , then  $E(\mathbf{x}) = \boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}} = \boldsymbol{\Sigma}$ .



**Figure 8.8.** Density of a bivariate normal distribution

For a proof see Georgii (2008, Th. 9.2, p. 242).

**Remark 8.38 (Univariate Normal Distribution)** The univariate normal distribution (see Def. 8.30) is in fact a special case of Definition 8.36. This can be seen as follows: If  $n = 1$  and  $X \sim \mathcal{N}_{\mu, \sigma^2}$ , then  $\boldsymbol{\mu} = [\mu] = [E(X)]$ ,  $\boldsymbol{\Sigma} = [\sigma^2]$ ,  $\det \boldsymbol{\Sigma} = \sigma^2$ , and  $\boldsymbol{\Sigma}^{-1} = \left[ \frac{1}{\sigma^2} \right]$ . Hence, the density (8.32) simplifies to (8.23).  $\triangleleft$

**Remark 8.39 (Bivariate Normal Distribution)** If  $n = 2$  and  $X = (X_1, X_2) \sim \mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ , then

$$\boldsymbol{\mu} = [\mu_1, \mu_2]' = E([X_1, X_2]') = [E(X_1), E(X_2)]' \quad (8.33)$$

and

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{xx} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \sigma_2 \\ \rho \cdot \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}, \quad (8.34)$$

where  $\rho = \text{Corr}(X_1, X_2)$ , and the density (8.32) can also be written

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{2(1-\rho^2)}\right), \quad (8.35)$$

where  $z_i := (x_i - \mu_i)/\sigma_i$  and  $\sigma_i = \sqrt{\sigma_i^2}$ , for  $i = 1, 2$  (see Fisz, 1963, section 5.11, p. 158). Figure 8.8 displays the density function of a bivariate normal distribution. The volume above the rectangle in the  $(x_1, x_2)$ -plane and the graph of the density is the probability that the bivariate random variable  $X = (X_1, X_2)$  takes on a value in that rectangle.  $\triangleleft$

**Remark 8.40 (Correlation and Independence)** If  $(X_1, X_2)$  has a bivariate normal distribution and  $\rho = 0$ , then Equation (8.35) simplifies to

$$\begin{aligned}
f_X(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} \cdot \exp\left[-\frac{1}{2}(z_1^2 + z_2^2)\right] \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}} \cdot \exp\left[-\frac{z_1^2}{2}\right] \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \cdot \exp\left[-\frac{z_2^2}{2}\right] \\
&= f_{X_1}(x_1) \cdot f_{X_2}(x_2).
\end{aligned} \tag{8.36}$$

Hence, this density is the product of the univariate normal densities of  $X_1$  and  $X_2$ , respectively [see Eq. (8.23)]. As shown in Exercise 8-5, this implies that  $X_1$  and  $X_2$  are independent if and only if  $\rho = 0$ . Hence, under bivariate normality,  $X_1$  and  $X_2$  are independent if and only if they are uncorrelated.  $\triangleleft$

**Theorem 8.41 (Linear Combinations)**

Let  $X = (X_1, \dots, X_n): (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$  be an  $n$ -variate real-valued random variable and  $\mathbf{x} = [X_1, \dots, X_n]'$ . Then  $X \sim \mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$  if and only if

$$\forall \mathbf{a} = [a_1, \dots, a_n]' \in \mathbb{R}^n: \mathbf{a}'\mathbf{x} \sim \mathcal{N}_{\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}. \tag{8.37}$$

For a proof and other characterizations of the multivariate normal distribution see Tong (1990, ch. 3).

**Remark 8.42 (Special Cases)** If  $X_1, \dots, X_n$  are independent random variables with  $X_i \sim \mathcal{N}_{\mu_i, \sigma_i^2}$ ,  $\mu_i \in \mathbb{R}$  and  $\sigma_i^2 > 0$ ,  $i = 1, \dots, n$ , then  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{x}}$  [see Eq. (7.36)] is diagonal and proposition (8.37) with  $\mathbf{a} = [1, \dots, 1]'$  yields

$$\sum_{i=1}^n X_i \sim \mathcal{N}_{\mu_s, \sigma_s^2}$$

with  $\mu_s = \sum_{i=1}^n \mu_i$  and  $\sigma_s^2 = \sum_{i=1}^n \sigma_i^2$ . In particular, if  $X_1, \dots, X_n$  is a sample, i. e., if  $X_1, \dots, X_n$  is a sequence of i. i. d. random variables with  $X_i \sim \mathcal{N}_{\mu, \sigma^2}$ ,  $i = 1, \dots, n$ ,  $\mu \in \mathbb{R}$ , and  $\sigma^2 > 0$ , then

$$\frac{(\bar{X}_n - \mu) \cdot \sqrt{n}}{\sigma} \sim \mathcal{N}_{0,1}, \tag{8.38}$$

where  $\bar{X}_n$  is the sample mean [see Eqs. (6.23) and (6.32)].  $\triangleleft$

**Theorem 8.43 (Linear Transformations)**

Let  $X = (X_1, \dots, X_n) \sim \mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$  and  $\mathbf{x} = [X_1, \dots, X_n]'$ . Furthermore, for  $m \leq n$ , let  $\mathbf{A}$  be an  $m \times n$ -matrix of real numbers of rank  $m$  and let  $\mathbf{c}$  be a column  $m$ -vector of real numbers. Then

$$\mathbf{y} := \mathbf{A}\mathbf{x} + \mathbf{c} \sim \mathcal{N}_{\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'}. \tag{8.39}$$

For a proof see Georgii (2008, Th. 9.5, pp. 243).

**Example 8.44 (Univariate Marginal Distribution)** For  $m = 1$ ,  $\mathbf{A} = [1, 0, \dots, 0]$ , and  $\mathbf{c} = [0]$ , Theorem 8.43 implies

$$X_1 \sim \mathcal{N}_{\mu_1, \sigma_1^2}, \quad (8.40)$$

where  $\mu_1$  is the first coordinate of  $\boldsymbol{\mu}$  and  $\sigma_1^2$  the first diagonal element of  $\boldsymbol{\Sigma}$ .  $\triangleleft$

**Example 8.45 (Bivariate Marginal Distribution)** For  $m = 2$ ,

$$\mathbf{A} = \begin{bmatrix} 1, 0, 0, \dots, 0 \\ 0, 1, 0, \dots, 0 \end{bmatrix},$$

and  $\mathbf{c} = [0, 0]'$ , Theorem 8.43 implies

$$\mathbf{Ax} := [X_1, X_2]' \sim \mathcal{N}_{\boldsymbol{\mu}_{12}, \boldsymbol{\Sigma}_{12}}, \quad (8.41)$$

where

$$\boldsymbol{\mu}_{12} := [\mu_1, \mu_2]' = E([X_1, X_2]') = [E(X_1), E(X_2)]'$$

and

$$\boldsymbol{\Sigma}_{12} := \boldsymbol{\Sigma}_{\mathbf{Ax}, \mathbf{Ax}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \sigma_2 \\ \rho \cdot \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix},$$

which is the variance-covariance matrix of  $(X_1, X_2)$ . [cf. Eqs. (8.33) and (8.34)].  $\triangleleft$

Examples 8.44 and 8.45 show that the appropriate choices of  $m$ ,  $\mathbf{A}$ , and  $\mathbf{c}$  immediately yield the following corollary:

**Corollary 8.46 (Marginal Distributions)**

*If  $X = (X_1, \dots, X_n) \sim \mathcal{N}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ , then all marginal distributions are normal. In particular,  $X_i \sim \mathcal{N}_{\mu_i, \sigma_i^2}$ ,  $i = 1, \dots, n$ , where  $\mu_i$  is the  $i$ th coordinate of  $\boldsymbol{\mu}$  and  $\sigma_i^2$  the  $i$ th diagonal element of  $\boldsymbol{\Sigma}$ .*

### 8.2.4 Central $\chi^2$ -Distribution

In the following definition  $\Gamma: \mathbb{R} \rightarrow \mathbb{R}$  denotes the *gamma function* defined by

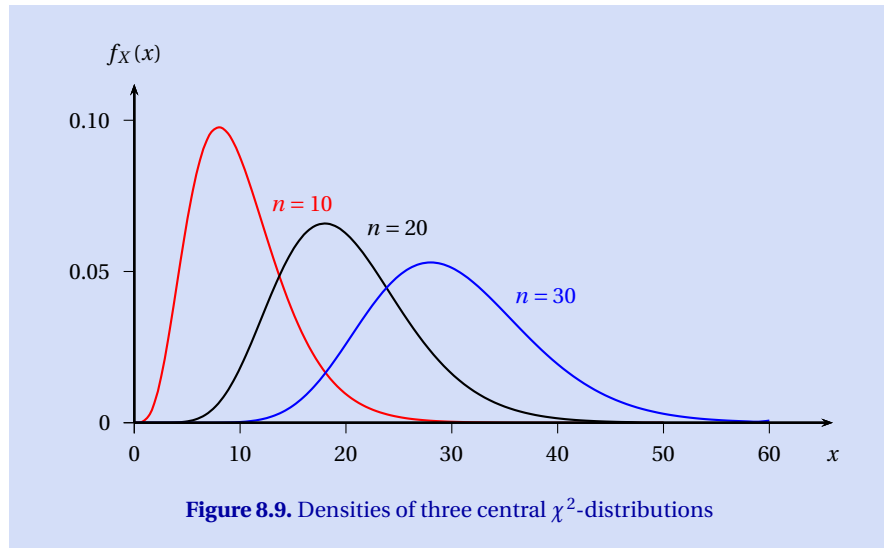
$$\Gamma(a) := \int_0^\infty t^{a-1} e^{-t} dt, \quad \forall a \in \mathbb{R}, \quad a > 0. \quad (8.42)$$

Note that,

$$\Gamma(a) = (a-1) \cdot \Gamma(a-1), \quad \text{for } a > 1. \quad (8.43)$$

Furthermore,

$$\Gamma(a) = (a-1)! \quad \text{for } a \in \mathbb{N}, \quad \text{and} \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \quad (8.44)$$

**Definition 8.47 (Central  $\chi^2$ -Distribution)**

Let  $n \in \mathbb{N}$ . A continuous nonnegative random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  has a central  $\chi^2$ -distribution with  $n$  degrees of freedom, abbreviated  $X \sim \chi_n^2$ , if  $X$  has a density satisfying

$$f_X(x) = \begin{cases} \frac{x^{n/2-1} \cdot e^{-x/2}}{2^{n/2} \cdot \Gamma(n/2)}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{cases} \quad \forall x \in \mathbb{R}. \quad (8.45)$$

Normal and  $\chi^2$ -distributions are related to each other as follows:

**Theorem 8.48 (Relationship Between Normal and  $\chi^2$ -Distributions)**

If  $X_1, \dots, X_n$  are i. i. d. random variables with standard normal distribution, then

$$X = \sum_{i=1}^n X_i^2 \quad (8.46)$$

has a central  $\chi^2$ -distribution with  $n$  degrees of freedom.

For a proof see Fisz (1963, section 9.4, pp. 339).

**Theorem 8.49 (Expectation and Variance)**

If  $X \sim \chi_n^2$ , then  $E(X) = n$  and  $\text{Var}(X) = 2n$ .

(Proof p. 270)

**Remark 8.50 ( $\chi^2$ -Distribution in Statistics)** Suppose that  $X_1, \dots, X_n$  is a sample of independent and normally distributed random variables with expectation  $E(X_i) = \mu$  and variance  $\text{Var}(X_i) = \sigma^2$ , for  $i = 1, \dots, n$ . Then

$$X := \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2, \quad (8.47)$$

and

$$Y := \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2, \quad (8.48)$$

(see Georgii, 2008, Th. 9.17, p. 249), where  $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$  is the *sample mean*.  $\triangleleft$

**8.2.5 Central  $t$ -Distribution**

In the following definition we again use the gamma function  $\Gamma$  defined by Equation (8.42).

**Definition 8.51 (Central  $t$ -Distribution)**

Let  $n \in \mathbb{N}$ . A continuous random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  has a *central  $t$ -distribution with  $n$  degrees of freedom*, denoted  $X \sim t_n$ , if  $X$  has a density satisfying

$$f_X(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \cdot \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad \forall x \in \mathbb{R}. \quad (8.49)$$

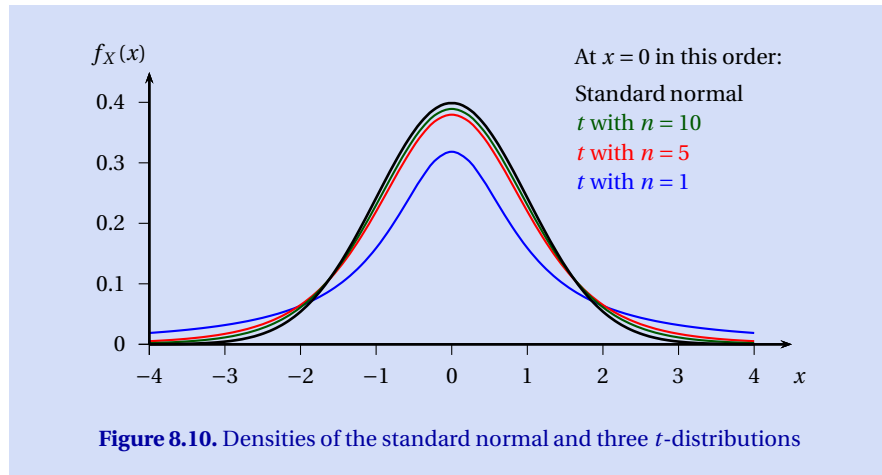
Figure 8.10 displays densities of three  $t$ -distributions with 1, 5, and 10 degrees of freedom, respectively.

**Theorem 8.52 (Expectation and Variance)**

If  $X \sim t_n$  and  $n > 1$ , then  $E(X) = 0$ , and if  $n > 2$ , then  $\text{Var}(X) = n/(n-2)$ .

For a proof see Johnson, Kotz, and Balakrishnan (1995, pp. 365, 366).

**Remark 8.53 (Cauchy Density)** If  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  has a  $t$ -distribution with  $n = 1$  degree of freedom, then the density  $f_X$  of  $X$  is also called the (*standard*) *Cauchy density*. In this case, the expectation of  $X$  does not exist (see Example 6.12).  $\triangleleft$



**Theorem 8.54 (Relationship Between  $t$ -, Normal and  $\chi^2$ -Distributions)**

If  $Z \sim \mathcal{N}_{0,1}$ ,  $Y \sim \chi_n^2$ , and  $Z$  and  $Y$  are independent, then

$$X := \frac{Z}{\sqrt{\frac{Y}{n}}} \quad (8.50)$$

has a  $t$ -distribution with  $n$  degrees of freedom.

For a proof see Fisz (1963, section 9.6, pp. 348) or Johnson et al. (1995, chapter 28, pp. 362).

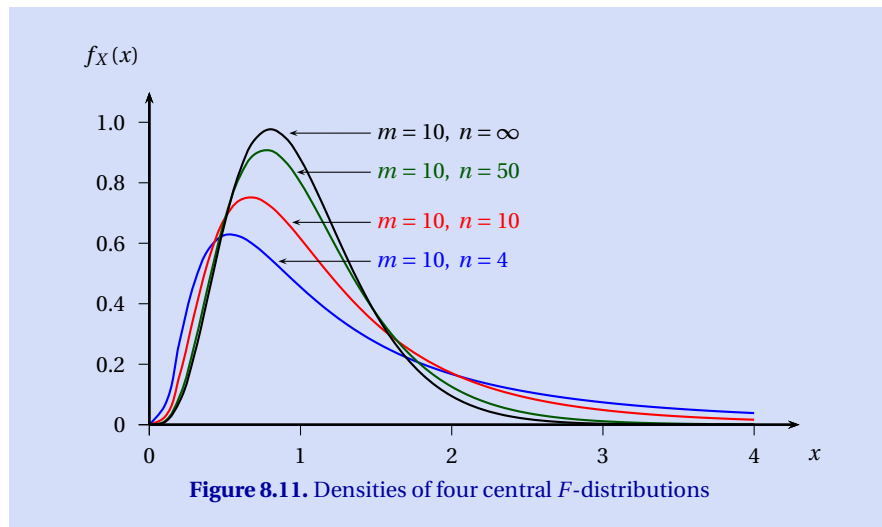
**Remark 8.55 (Convergence of Densities)** The sequence  $f_{X_n}$ ,  $n = 1, 2, \dots$ , of densities of the  $t$ -distributions with  $n$  degrees of freedom converges to the density of a standard normal distribution for  $n \rightarrow \infty$ . That is, if  $f_{X_n}$ ,  $n = 1, 2, \dots$ , satisfy Equation (8.49), then

$$\lim_{n \rightarrow \infty} f_{X_n}(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-x^2}{2}\right), \quad \forall x \in \mathbb{R} \quad (8.51)$$

(see Fig. 8.10 and Exercise 8-6). ◁

**Remark 8.56 ( $t$ -Distribution in Statistics)** Suppose that  $X_1, \dots, X_n$  are independent and normally distributed random variables with expectation  $E(X_i) = \mu$  and variance  $\text{Var}(X_i) = \sigma^2$ , for  $i = 1, \dots, n$ . Then

$$X := \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} \sim t_{n-1}, \quad (8.52)$$



where

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \text{and} \quad S := \sqrt{S^2}. \quad (8.53)$$

For the proof of independence of  $\bar{X}$  and  $S^2$  see Georgii (2008, Th. 9.17, p. 249). Applying Theorem 8.54, (8.48), and (8.38) then yield the result.  $\triangleleft$

### 8.2.6 Central $F$ -Distribution

In the following definition we again use the gamma function  $\Gamma$  defined by Equation (8.42).

#### **Definition 8.57 (Central $F$ -Distribution)**

Let  $m, n \in \mathbb{N}$  and let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be a continuous nonnegative random variable. Then  $X$  has a central  $F$ -distribution with  $m$  and  $n$  degrees of freedom, abbreviated  $X \sim F_{m,n}$ , if it has a density satisfying

$$f_X(x) = \begin{cases} \frac{\Gamma((m+n)/2) \cdot m^{m/2} \cdot n^{n/2} \cdot x^{m/2-1}}{\Gamma(m/2) \cdot \Gamma(n/2) \cdot (n+mx)^{(m+n)/2}}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{cases} \quad \forall x \in \mathbb{R}. \quad (8.54)$$

Figure 8.11 displays the densities of four  $F$ -distributions.

**Theorem 8.58 (Expectation and Variance)**

If  $X \sim F_{m,n}$ , then, for  $n \geq 3$ , the expectation of  $X$  is  $E(X) = n/(n-2)$ . For  $n \leq 2$ , the expectation does not exist. If  $n \geq 5$  the variance of  $X$  is

$$\text{Var}(X) = \frac{2n^2 \cdot (m+n-2)}{m \cdot (n-2)^2 \cdot (n-4)}. \quad (8.55)$$

For  $n \leq 4$ , the variance of  $X$  does not exist.

For a proof see Johnson et al. (1995, chapter 27, pp. 322).

**Theorem 8.59 (Relationship Between  $F$ - and  $\chi^2$ -Distributions)**

If  $Z \sim \chi_m^2$  and  $Y \sim \chi_n^2$  are independent, then

$$X := \frac{Z/m}{Y/n} \sim F_{m,n}. \quad (8.56)$$

For a proof see Fisz (1963, section 9.7, pp. 354).

**Remark 8.60 ( $F$ -Distribution in Statistics)** Let  $Z_1, \dots, Z_{n_1}, Y_1, \dots, Y_{n_2}$  be independent normally distributed random variables with expectations  $E(Z_i) = \mu_Z, E(Y_j) = \mu_Y$ , and variances  $\text{Var}(Z_i) = \sigma_Z^2, \text{Var}(Y_j) = \sigma_Y^2, i = 1, \dots, n_1, j = 1, \dots, n_2$ , respectively. If  $\sigma_Z^2 = \sigma_Y^2$ , then

$$X := \frac{S_Z^2}{S_Y^2} = \frac{\frac{1}{n_1-1} \sum_{i=1}^{n_1} (Z_i - \bar{Z})^2}{\frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2} \sim F_{n_1-1, n_2-1}, \quad (8.57)$$

where

$$\bar{Z} := \frac{1}{n_1} \sum_{i=1}^{n_1} Z_i \quad \text{and} \quad \bar{Y} := \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j \quad (8.58)$$

are the sample means [see (8.48) and Th. 8.59]. ◁

### 8.3 Proofs

**Proof of Theorem 8.9**

Let  $x \in \{0, 1, \dots, n\}$ . Then for any subset  $I_x$  of  $\{1, \dots, n\}$  that has  $x$  elements the assumptions of independence and identical distributions of the  $X_1, \dots, X_n$  imply

$$\begin{aligned} P(\forall i \in I_x: X_i = 1, \forall i \in \{1, \dots, n\} \setminus I_x: X_i = 0) &= \prod_{i \in I_x} P(X_i = 1) \cdot \prod_{i \in \{1, \dots, n\} \setminus I_x} [1 - P(X_i = 1)] \\ &= p^x (1-p)^{n-x}. \end{aligned}$$

As there are  $\binom{n}{x}$  subsets  $I_x$  of  $\{1, \dots, n\}$  with  $x$  elements, additivity of  $P$  implies

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

### ***Proof of Corollary 8.13***

According to Rule (x) of Box 4.1, the distribution  $P_X$  of  $X$ , which is a probability measure on  $(\Omega', \mathcal{A}')$ , is uniquely defined by the probability function  $p_X$  (see Def. 5.56). Furthermore, considering  $g(X) := X$  or  $g(X) := [X - E(X)]^2$ , Corollary 6.17 shows that the expectation and the variance of  $X$  solely depend on its distribution  $P_X$ . Therefore, it suffices to derive the expectation and the variance of the random variable  $X := \sum_{i=1}^n X_i$  on  $(\Omega, \mathcal{A}, P)$  defined in Theorem 8.9 with independent and identically distributed  $X_1, \dots, X_n$ . Hence, the expectation of  $X$  is

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^n X_i\right) && \text{[def. of } X\text{]} \\ &= \sum_{i=1}^n E(X_i) && \text{[Box 6.1 (vi)]} \\ &= \sum_{i=1}^n P(X_i=1) && \text{[(6.4), (6.5)]} \\ &= \sum_{i=1}^n p = np. && \text{[} P(X_i=1) = p \text{]} \end{aligned}$$

Similarly, the variance of  $X$  is

$$\begin{aligned} \text{Var}(X) &= \text{Var}\left(\sum_{i=1}^n X_i\right) && \text{[def. of } X\text{]} \\ &= \sum_{i=1}^n \text{Var}(X_i) && \text{[Box 6.2 (vi), } \perp\!\!\!\perp X_i\text{]} \\ &= \sum_{i=1}^n p(1-p) = np(1-p). && \text{[Example 6.30, } P(X_i=1) = p\text{]} \end{aligned}$$

### ***Proof of Theorem 8.16***

If  $X$  has a Poisson distribution with parameter  $\lambda$ , then

$$E(X) = \sum_{x=0}^{\infty} x \cdot p_X(x) = \sum_{x=1}^{\infty} x \cdot p_X(x) \quad [(6.6)]$$

$$= \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} \cdot e^{-\lambda} \quad [(8.11)]$$

$$= e^{-\lambda} \sum_{x=1}^{\infty} x \cdot \frac{\lambda \cdot \lambda^{x-1}}{x \cdot (x-1)!}$$

$$= e^{-\lambda} \lambda \cdot \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= e^{-\lambda} \lambda \cdot \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \lambda e^{\lambda} \quad \left[ e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \right]$$

$$= \lambda. \quad [e^{-\lambda} \cdot e^{\lambda} = e^{\lambda-\lambda}]$$

Because  $E(X^2) = E[X \cdot (X-1) + X] = E[X \cdot (X-1)] + E(X)$  and  $\text{Var}(X) = E(X^2) - E(X)^2$  [see Box 6.2 (i)], we consider

$$E[X \cdot (X-1)] = \sum_{x=0}^{\infty} x(x-1) \cdot p_X(x) \quad [(6.16)]$$

$$= \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} \cdot e^{-\lambda} \quad [(8.11)]$$

$$= e^{-\lambda} \sum_{x=2}^{\infty} x(x-1) \cdot \frac{\lambda^2 \cdot \lambda^{x-2}}{x \cdot (x-1) \cdot (x-2)!}$$

$$= e^{-\lambda} \lambda^2 \cdot \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!}$$

$$= e^{-\lambda} \lambda^2 \cdot \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \lambda^2 e^{\lambda} \quad \left[ e^{\lambda} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \right]$$

$$= \lambda^2. \quad [e^{-\lambda} \cdot e^{\lambda} = e^{\lambda-\lambda}]$$

Because  $E(X) = \lambda$ , we receive  $\text{Var}(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ .

### ***Proof of Theorem 8.22***

If  $X$  has a geometric distribution with parameter  $p$  and we define  $q := 1 - p$ , then

$$E(X) = \sum_{x=1}^{\infty} x p_X(x) \quad [(6.6)]$$

$$= \sum_{x=1}^{\infty} x q^{x-1} \cdot p \quad [(8.15)]$$

$$= p \cdot \sum_{x=1}^{\infty} x q^{x-1}$$

$$= p \cdot \sum_{x=1}^{\infty} \frac{d}{dq} q^x \quad \left[ \frac{d}{dq} q^x = x q^{x-1} \right]$$

$$= p \cdot \frac{d}{dq} \sum_{x=1}^{\infty} q^x$$

$$= p \cdot \frac{d}{dq} \left( \frac{1}{1-q} - 1 \right) \quad \left[ \sum_{x=1}^{\infty} q^x = \frac{1}{1-q} - 1 \right]$$

$$= p \cdot \frac{1}{(1-q)^2}$$

$$= p \cdot \frac{1}{p^2} = \frac{1}{p}. \quad [p = 1 - q]$$

In the fifth equation we used the fact that power series can be differentiated term-by-term within their radius of convergence.

The second moment  $E(X^2) = \sum_{x=1}^{\infty} x^2 p_X(x)$  can be calculated analogously to  $E(X)$ , except for using the second derivative with respect to  $q$ . The variance is then obtained by  $\text{Var}(X) = E(X^2) - E(X)^2$ .

### ***Proof of Theorem 8.23***

Let  $X_1, X_2, \dots$  be an infinite sequence of Bernoulli variables and  $X$  be defined by Equation (8.16). Then, for all  $x \in \mathbb{N}$ ,

$$\begin{aligned} p_X(x) &= P(X=x) = P(X_1=0, X_2=0, \dots, X_x=1) \\ &= \left[ \prod_{i=0}^{x-1} P(X_i=0) \right] \cdot P(X_x=1) \quad \left[ \prod_{i=1}^x X_i, (5.28) \right] \\ &= (1-p)^{x-1} \cdot p. \quad [(8.14)] \end{aligned}$$

According to Definition 8.20, this is the probability function of the geometric distribution.

### ***Proof of Theorem 8.49***

Let  $X_1, \dots, X_n$  be i.i.d. and  $X_i \sim \mathcal{N}_{0,1}$ ,  $i = 1, \dots, n$ , then  $E(X_1) = 0$ ,  $\text{Var}(X_1) = E(X_1^2) - E(X_1)^2 = E(X_1^2) = 1$ . We start calculating

$$E(X_1^4) = \int_{-\infty}^{\infty} x^4 f_{X_1}(x) dx \quad [(6.13)]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-x^2/2} dx \quad [(8.24)]$$

$$= \frac{1}{\sqrt{2\pi}} \left( -x^3 e^{-x^2/2} \Big|_{-\infty}^{\infty} + 3 \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \right) \quad [\text{integration by parts}]$$

$$= \frac{1}{\sqrt{2\pi}} \cdot (0 + 3 \cdot \sqrt{2\pi} \cdot E(X_1^2)) \quad [(6.22)]$$

$$= \frac{1}{\sqrt{2\pi}} \cdot (3 \cdot \sqrt{2\pi}) = 3.$$

Using this result we obtain

$$\text{Var}(X_1^2) = E(X_1^4) - E(X_1^2)^2 = 3 - 1 = 2.$$

If  $X \sim \chi_n^2$ , then this equation,  $P_X = P_{\sum_{i=1}^n X_i^2}$  [see (8.46)], Corollary 6.17,  $\frac{\perp}{p} X_1, \dots, X_n$ , Box 6.1 (vii), and Box 6.2 (vi) yield

$$E(X) = E\left(\sum_{i=1}^n X_i^2\right) = n \cdot E(X_1^2) = n$$

and

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^n X_i^2\right) = n \cdot \text{Var}(X_1^2) = 2n.$$

## 8.4 Exercises

▷ **Exercise 8-1** Consider Definition 8.4 and show that (8.6) holds.

▷ **Exercise 8-2** Consider the random variable  $X$  defined in Example 8.11 for  $n = 4$  trials. Determine the elements of  $\Omega = \{h, t\}^4$  that yield the value  $X = 3$ . Determine the probabilities for each  $\{\omega\}$ ,  $\omega \in \{X = 3\}$ . Identify the four paths in Figure 8.2 (p. 249) that lead from knot 1 to knot  $4p^3q$ . Also use this figure in order determine  $P(X=3)$ .

▷ **Exercise 8-3** Show that Equation (8.17) is the distribution function of  $X$  if  $X \sim \mathcal{G}_p$ .

▷ **Exercise 8-4** Show that the expectation of a random variable  $Z$  that has a standard normal distribution is 0.

▷ **Exercise 8-5** Use Corollary 5.100 to show that  $X_1$  and  $X_2$  are independent if and only if  $\rho = 0$ , provided that  $(X_1, X_2)$  has a bivariate normal distribution with density (8.35).

▷ **Exercise 8-6** Prove the proposition of Remark 8.55.

## Solutions

▷ **Solution 8-1**

$$\begin{aligned}
 p_X(x_1, \dots, x_n) &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\
 &= p^{\sum_{i=1}^n x_i} \cdot (1-p)^{\sum_{i=1}^n (1-x_i)} \\
 &= p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n - \sum_{i=1}^n x_i}.
 \end{aligned}$$

▷ **Solution 8-2** If we consider  $n = 4$  trials, then the value 3 of  $X$  is obtained by the  $\binom{4}{3} = 4$  outcomes:  $\omega_1 = (h, h, h, t)$ ,  $\omega_2 = (h, h, t, h)$ ,  $\omega_3 = (h, t, h, h)$ , and  $\omega_4 = (t, h, h, h)$ . The corresponding probabilities are  $P(\{\omega_1\}) = p^3 q$ ,  $P(\{\omega_2\}) = p^2 q p$ ,  $P(\{\omega_3\}) = p q p^2$ , and  $P(\{\omega_4\}) = q p^3$ . Summing these four probabilities yields  $P(X=3) = 4 \cdot p^3 q$ . The four paths from knot 1 to knot  $4p^3 q$  are:  $1 \rightarrow p \rightarrow p^2 \rightarrow p^3 \rightarrow 4p^3 q$ ,  $1 \rightarrow p \rightarrow p^2 \rightarrow 3qp^2 \rightarrow 4p^3 q$ ,  $1 \rightarrow p \rightarrow 2qp \rightarrow 3qp^2 \rightarrow 4p^3 q$ , and  $1 \rightarrow q \rightarrow 2qp \rightarrow 3qp^2 \rightarrow 4p^3 q$ . Using Figure 8.2 yields  $P(X=3) = 4p^3 q = \binom{4}{3} p^3 q^{4-3}$ . Because, in this example  $p = q = .5$ , this yields  $P(X=3) = 4p^3 q = 4 \cdot .5^3 \cdot .5 = .25$ .

▷ **Solution 8-3** For  $x \in \mathbb{N}$ ,

$$\begin{aligned}
 1 - F_X(x) &= P(X > x) && [(5.51)] \\
 &= P(\min\{n \in \mathbb{N} : X_n(\omega) = 1\} > x) && [\text{Th. 8.23}] \\
 &= P(X_1 = \dots = X_x = 0) \\
 &= P(X_1=0) \cdot \dots \cdot P(X_x=0) && [(5.28)] \\
 &= (1-p)^x. && [\text{Def. 8.19}]
 \end{aligned}$$

Because  $X$  is a discrete random variable with values in  $\mathbb{N}$ , its distribution function is a right-continuous step function with jumps at  $x \in \mathbb{N}$ , which yields Equation (8.17).

▷ **Solution 8-4** According to Theorem 6.11 and Equation (8.26) we have to show that the function

$$h(z) := z \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}, \quad (8.59)$$

is integrable and that its integral is 0. For these purposes we consider the positive and the negative parts of this function. The positive part is

$$h^+(z) := \begin{cases} 0, & \text{if } z < 0 \\ z \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}, & \text{if } z \geq 0, \end{cases}$$

and its integral is

$$\begin{aligned}
 \int_{-\infty}^{\infty} h^+(z) dz &= \frac{1}{\sqrt{2\pi}} \cdot \int_0^{\infty} (z \cdot e^{-z^2/2}) dz \\
 &= \frac{1}{\sqrt{2\pi}} \cdot \left( -e^{-z^2/2} \Big|_0^{\infty} \right) \\
 &= \frac{1}{\sqrt{2\pi}} \cdot [0 - (-1)] = \frac{1}{\sqrt{2\pi}}.
 \end{aligned}$$

The negative part is

$$h^-(z) := \begin{cases} -z \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}, & \text{if } z \leq 0 \\ 0 & \text{if } z > 0, \end{cases}$$

and its integral is

$$\begin{aligned} \int_{-\infty}^{\infty} h^-(z) dz &= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^0 (-z \cdot e^{-z^2/2}) dz \\ &= \frac{1}{\sqrt{2\pi}} \cdot \left( e^{-z^2/2} \Big|_{-\infty}^0 \right) \\ &= \frac{1}{\sqrt{2\pi}} \cdot (1 - 0) = \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

Hence, according to Definition 3.28, the function  $h(z)$  is integrable, and Equations (3.27) and (3.68) imply

$$\int_{-\infty}^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} - \frac{1}{\sqrt{2\pi}} = 0.$$

▷ **Solution 8-5** If  $\rho = 0$ , then Equation (8.36) and Corollary 5.100 imply that  $X_1$  and  $X_2$  are independent. If  $\rho \neq 0$ , then  $\text{Cov}(X_1, X_2) \neq 0$  [see (7.18)] and, by contraposition, Rule (vi) of Box 7.1 implies that  $X_1$  and  $X_2$  are not independent.

▷ **Solution 8-6**

$$\begin{aligned} \lim_{n \rightarrow \infty} \left( 1 + \frac{x^2}{n} \right)^{-\frac{n+1}{2}} &= \lim_{n \rightarrow \infty} \left[ \left( 1 + \frac{x^2}{n} \right)^n \right]^{-\frac{1}{2}} \left[ 1 + \frac{x^2}{n} \right]^{-\frac{1}{2}} \\ &= \left[ \lim_{n \rightarrow \infty} \left( 1 + \frac{x^2}{n} \right)^n \right]^{-\frac{1}{2}} \left[ \lim_{n \rightarrow \infty} \left( 1 + \frac{x^2}{n} \right) \right]^{-\frac{1}{2}} \\ &= \left( e^{x^2} \right)^{-1/2} \cdot 1^{-1/2} \\ &= e^{-x^2/2} \end{aligned}$$

(see Harris & Stocker, 1998, pp. 123, 124, 234).

Using the Stirling formula

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1$$

for all even  $n \in \mathbb{N}$ ,

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{\Gamma\left(\frac{n}{2} + \frac{1}{2}\right)}{\sqrt{\pi} \sqrt{n} \Gamma\left(\frac{n}{2}\right)} \\ &= \lim_{n \rightarrow \infty} \frac{1 \cdot 3 \cdot 5 \cdots (n-1) \sqrt{\pi}}{\sqrt{\pi} \sqrt{n} \left(\frac{n}{2} - 1\right)! 2^{n/2}} \quad [(8.43), (8.44)] \\ &= \lim_{n \rightarrow \infty} \frac{\frac{n}{2} \cdot n!}{\sqrt{n} \left(\frac{n}{2}\right)! 2^{n/2} 2^{n/2} \left(\frac{n}{2}\right)!} \quad \left[ 1 \cdot 3 \cdot 5 \cdots (n-1) = \frac{n!}{2^{n/2} \left(\frac{n}{2}\right)!} \right] \\ &= \lim_{n \rightarrow \infty} \frac{n \sqrt{2\pi n} n^n e^{-n}}{2 \sqrt{n} 2^n \sqrt{\pi n} \left(\frac{n}{2}\right)^{(n/2)} e^{-(n/2)} \sqrt{\pi n} \left(\frac{n}{2}\right)^{(n/2)} e^{-(n/2)}} \quad [\text{Stirling for } n!, (n/2)!] \\ &= \lim_{n \rightarrow \infty} \frac{n \sqrt{2\pi n} n^n e^{-n}}{2 \sqrt{n} 2^n \pi n n^n 2^{-n} e^{-n}} \\ &= \lim_{n \rightarrow \infty} \frac{\sqrt{2\pi}}{2\pi} = \frac{1}{\sqrt{2\pi}}. \end{aligned}$$

The proof for all odd  $n \in \mathbb{N}$  is analogous. Combining the two limits yields the proposition.



**Part III**  
**Conditional Expectation and Regression**



## Chapter 9

# Conditional Expectation Value and Discrete Conditional Expectation

In chapter 6 we introduced the concepts covariance and correlation, which quantify the strength of the kind of dependence that can be described by a linear quasi-regression. In the next five chapters we introduce the concept of a *conditional expectation* and a ‘genuine’ regression. These concepts can be used to describe how the  $(X=x)$ -conditional expectation values of a numerical random variable  $Y$  depend on the values of a (numerical, non-numerical, multivariate) random variable  $X$ . In this chapter we start with the concepts  $(X=x)$ -*conditional expectation value* and *discrete conditional expectation*, presuming that  $X$  is a discrete random variable. In this case, the conditional expectation  $E(Y|X)$  is easily defined as that random variable whose values are the conditional expectation values  $E(Y|X=x)$ . In chapter 10 we introduce the general concept of a conditional expectation, dropping the assumption that  $X$  is discrete. Chapters 12 and 13 deal with parametrizations of a conditional expectation, and Chapter 11 is devoted to the concepts *residual with respect to a conditional expectation*, *conditional variance*, *conditional covariance*, and partial correlation.

### 9.1 Conditional Expectation Value

Remember, the expectation of a numerical random variable  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  is defined by  $E(Y) = \int Y dP$ , using the probability measure  $P$ . Now we choose an event  $B \in \mathcal{A}$  with  $P(B) > 0$  and, instead of  $P$ , we consider the  $B$ -*conditional probability measure*  $P^B: \mathcal{A} \rightarrow [0, 1]$  defined by

$$P^B(A) := P(A|B), \quad \forall A \in \mathcal{A}, \quad (9.1)$$

(see Def. 4.29). Referring to this measure,

$$E^B(Y) := \int Y dP^B, \quad (9.2)$$

defines the  $P^B$ -*expectation of*  $Y$ , i. e., the expectation of  $Y$  with respect to the measure  $P^B$ . Reading the following definition, also remember that

$$\{X=x\} = X^{-1}(\{x\}) = \{\omega \in \Omega: X(\omega) = x\}$$

denotes the event that a random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  takes on the value  $x$  and that we defined  $P(X=x) := P(\{X=x\})$  (see Rem. 5.4). Assuming  $P(X=x) > 0$  and  $A \in \mathcal{A}$ , we denote

$$P(A|X=x) := P(A|\{X=x\}), \quad (9.3)$$

Furthermore, we denote  $P^{X=x} := P^{\{X=x\}}$ . Hence, according to Equation (9.1),

$$P^{X=x}(A) := P(A|X=x), \quad \forall A \in \mathcal{A}. \quad (9.4)$$

**Remark 9.1 ( $P^{X=x}$ -Equivalence of  $f(X)$  and  $f(x)$ )** Let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable,  $x \in \Omega'_X$ ,  $\{x\} \in \mathcal{A}'_X$ , and  $P(X=x) > 0$ . Furthermore, let  $f: (\Omega'_X, \mathcal{A}'_X) \rightarrow (\Omega', \mathcal{A}')$  be a measurable mapping, then  $f(X) \stackrel{P^{X=x}}{=} f(x)$  (see Exercise 9-1).  $\triangleleft$

The probability measures  $P^B$  and  $P^{X=x}$  are now used to define the conditional expectation value.

**Definition 9.2 (Conditional Expectation Value)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a random variable.

- (i) If  $B \in \mathcal{A}$  with  $P(B) > 0$  and  $Y$  is quasi-integrable with respect to  $P^B$ , then we define

$$E(Y|B) := E^B(Y) = \int Y dP^B, \quad (9.5)$$

call it the conditional expectation value of  $Y$  given the event  $B$  (or the  $B$ -conditional expectation value of  $Y$ ), and say that it exists.

- (ii) If  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  is a random variable,  $x \in \Omega'_X$  with  $\{x\} \in \mathcal{A}'_X$  and  $P(X=x) > 0$ , and  $Y$  is quasi-integrable with respect to  $P^{X=x}$ , then we define

$$E(Y|X=x) := E(Y|\{X=x\}) \quad (9.6)$$

and call it the conditional expectation value of  $Y$  given  $X=x$  (or the  $(X=x)$ -conditional expectation value of  $Y$ ), and say that it exists.

Note that  $E(Y|B)$  can be infinite. The only restriction is  $B \in \mathcal{A}$  with  $P(B) > 0$  and that  $Y$  is quasi-integrable with respect to  $P^B$ . Otherwise the integral  $\int Y dP^B$  is not defined.

**Remark 9.3 (Multivariate Numerical or Qualitative  $X$ )** Also note that the random variable  $X$  in Definition 9.2 (ii) can be numerical, i. e.,  $\Omega'_X \subset \overline{\mathbb{R}}$ , multivariate (see section 5.3), or nonnumerical. If  $X = (X_1, \dots, X_n)$  is a multivariate random variable, then we also use the notation  $E(Y|X_1=x_1, \dots, X_n=x_n)$  instead of  $E(Y|X=x)$  for the conditional expectation value of  $Y$  given  $(X_1, \dots, X_n) = (x_1, \dots, x_n)$  and call it the  $(X_1=x_1, \dots, X_n=x_n)$ -conditional expectation value of  $Y$  or the conditional expectation value of  $Y$  given  $X_1=x_1, \dots, X_n=x_n$ .  $\triangleleft$

The following theorem addresses the relationship between the  $B$ -conditional expectation value of  $Y$  and the expectation of  $1_B \cdot Y$  with respect to  $P$ .

**Theorem 9.4 (B-Conditional Expectation Value and the P-Expectation)**

Let the assumptions of Definition 9.2 (i) hold. Then:

- (i)  $E(Y)$  exists  $\Rightarrow E(Y|B)$  exists.
- (ii)  $E(Y)$  is finite  $\Rightarrow E(Y|B)$  is finite.
- (iii) Furthermore, if  $E(Y^2)$  is finite, then

$$E(Y|B) = \frac{1}{P(B)} \cdot \int 1_B \cdot Y \, dP = \frac{1}{P(B)} \cdot E(1_B \cdot Y) \quad (9.7)$$

$$= \frac{1}{P(B)} \cdot \text{Cov}(Y, 1_B) + E(Y). \quad (9.8)$$

(Proof p. 291)

Recall the following notation:

$$1_{X=x} := 1_{\{X=x\}}, \quad P(X=x) := P(\{X=x\}), \quad (9.9)$$

and

$$E^{X=x}(Y) := E^{\{X=x\}}(Y). \quad (9.10)$$

Using this notation, Equations (9.5), (9.6), and (9.7) yield the following corollary.

**Corollary 9.5 ((X=x)-Conditional Expectation Value and P-Expectation)**

If the assumptions of Definition 9.2 (ii) hold, then

$$\begin{aligned} E(Y|X=x) &= E^{X=x}(Y) = \int Y \, dP^{X=x} \\ &= \frac{1}{P(X=x)} \cdot \int 1_{X=x} \cdot Y \, dP = \frac{1}{P(X=x)} \cdot E(1_{X=x} \cdot Y). \end{aligned} \quad (9.11)$$

**Remark 9.6 (B-Conditional Probability)** If  $A \in \mathcal{A}$  and  $P(B) > 0$ , then

$$E(1_A|B) = \frac{1}{P(B)} \cdot E(1_B \cdot 1_A) \quad [(9.7)]$$

$$= \frac{1}{P(B)} \cdot P(A \cap B) \quad [\text{Box 6.1 (iv)}] \quad (9.12)$$

$$= P(A|B) = P^B(A). \quad [(4.2), (9.1)]$$

Because  $P^B$  is a probability measure, these equations imply  $0 \leq E(1_A|B) \leq 1$ .  $\triangleleft$

**Remark 9.7 ((X=x)-Conditional Probability)** For  $B = \{X=x\}$ , using the notation (9.3) and (9.4), Equation (9.12) implies

$$\begin{aligned} E(1_A|X=x) &= \frac{1}{P(X=x)} \cdot E(1_{X=x} \cdot 1_A) \\ &= \frac{1}{P(X=x)} \cdot P(A \cap \{X=x\}) \\ &= P(A|X=x) = P^{X=x}(A), \end{aligned} \quad (9.13)$$

provided that  $A \in \mathcal{A}$  and  $P(X=x) > 0$ . Equations (9.13) show that  $E(1_A | X=x)$  is identical to the conditional probability  $P(A | \{X=x\})$  of  $A$  given the event  $\{X=x\}$  (see Def. 4.12). The term  $P(A | X=x)$  is also called the *conditional probability of  $A$  given  $X=x$*  or the  *$(X=x)$ -conditional probability of  $A$* .  $\triangleleft$

**Remark 9.8 (( $X=x$ )-Conditional Probability of  $\{Y=y\}$ )** If  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_Y, \mathcal{A}'_Y)$  is a random variable and the assumptions of Definition 9.2 (ii) hold, then we also use the notation

$$P(Y=y | X=x) := P(\{Y=y\} | X=x) = P(1_{Y=y}=1 | X=x) = E(1_{Y=y} | X=x) \quad (9.14)$$

and call it the  *$(X=x)$ -conditional probability of  $\{Y=y\}$*  [see Eqs. (9.9)]. Hence, Equations (9.3), (9.14), and (4.2) yield

$$P(Y=y | X=x) = \frac{P(Y=y, X=x)}{P(X=x)}. \quad (9.15)$$

$\triangleleft$

## 9.2 Transformation Theorem

If  $P_Y^{X=x}: \mathcal{A}'_Y \rightarrow [0, 1]$  denotes the distribution of  $Y$  with respect to the  $(X=x)$ -conditional-probability measure  $P^{X=x}$  and  $E_Y^{X=x}(g)$  the expectation of  $g$  with respect to the distribution  $P_Y^{X=x}$ , then the transformation theorem (cf. Th. 6.13) for the conditional expectation value  $E(Y | X=x)$  can be formulated as follows:

### Theorem 9.9 (Transformation Theorem for $E(Y | X=x)$ )

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_Y, \mathcal{A}'_Y)$ ,  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be random variables, and  $g: (\Omega'_Y, \mathcal{A}'_Y) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a measurable function. Furthermore, let  $x \in \Omega'_X$  with  $\{x\} \in \mathcal{A}'_X$  and  $P(X=x) > 0$ .

(i) If  $g$  is nonnegative or with finite expectation  $E_Y^{X=x}(g)$ , then

$$\begin{aligned} E_Y^{X=x}(g) &= \int g \, dP_Y^{X=x} = \int g(y) P_Y^{X=x}(dy) \\ &= \int g(Y) \, dP^{X=x} = E^{X=x}[g(Y)] = E[g(Y) | X=x]. \end{aligned} \quad (9.16)$$

(ii)  $E_Y^{X=x}(g)$  is finite if and only if  $E^{X=x}[g(Y)]$  is finite.

There are two important points in Equations (9.16). *First*, these equations show the relationship between integrals of the composition  $g(Y)$  with respect to the conditional-probability measure  $P^{X=x}$  on  $(\Omega, \mathcal{A})$  on one side, and the distribution  $P_Y^{X=x}$  of  $Y$  with respect to  $P^{X=x}$  on the other side. *Second*,  $E[g(Y) | X=x]$  is identical to the expectation of  $g$  with respect to the distribution  $P_Y^{X=x}$ , i. e., the

distribution of  $Y$  with respect to the probability measure  $P^{X=x}$ . Thus, using the distribution  $P_{g(Y)}^{X=x}$  of  $g(Y)$  with respect to  $P^{X=x}$  is not necessary.

**Remark 9.10 (( $X=x$ )-Conditional Expectation Value of  $g(Y)$  and  $P$ -Expectation)** Equations (9.16) and Equations (9.11) imply

$$E[g(Y)|X=x] = \frac{1}{P(X=x)} \cdot \int 1_{X=x} \cdot g(Y) dP = \frac{1}{P(X=x)} \cdot E[1_{X=x} \cdot g(Y)]. \quad (9.17)$$

&lt;

**Remark 9.11 (A Special Case of the Transformation Theorem)** Let  $(\Omega'_Y, \mathcal{A}'_Y) = (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  and  $g$  be the identity function  $id: \bar{\mathbb{R}} \rightarrow \bar{\mathbb{R}}$ , defined by  $id(y) = y$  for all  $y \in \bar{\mathbb{R}}$ , which implies  $id(Y) = Y$ . If we assume that  $x \in \Omega'_X$  with  $\{x\} \in \mathcal{A}'_X$  and  $P(X=x) > 0$ , and that  $Y$  is nonnegative or with finite expectation  $E^{X=x}(Y)$ , then Equations (9.16) yield

$$E(Y|X=x) = E^{X=x}(Y) = \int Y dP^{X=x} = \int y P_Y^{X=x}(dy) = \int id dP_Y^{X=x}. \quad (9.18)$$

&lt;

Using the notation introduced in Equation (9.14), Equations (9.16), Theorem 9.9 (i), Equations (6.3) and (6.6) imply the following corollary.

**Corollary 9.12 ( $Y$  Discrete,  $g(Y)$  Real-Valued)**

Let the assumptions of Theorem 9.9 (i) hold and let  $g: (\Omega'_Y, \mathcal{A}'_Y) \rightarrow (\mathbb{R}, \mathcal{B})$ .

- (i) If  $Y$  is discrete and we assume that there is a finite set  $\{y_1, \dots, y_n\} \subset \Omega'_Y$ ,  $\{y_1\}, \dots, \{y_n\} \in \mathcal{A}'_Y$  with  $P_Y(\{y_1, \dots, y_n\}) = 1$ , then

$$E[g(Y)|X=x] = \sum_{i=1}^n g(y_i) \cdot P(Y=y_i|X=x). \quad (9.19)$$

- (ii) If  $Y$  is discrete and we assume that there is a countably infinite set  $\{y_1, y_2, \dots\} \subset \Omega'_Y$ ,  $\{y_i\} \in \mathcal{A}'_Y$ ,  $i \in \mathbb{N}$ , with  $P_Y(\{y_1, y_2, \dots\}) = 1$ , then

$$E[g(Y)|X=x] = \sum_{i=1}^{\infty} g(y_i) \cdot P(Y=y_i|X=x). \quad (9.20)$$

Note that, in this corollary,  $Y$  does not have to be real-valued or numerical. We only assume that  $g(Y)$  is real-valued. In contrast, in the following theorem we have to assume that  $Y$  itself is real-valued. Remember that  $\{y\} \in \bar{\mathcal{B}}$  if  $y \in \mathbb{R}$  [see (1.22)].

**Corollary 9.13 (Discrete and Real-Valued  $Y$ )**

Let the assumptions of Definition 9.2 (ii) hold.

- (i) If  $Y$  is discrete and there is a finite set  $\{y_1, \dots, y_n\} \subset \mathbb{R}$  of real numbers with  $P_Y(\{y_1, \dots, y_n\}) = 1$ . Then

$$E(Y|X=x) = \sum_{i=1}^n y_i \cdot P(Y=y_i | X=x). \quad (9.21)$$

- (ii) If  $Y$  is discrete and there is a countably infinite set  $\{y_1, y_2, \dots\} \subset \mathbb{R}$  of real numbers with  $P_Y(\{y_1, y_2, \dots\}) = 1$ , then

$$E(Y|X=x) = \sum_{i=1}^{\infty} y_i \cdot P(Y=y_i | X=x). \quad (9.22)$$

**9.3 Other Properties**

Because  $E(Y|B)$  is defined as the expectation  $E^B(Y)$  of  $Y$  with respect to the probability measure  $P^B$ , all properties of the expectation with respect to  $P$  can be translated to  $E(Y|B)$ , simply replacing  $P$  by  $P^B$  and  $E(Y)$  by  $E^B(Y) = E(Y|B)$ . Box 9.1 is such a translation of Box 6.1. Note that, according to Theorem 9.4 (i), the expectation  $E(Y|B)$  exists if  $E(Y)$  exists, provided that  $P(B) > 0$ . Of course, the rules for  $E(Y|B)$  also apply to the  $(X=x)$ -conditional expectation value  $E(Y|X=x)$  [see Def. 9.2 (ii)].

However, there are additional properties dealing with the relationship between the expectation and the conditional expectation value. Some of these have already been formulated in Theorem 9.4. Other additional properties are summarized in Box 9.2 and proved in Exercise 9-2.

Rule (ii) shows how the  $(X=x)$ -conditional expectation values  $E(Y|X=x)$  can be computed from the conditional expectation values  $E(Y|X=x, Z=z_i)$  and the conditional probabilities  $P(Z=z_i | X=x)$ . Hence, considering Equation (9.21) and Rule (ii) in Box 9.2 shows that we have two different equations for computing the conditional expectation value  $E(Y|X=x)$ . Finally, note that a special case of Rule (ii) is

$$E(Y) = \sum_{i=1}^m E(Y|Z=z_i) \cdot P(Z=z_i) \quad (9.23)$$

(see Exercise 9-3). According to this equation, we can also compute the expectation of  $Y$  from the conditional expectations  $E(Y|Z=z_i)$  and the probabilities  $P(Z=z_i)$ .

**Box 9.1 Rules of Computation for  $B$ -Conditional Expectation Values**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a random variable, let  $A, B, C \in \mathcal{A}$  with  $P(B) > 0$ , let the conditional expectation value  $E(Y|B)$  exist, and let  $\alpha \in \mathbb{R}$ . Then:

$$Y \stackrel{p^B}{=} \alpha \Rightarrow E(Y|B) = \alpha. \quad (\text{i})$$

$$E(\alpha + Y|B) = \alpha + E(Y|B). \quad (\text{ii})$$

$$E(\alpha \cdot Y|B) = \alpha \cdot E(Y|B). \quad (\text{iii})$$

$$E(1_A \cdot 1_C|B) = P(A \cap C|B). \quad (\text{iv})$$

For  $i = 1, \dots, n$ , let  $Y_i: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$  be random variables with finite  $B$ -conditional expectation values  $E(Y_i|B)$  and  $\alpha_i \in \mathbb{R}$ . Then

$$E\left(\sum_{i=1}^n \alpha_i \cdot Y_i \mid B\right) = \sum_{i=1}^n \alpha_i \cdot E(Y_i|B). \quad (\text{v})$$

Let  $X, Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be random variables that are nonnegative or with finite  $B$ -conditional expectation values. Then:

$$X \stackrel{p^B}{=} Y \Rightarrow E(X|B) = E(Y|B). \quad (\text{vi})$$

$$X \stackrel{p^B}{=} Y \Leftrightarrow \forall A \in \mathcal{A}: E(1_A X|B) = E(1_A Y|B). \quad (\text{vii})$$

$$X \stackrel{p^B}{\perp\!\!\!\perp} Y \Rightarrow E(X \cdot Y|B) = E(X|B) \cdot E(Y|B). \quad (\text{viii})$$

**9.4 Discrete Conditional Expectation**

The *discrete conditional expectation*  $E(Y|X)$  of a numerical random variable  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  given a random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  is now defined as that random variable on  $(\Omega, \mathcal{A}, P)$  whose values are identical to the conditional expectation values  $E(Y|X=x)$ . In this definition we have to assume that  $X$  is *discrete*, i. e., we assume that there is a finite or countable set  $\Omega'_0 \subset \Omega'_X$  such that  $P_X(\Omega'_0) = 1$  and  $P(X=x) > 0$  for all  $x \in \Omega'_0$  (see Def. 5.56). In chapter 10 this limitation is dropped.

**Definition 9.14 (Discrete Conditional Expectation)**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  be a numerical random variable that is nonnegative or has a finite expectation and let the random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be discrete.

- (i) If  $\{x_1, \dots, x_m\} = \Omega'_0 \subset \Omega'_X$  such that  $P_X(\Omega'_0) = 1$  and, for all  $i = 1, \dots, m$ ,  $\{x_i\} \in \mathcal{A}'_X$  and  $P(X=x_i) > 0$ , then the *discrete conditional expectation*

**Box 9.2 Rules of Computation for  $(X=x)$ -Conditional Expectation Values**

Let  $Y: (\Omega, \mathcal{A}, P) \rightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$  be a numerical random variable. Furthermore, let  $X: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_X, \mathcal{A}'_X)$  be a random variable, and let  $x \in \Omega'_X$  with  $\{x\} \in \mathcal{A}'_X$  and  $P(X=x) > 0$ . If  $E(Y|X=x)$  exists and  $f: (\Omega'_X, \mathcal{A}'_X) \rightarrow (\mathbb{R}, \mathcal{B})$  is a measurable function and  $E(Y^2), E[f(X)^2] < \infty$ , then

$$E[f(X) \cdot Y | X=x] = f(x) \cdot E(Y | X=x) = E[f(x) \cdot Y | X=x]. \quad (\text{i})$$

If  $Z: (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_Z, \mathcal{A}'_Z)$  is a random variable and  $z_1, \dots, z_m \in \Omega'_Z$  such that  $P_Z(\{z_1, \dots, z_m\}) = 1$  and, for all  $i = 1, \dots, m$ ,  $\{z_i\} \in \mathcal{A}'_Z$  and  $P(X=x, Z=z_i) > 0$ , then

$$E(Y | X=x) = \sum_{i=1}^m E(Y | X=x, Z=z_i) \cdot P(Z=z_i | X=x). \quad (\text{ii})$$

Correspondingly, if  $z_1, z_2, \dots \in \Omega'_Z$  such that  $P_Z(\{z_1, z_2, \dots\}) = 1$  and, for all  $i = 1, 2, \dots$ ,  $\{z_i\} \in \mathcal{A}'_Z$  and  $P(X=x, Z=z_i) > 0$ , then

$$E(Y | X=x) = \sum_{i=1}^{\infty} E(Y | X=x, Z=z_i) \cdot P(Z=z_i | X=x). \quad (\text{iii})$$

*tation of  $Y$  given  $X$  is defined by*

$$E(Y | X) := \sum_{i=1}^m E(Y | X=x_i) \cdot \mathbf{1}_{X=x_i}. \quad (9.24)$$

(ii) *If  $\{x_1, x_2, \dots\} = \Omega'_0 \subset \Omega'_X$  such that  $P_X(\Omega'_0) = 1$  and, for all  $i = 1, 2, \dots$ ,  $\{x_i\} \in \mathcal{A}'_X$  and  $P(X=x_i) > 0$ , then the discrete conditional expectation of  $Y$  given  $X$  is defined by*

$$E(Y | X) := \sum_{i=1}^{\infty} E(Y | X=x_i) \cdot \mathbf{1}_{X=x_i}. \quad (9.25)$$

Hence, in contrast to a conditional expectation value  $E(Y|X=x)$ , a discrete conditional expectation  $E(Y|X)$  is a discrete *random variable* (see Def. 5.56) on  $(\Omega, \mathcal{A}, P)$ . Note that  $E(Y|X)$  is a random variable taking a numerical value for each  $\omega \in \Omega$ . This means that we might look at its expectation, variance, covariance and correlation with other random variables (see, e.g., Box 10.2 and section 11.2).

**Remark 9.15 ( $X$ -Conditional Probability)** If  $A \in \mathcal{A}$ , then we use the notation

$$P(A|X) := E(\mathbf{1}_A | X) \quad (9.26)$$

and call it the *X*-conditional probability of *A*. If *Y* is dichotomous with values 0 and 1, we also use the notation  $P(Y=1|X)$  for the *X*-conditional probability of the event  $\{Y=1\}$ . If *Y* is dichotomous with values 0 and 1, Equations (9.21) and (9.26) then yield

$$P(Y=1|X) = E(Y|X). \quad (9.27)$$

◁

**Remark 9.16 (Uniqueness and Values of the Conditional Expectation)** An alternative way to write Equations (9.24) and (9.25) is

$$E(Y|X)(\omega) = \begin{cases} E(Y|X=x), & \text{if } \omega \in X^{-1}(\{x\}), \quad \forall x \in \Omega'_0, \\ 0, & \text{otherwise.} \end{cases} \quad (9.28)$$

Hence, the values of the conditional expectation  $E(Y|X)$  are uniquely defined by Equations (9.24) and (9.25) for all  $\omega \in \Omega$  (see Example 9.22). Assigning the value  $E(Y|X)(\omega) = 0$  if  $\omega \in \Omega \setminus X^{-1}(\Omega'_0)$  is arbitrary, but note that  $P(\Omega \setminus X^{-1}(\Omega'_0)) = 0$ . According to Equation (9.28) and Definition 5.56 this arbitrary assignment does not occur if  $P(X=x) > 0$  for all  $x \in X(\Omega)$ , i. e., if  $\Omega'_0 = X(\Omega)$ . ◁

## 9.5 Discrete Regression

**Remark 9.17 (Measurability and Factorization)** Definition 9.14 implies that the discrete conditional expectation  $E(Y|X)$  is a random variable on  $(\Omega, \mathcal{A}, P)$  that is measurable with respect to *X*. In more formal terms,  $E(Y|X): (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  and  $\sigma[E(Y|X)] \subset \sigma(X)$ . The reason is that there is a measurable function  $g: (\Omega'_X, \mathcal{A}'_X) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$  that is defined by

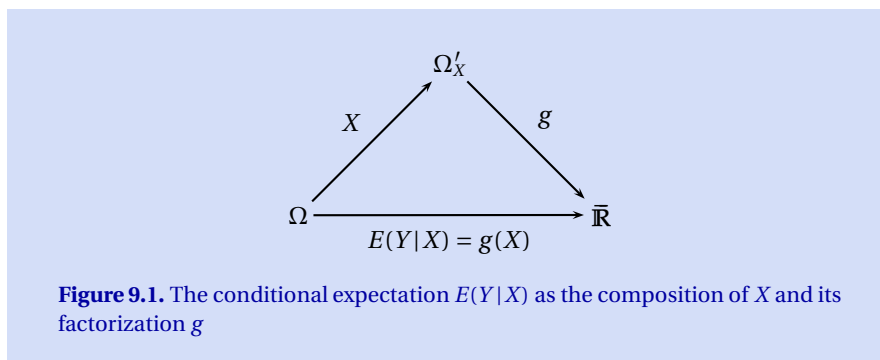
$$g(x) = \begin{cases} E(Y|X=x), & \forall x \in \Omega'_0, \\ 0, & \text{otherwise.} \end{cases} \quad (9.29)$$

Hence,  $E(Y|X) = g(X)$  (see Fig. 9.1) and Lemma 2.52 implies that  $E(Y|X)$  is measurable with respect to *X*. The function *g* is called the *factorization of E(Y|X)* or the *discrete regression of Y on X*. ◁

### Definition 9.18 (Discrete Regression)

Under the assumptions specified in Definition 9.14, the function  $g: \Omega'_X \rightarrow \mathbb{R}$  defined by Equation (9.29) is called the *discrete regression of Y on X*.

**Remark 9.19 (Regressand and Regressor)** Considering the conditional expectation  $E(Y|X)$ , we call *Y* the *regressand* and *X* the *regressor*. Whereas the regressand *Y* has to be numerical, the regressor *X* can be *any* random variable on  $(\Omega, \mathcal{A}, P)$  as long as it is discrete and satisfies  $P(X=x) > 0$  for all  $x \in \Omega'_0$ . ◁



**Figure 9.1.** The conditional expectation  $E(Y|X)$  as the composition of  $X$  and its factorization  $g$

**Table 9.1.** Joe and Ann With Random Assignment: Conditional Expectations

Elements of $\Omega$			Random variables			Conditional expectations		
Unit	Treatment	Success	Person variable $U$	Treatment variable $X$	Outcome variable $Y$	$E(Y X, U)$	$E(Y X)$	$P(X=1 U)$
(Joe, no, -)			Joe	0	0	.70	.45	.40
(Joe, no, +)			Joe	0	1	.70	.45	.40
(Joe, yes, -)			Joe	1	0	.80	.60	.40
(Joe, yes, +)			Joe	1	1	.80	.60	.40
(Ann, no, -)			Ann	0	0	.20	.45	.40
(Ann, no, +)			Ann	0	1	.20	.45	.40
(Ann, yes, -)			Ann	1	0	.40	.60	.40
(Ann, yes, +)			Ann	1	1	.40	.60	.40

**Remark 9.20 (Multivariate Numerical or Qualitative Regressors)** Note that the codomain  $\Omega'_X$  of  $X$  can be *any* set as long as there is a subset  $\Omega'_0 \subset \Omega'_X$  such that  $\Omega'_0$  is finite, or countable with  $P(X \in \Omega'_0) = 1$  and  $P(X=x) > 0$  for all  $x \in \Omega'_0$ . Hence, the *regressor*  $X$  can be uni- or multivariate (see Examples 9.21 and 9.22). If  $X = (X_1, \dots, X_n)$  is a discrete multivariate random variable, then we also use the notation  $E(Y|X_1, \dots, X_n)$  instead of  $E(Y|X)$  for the conditional expectation of  $Y$  given  $X$ . ◁

## 9.6 Examples

We treat two examples in some detail. Example 9.21 is straightforward, whereas Example 9.22 exemplifies that the values of a conditional expectation are uniquely defined by Equation (9.24) for all  $\omega \in \Omega$ .

**Example 9.21 (Joe and Ann With Random Assignment – continued)** Table 9.1 contains three conditional expectations we may consider in the example already used in Example 1.9. All of them are random variables taking a numerical value for each  $\omega \in \Omega$ . According to Remark 9.16, the values of  $E(Y|X)$  are the conditional expectation values  $E(Y|X=x)$  for all  $x \in \Omega'_0$ , and they are 0 for all  $x \in \Omega'_X \setminus \Omega'_0$ .

We start illustrating the conditional expectation of  $Y$  given  $X$ . Both random variables,  $X$  and  $Y$ , are specified in Table 9.1. We consider the random variable  $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ . In this case,  $\Omega'_X = \mathbb{R}$  and  $\Omega'_0 = \{0, 1\}$ . Because  $X$  takes on a value in  $\Omega'_0 = \{0, 1\}$  for *all*  $\omega \in \Omega$ , the conditional expectation  $E(Y|X)$  either takes on the value  $E(Y|X=0)$  or the value  $E(Y|X=1)$ . It does not take on the value 0, because  $\{X=x\} = \emptyset$  for all  $x \in \mathbb{R} \setminus \{0, 1\}$  [see Eq. (9.28)].

Because  $Y$  is an indicator,  $E(Y|X=x) = P(Y=1|X=x)$  [see Eqs. (9.14) and (9.27)]. Hence, if we want to compute the values of  $E(Y|X) = P(Y=1|X)$ , we have to compute the conditional probabilities  $P(Y=1|X=x)$ . For  $x=0$  we receive

$$P(Y=1|X=0) = \frac{P(Y=1, X=0)}{P(X=0)} = \frac{.21 + .06}{.09 + .21 + .24 + .06} = \frac{.27}{.60} = .45,$$

and for  $x=1$ ,

$$P(Y=1|X=1) = \frac{P(Y=1, X=1)}{P(X=1)} = \frac{.16 + .08}{.04 + .16 + .12 + .08} = \frac{.24}{.40} = .60.$$

Now we consider the conditional expectation  $E(Y|X, U)$ . The regressor is the random variable  $(X, U): (\Omega, \mathcal{A}, P) \rightarrow [\mathbb{R} \times \Omega_U, \mathcal{B} \otimes \mathcal{P}(\Omega_U)]$ , where  $\Omega_U = \{Joe, Ann\}$ , and  $\Omega'_0 = \{0, 1\} \times \Omega_U$ . Because,

$$\forall (x, u) \in (\mathbb{R} \times \Omega_U) \setminus (\{0, 1\} \times \Omega_U): \{(X=x, U=u)\} = \emptyset,$$

$E(Y|X, U)$  takes on a value in  $\mathbb{R}$  for *all*  $\omega \in \Omega$ . Furthermore, because  $Y$  is an indicator,  $E(Y|X, U) = P(Y=1|X, U)$ , and this conditional expectation has only four different values: the conditional probabilities  $P(Y=1|X=x, U=u)$ . For  $x=0$ ,  $u=Joe$  we receive

$$P(Y=1|X=0, U=Joe) = \frac{P(Y=1, X=0, U=Joe)}{P(X=0, U=Joe)} = \frac{.21}{.09 + .21} = .70,$$

for  $x=1$ ,  $u=Joe$ ,

$$P(Y=1|X=1, U=Joe) = \frac{P(Y=1, X=1, U=Joe)}{P(X=1, U=Joe)} = \frac{.16}{.04 + .16} = .80,$$

for  $x=0$ ,  $u=Ann$ ,

$$P(Y=1|X=0, U=Ann) = \frac{P(Y=1, X=0, U=Ann)}{P(X=0, U=Ann)} = \frac{.06}{.24 + .06} = .20,$$

and for  $x=1, u=Ann$ ,

$$P(Y=1|X=1, U=Ann) = \frac{P(Y=1, X=1, U=Ann)}{P(X=1, U=Ann)} = \frac{.08}{.12 + .08} = .40.$$

◁

**Example 9.22 (No Treatment for Joe)** Let us use a second example in order to illustrate the concepts introduced above. Again the random experiment consists of sampling a person, observing the value  $x$  of the treatment variable  $X$ , i. e., whether or not the sampled person receives a treatment ( $x=1$  vs.  $x=0$ ), and observing whether or not a success criterion is reached some time after treatment. In this new example, we fixed new probabilities of the elementary events. For instance, now the probability that Joe receives treatment is zero. This is useful to illustrate some general properties of conditional expectations. Also note that the probabilities of the other elementary events have been changed as well. The only restriction on the probabilities of the elementary events in such a hypothetical example is that they sum up to one.

Using the probabilities displayed in Table 9.2, Equation (9.21) yields:

$$\begin{aligned} E(Y|X=1) &= P(Y=1|X=1) = \frac{P(Y=1, X=1)}{P(X=1)} \\ &= \frac{0 + .152}{0 + 0 + .228 + .152} = .40 \end{aligned}$$

for the treatment condition  $x=1$ . Applying the corresponding formula to the control condition  $x=0$  yields  $E(Y|X=0) = (.348 + .024) / (.152 + .348 + .096 + .024) = .60$ . Note that the conditional probabilities  $P(Y=1|X=1)$  and  $P(Y=1|X=0)$  do *not* necessarily add up to 1. In contrast, the sum of  $P(Y=1|X=1)$  and  $P(Y=0|X=1)$  and the sum of  $P(Y=1|X=0)$  and  $P(Y=0|X=0)$  is always equal to 1, provided that  $Y$  is dichotomous with values 0 and 1.

Table 9.2 also displays the conditional probability  $P(X=1|U)$ , whose values are the treatment probabilities of Joe and Ann. For Joe, the treatment probability is  $P(X=1|U=Joe) = 0$ , and for Ann it is

$$P(X=1|U=Ann) = (.228 + .152) / (.096 + .024 + .228 + .152) = .76.$$

Finally, we compute the conditional expectations  $E(Y|X=x, U=u) = P(Y=1|X=x, U=u)$  [see Eq. (9.14)]:

$$\begin{aligned} P(Y=1|X=0, U=Joe) &= \frac{P(Y=1, X=0, U=Joe)}{P(X=0, U=Joe)} = \frac{.348}{.152 + .348} = .696, \\ P(Y=1|X=0, U=Ann) &= \frac{P(Y=1, X=0, U=Ann)}{P(X=0, U=Ann)} = \frac{.024}{.096 + .024} = .20, \end{aligned}$$

**Table 9.2.** No Treatment for Joe With Conditional Expectations

Elements of $\Omega$			Observables			Conditional expectations			
Unit	Treatment	Success	$P(\{\omega\})$	Person variable $U$	Treatment variable $X$	Outcome variable $Y$	$E(Y X,U)$	$E(Y X)$	$P(X=1 U)$
(Joe, no, -)			.152	Joe	0	0	.696	.60	0
(Joe, no, +)			.348	Joe	0	1	.696	.60	0
(Joe, yes, -)			0	Joe	1	0	0	.40	0
(Joe, yes, +)			0	Joe	1	1	0	.40	0
(Ann, no, -)			.096	Ann	0	0	.20	.60	.76
(Ann, no, +)			.024	Ann	0	1	.20	.60	.76
(Ann, yes, -)			.228	Ann	1	0	.40	.40	.76
(Ann, yes, +)			.152	Ann	1	1	.40	.40	.76

$$P(Y=1|X=1, U=Ann) = \frac{P(Y=1, X=1, U=Ann)}{P(X=1, U=Ann)} = \frac{.152}{.228 + .152} = .40.$$

Note that  $E(Y|X=1, U=Joe)$  is not defined, because  $P(X=1, U=Joe) = 0$  (see Def. 9.2). However, according to Definition 9.14 (i),  $E(Y|X,U)(\omega) = 0$  for  $\omega \in \Omega \setminus \Omega'_0 = \{(Joe, yes, -), (Joe, yes, +)\}$ . Thus  $E(Y|X,U)$  is uniquely defined, i.e., the values  $E(Y|X,U)(\omega)$  are uniquely defined for all  $\omega \in \Omega$ .  $\triangleleft$

**Example 9.23 (No Treatment for Joe – continued)** Using the results obtained in Example 9.22, Equation (9.24) yields:

$$\begin{aligned} E(Y|X) &= \sum_{x=0}^1 E(Y|X=x) \cdot 1_{X=x} \\ &= E(Y|X=0) \cdot 1_{X=0} + E(Y|X=1) \cdot 1_{X=1} \\ &= .60 \cdot 1_{X=0} + .40 \cdot 1_{X=1}. \end{aligned}$$

Hence, the values  $E(Y|X)(\omega)$  of the  $X$ -conditional expectation of  $Y$  are

$$E(Y|X)(\omega) = .60 \cdot 1_{X=0}(\omega) + .40 \cdot 1_{X=1}(\omega) = .60, \quad \text{if } X(\omega) = 0,$$

and

$$E(Y|X)(\omega) = .60 \cdot 1_{X=0}(\omega) + .40 \cdot 1_{X=1}(\omega) = .40, \quad \text{if } X(\omega) = 1.$$

These are the only two values that  $E(Y|X)$  takes on depending on the outcome  $\omega \in \Omega$  of the random experiment considered (see the first column of Table 9.2). This example illustrates that  $E(Y|X)$  is a random variable on  $(\Omega, \mathcal{A}, P)$  just like  $X$ ,

$Y$ , and  $U$ . Note again that the two values of  $E(Y|X) = P(Y=1|X)$  add up to 1 only by coincidence.

Table 9.2 shows two additional conditional expectations,  $E(X|U) = P(X=1|U)$  as well as  $E(Y|X, U) = P(Y=1|X, U)$ . Again using Equation (9.24) and the results obtained in Example 9.22, the conditional expectation  $P(X=1|U)$  is

$$\begin{aligned} P(X=1|U) &= \sum_u P(X=1|U=u) \cdot 1_{U=u} \\ &= P(X=1|U=Joe) \cdot 1_{U=Joe} + P(X=1|U=Ann) \cdot 1_{U=Ann} \\ &= .00 \cdot 1_{U=Joe} + .76 \cdot 1_{U=Ann} = .76 \cdot 1_{U=Ann}. \end{aligned}$$

Hence, the values  $P(X=1|U)(\omega)$  of the  $U$ -conditional expectation of  $X$  are

$$P(X=1|U)(\omega) = .00 \cdot 1_{U=Joe}(\omega) + .76 \cdot 1_{U=Ann}(\omega) = .00, \quad \text{if } U(\omega) = Joe$$

and

$$P(X=1|U)(\omega) = .00 \cdot 1_{U=Joe}(\omega) + .76 \cdot 1_{U=Ann}(\omega) = .76, \quad \text{if } U(\omega) = Ann.$$

These are the only two values that  $P(X=1|U)$  takes on. Again, Table 9.2 shows how the values of  $P(X=1|U)$  are assigned to the outcomes  $\omega \in \Omega$ .

Finally, let us turn to the conditional expectation  $E(Y|X, U)$ . Using the results obtained in Example 9.22, its values are

$$E(Y|X, U)(\omega) = E(Y|X=0, U=Joe) = .696, \quad \text{if } X(\omega) = 0 \text{ and } U(\omega) = Joe$$

$$E(Y|X, U)(\omega) = E(Y|X=0, U=Ann) = .20, \quad \text{if } X(\omega) = 0 \text{ and } U(\omega) = Ann$$

and

$$E(Y|X, U)(\omega) = E(Y|X=1, U=Ann) = .40, \quad \text{if } X(\omega) = 1 \text{ and } U(\omega) = Ann,$$

whereas

$$E(Y|X, U)(\omega) = 0, \quad \text{if } X(\omega) = 1 \text{ and } U(\omega) = Joe.$$

Note that the value  $E(Y|X, U)(\omega)$  is defined for  $\omega \in \{X=1, U=Joe\}$  although the conditional expectation value  $E(Y|X=1, U=Joe)$  is *not* defined. Also note that in this case the value  $E(Y|X, U)(\omega) = 0$  is arbitrarily fixed and has no substantive meaning. However, because  $P(X=1, U=Joe) = 0$ , this has no disadvantageous consequences. [In chapter 10, we show that the values of a conditional expectation can arbitrarily be fixed for all elements  $\omega$  of a subset  $A$  of  $\Omega$  for which  $P(A) = 0$ .]

According to Equation (9.24), the conditional expectation  $E(Y|X, U)$  is

$$\begin{aligned} E(Y|X, U) &= \sum_{(x,u)} E(Y|X=x, U=u) \cdot 1_{X=x, U=u} \\ &= .696 \cdot 1_{X=0, U=Joe} + .20 \cdot 1_{X=0, U=Ann} + .40 \cdot 1_{X=1, U=Ann}. \end{aligned} \tag{9.30}$$

The pair  $(1, Joe)$  is not an element of the set  $\Omega'_0$  (see Def. 9.14), and therefore the corresponding indicator  $1_{X=1, U=Joe}$  does not occur in this sum. Hence, if

$$\omega \in \{X=1, U=Joe\} = \{(Joe, yes, -), (Joe, yes, +)\},$$

then all three indicators occurring in Equation (9.30) take on the value 0, implying that  $E(Y|X, U)(\omega) = 0$  for these two elements of  $\Omega$ .  $\triangleleft$

## 9.7 Proofs

### Proof of Theorem 9.4

$$\begin{aligned} E(Y|B) &= E^B(Y) = \int Y dP^B && [(9.5)] \\ &= \int Y d\left(\frac{1}{P(B)} \cdot 1_B\right) \circ P && [\text{Lem. 4.32 (ii)}] \\ &= \frac{1}{P(B)} \cdot \int 1_B \cdot Y dP && [(3.72), (3.32)] \\ &= \frac{1}{P(B)} \cdot E(1_B \cdot Y), && [(6.1)] \end{aligned}$$

which yields Equation (9.7). Using Lemma 3.33 (i) yields Theorem 9.4 (i) and Lemma 3.33 (ii) implies Theorem 9.4 (ii).

If  $E(Y^2)$  is finite, then  $E(Y)$  and  $E(1_B \cdot Y)$  are finite as well [see Rem. 6.25 (iii) and Lemma 3.33 (ii)]. Therefore, according to Box 7.1 (i),  $Cov(1_B, Y) = E(1_B \cdot Y) - E(1_B) \cdot E(Y)$  is finite. Hence,

$$\begin{aligned} E(Y|B) &= \frac{1}{P(B)} \cdot E(1_B \cdot Y) && [(9.7)] \\ &= \frac{1}{P(B)} \cdot [Cov(1_B, Y) + E(1_B) \cdot E(Y)] && [\text{Box 7.1 (i)}] \\ &= \frac{1}{P(B)} \cdot Cov(1_B, Y) + E(Y). && [(6.4)] \end{aligned}$$

## 9.8 Exercises

- ▷ **Exercise 9-1** Prove the proposition of Remark 9.1
- ▷ **Exercise 9-2** Prove the rules of computation of conditional expectation values displayed in Box 9.2.
- ▷ **Exercise 9-3** Show that Equation (9.23) is a special case of Rule (ii) of Box 9.2.

## Solutions

▷ **Solution 9-1** It is sufficient to prove  $P^{X=x}(\{f(X) = f(x)\}^c) = 0$ , which is equivalent to  $P^{X=x}(\{f(X) = f(x)\}) = 1$ .

$$P^{X=x}[f(X) = f(x)] = \frac{P(f(X) = f(x), X=x)}{P(X=x)} \quad [(9.4), (9.15)]$$

$$\begin{aligned}
&= \frac{P(X=x)}{P(X=x)} && \{X=x\} \subset \{f(X)=f(x)\}, (9.9) \\
&= 1.
\end{aligned}$$

▷ **Solution 9-2**

$$\begin{aligned}
\text{(i)} \quad E[f(X) \cdot Y | X=x] &= \frac{1}{P(X=x)} \cdot E[1_{X=x} \cdot f(X) \cdot Y] && [(9.11)] \\
&= \frac{1}{P(X=x)} \cdot E[1_{X=x} \cdot f(x) \cdot Y] && [1_{X=x} \cdot f(X) = 1_{X=x} \cdot f(x)] \\
&= f(x) \cdot \frac{1}{P(X=x)} \cdot E(1_{X=x} \cdot Y) && [\text{Box 6.1 (iii)}] \\
&= f(x) \cdot E(Y | X=x) && [(9.7)] \\
&= E[f(x) \cdot Y | X=x]. && [\text{Box 9.1 (iii)}]
\end{aligned}$$

$$\begin{aligned}
\text{(ii)} \quad E(Y | X=x) &= \frac{1}{P(X=x)} \cdot E(1_{X=x} \cdot Y) && [(9.11)] \\
&= \frac{1}{P(X=x)} \cdot E\left(1_{X=x} \cdot \sum_{i=1}^m 1_{Z=z_i} \cdot Y\right) && \left[1 \stackrel{p}{=} \sum_{i=1}^m 1_{Z=z_i}, (5.33), \text{Rem. 6.27}\right] \\
&= \sum_{i=1}^m \frac{1}{P(X=x)} \cdot E(1_{X=x} \cdot 1_{Z=z_i} \cdot Y) && [\text{Box 6.1 (vii)}] \\
&= \sum_{i=1}^m \frac{P(X=x, Z=z_i)}{P(X=x)} \cdot \frac{1}{P(X=x, Z=z_i)} \cdot E(1_{X=x, Z=z_i} \cdot Y) \\
&= \sum_{i=1}^m E(Y | X=x, Z=z_i) \cdot P(Z=z_i | X=x). && [(9.11), (9.15)]
\end{aligned}$$

$$\begin{aligned}
\text{(iii)} \quad &\sum_{i=1}^{\infty} E(Y | X=x, Z=z_i) \cdot P(Z=z_i | X=x) \\
&= \sum_{i=1}^{\infty} P(Z=z_i | X=x) \cdot \frac{1}{P(X=x, Z=z_i)} \cdot E(1_{X=x, Z=z_i} \cdot Y) && [(9.11)] \\
&= \sum_{i=1}^{\infty} \frac{1}{P(X=x)} \cdot E(1_{X=x, Z=z_i} \cdot Y) && [(9.15)] \\
&= \frac{1}{P(X=x)} \sum_{i=1}^{\infty} \int 1_{X=x, Z=z_i} \cdot Y \, dP && [\text{Def. 6.1}] \\
&= \frac{1}{P(X=x)} \sum_{i=1}^{\infty} \left[ \int 1_{Z=z_i} \cdot 1_{X=x} \cdot Y^+ \, dP - \int 1_{Z=z_i} \cdot 1_{X=x} \cdot Y^- \, dP \right] && [\text{Def. 3.28}] \\
&= \frac{1}{P(X=x)} \left[ \sum_{i=1}^{\infty} \int 1_{Z=z_i} \cdot 1_{X=x} \cdot Y^+ \, dP - \sum_{i=1}^{\infty} \int 1_{Z=z_i} \cdot 1_{X=x} \cdot Y^- \, dP \right] \\
& && [1_{X=x} \cdot Y \text{ is quasi-integrable}] \\
&= \frac{1}{P(X=x)} \left[ \int 1_{X=x} \cdot Y^+ \, dP - \int 1_{X=x} \cdot Y^- \, dP \right] && \left[ (3.65), 1 \stackrel{p}{=} \sum_{i=1}^{\infty} 1_{Z=z_i}, (5.33), (3.44) \right]
\end{aligned}$$

$$\begin{aligned} &= \frac{1}{P(X=x)} \int 1_{X=x} \cdot Y \, dP && \text{[Def. 3.28]} \\ &= E(Y|X=x). && \text{[(9.11)]} \end{aligned}$$

▷ **Solution 9-3** This is easily seen considering the special case  $X=\alpha$ ,  $\alpha \in \Omega'_X$ , i. e., the case in which  $X$  is a constant. Then  $\{X=\alpha\} = \Omega$ , and Equation (9.11) yields  $E(Y|X=\alpha) = E(Y)$ , and  $E(Y|X=\alpha, Z=z_i) = E(Y|Z=z_i)$ . Hence, Rule (ii) of Box 9.2 yields

$$E(Y) = \sum_{i=1}^m E(Y|Z=z_i) \cdot P(Z=z_i).$$



## References

- Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.
- Bauer, H. (1996). *Probability theory*. Berlin, Germany and New York, NY: de Gruyter.
- Bauer, H. (2001). *Measure and integration theory*. Berlin, Germany and New York, NY: de Gruyter.
- Billingsley, P. (1995). *Probability and measure* (3rd ed.). New York, NY: Wiley-Interscience.
- Bronshstein, I. N., Semendyayev, K. A., Musiol, G., & Mühlig, H. (2015). *Handbook of mathematics* (6th ed.). New York, NY: Springer.
- Ellis, R., & Gulick, D. (2006). *Calculus* (6<sup>th</sup> ed.). Mason, OH: Thomson.
- Elstrodt, J. (2007). *Maß- und Integrationstheorie [Measure and integration theory]* (4th ed.). Berlin, Germany: Springer.
- Fisz, M. (1963). *Probability theory and mathematical statistics*. New York, NY: Wiley.
- Fristedt, B., & Gray, L. (1997). *A modern approach to probability theory*. Boston, MA: Birkhäuser.
- Georgii, H.-O. (2008). *Stochastics – Introduction to probability and statistics*. Berlin, Germany: de Gruyter.
- Harris, J. W., & Stocker, H. (1998). *Handbook of mathematics and computational science*. New York, NY: Springer.
- Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics* (Vol. 1). New York, NY: Chapman & Hall.
- Horn, R. A., & Johnson, C. R. (1991). *Matrix analysis*. Cambridge, UK: Cambridge University Press.
- Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (3rd ed.). New York, NY: Wiley.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 2). New York, NY: Wiley.
- Kheyfits, A. (2010). *A primer in combinatorics*. Berlin/New York: De Gruyter.
- Klenke, A. (2013). *Probability theory – A comprehensive course* (2nd ed.). London, England: Springer. doi: 10.1007/978-1-4471-5361-0
- Kolmogoroff, A. N. (1933/1977). *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Reprinted ed.). Berlin, Germany: Springer.
- Kolmogorov, A. N. (1933/1977). *Grundbegriffe der Wahrscheinlichkeitsrechnung*

- [*Foundations of the Theory of Probability*] (Reprinted ed.). Berlin, Germany: Springer.
- Kolmogorov, A. N. (1956). *Foundations of the theory of probability* (2nd ed.; N. Morrison, Trans.). New York, NY: Chelsea.
- Mayer, A., Thoemmes, F., Rose, N., Steyer, R., & West, S. G. (2014). Theory and analysis of total, direct, and indirect causal effects. *Multivariate Behavioral Research*, 49(5), 425–442.
- McCullagh, P., & Nelder, J. A. (1989). *Monographs on statistics and applied probability: Vol. 37. Generalized linear models* (2nd ed.; D. R. Cox, D. V. Hinkley, N. Reid, D. B. Rubin, & D. V. Silverman, Eds.). Chapman & Hall.
- Michel, H. (1978). *Maß- und Integrationstheorie I. [Measure and integration theory I]*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Rao, C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). New York, NY: Wiley.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago, IL: University of Chicago. (Originally published by Nissen & Lydicke, 1960. Kopenhagen.)
- Rosen, K. (2012). *Discrete mathematics and its applications* (7<sup>th</sup> ed.). New York, NY: McGraw-Hill.
- Steyer, R. (2003). *Wahrscheinlichkeit und Regression [Probability and regression]*. Berlin, Germany: Springer.
- Steyer, R., Mayer, A., & Fiege, C. (2014). Causal inference on total, direct, and indirect effects. In Michalos, A. C. (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 606–631). Dordrecht, Netherlands: Springer.
- Süli, E., & Mayers, D. F. (2003). *An introduction to numerical analysis*. Cambridge: Cambridge University Press.
- Tong, Y. L. (1990). *The multivariate normal distribution*. New York, NY: Springer.
- Wise, G. L., & Hall, E. B. (1993). *Counterexamples in probability and real analysis*. New York: Oxford University Press.

## Author Index

- Agresti, A., 382, 388, 541
- Balakrishnan, N., 264, 541
- Bauer, H., VI, 16, 24, 91–93, 98, 101, 102, 104, 106, 107, 109, 133, 174, 206, 251, 298, 319, 326, 426, 485, 498, 523, 541
- Billingsley, P., VI, 541
- Bronshtein, I. N., VIII, 541
- Ellis, R., VIII, 104, 188, 239, 394, 514, 541
- Elstrodt, J., 106, 541
- Fiege, C., 133, 542
- Fisz, M., 260, 263, 265, 267, 514, 541
- Fristedt, B., VI, 541
- Georgii, H.-O., 4, 7, 12, 145, 256, 258–260, 262, 264, 266, 541
- Gray, L., VI, 541
- Gulick, D., VIII, 104, 188, 239, 394, 514, 541
- Hall, E. B., 7, 23, 542
- Harris, J. W., VIII, 273, 541
- Hoffmann-Jørgensen, J., VI, 541
- Horn, R. A., 373, 541
- Johnson, C. R., 373, 541
- Johnson, N. L., 253, 264, 265, 267, 541
- Kemp, A. W., 253, 541
- Kheyfits, A., 8, 541
- Klenke, A., VI, 12, 13, 15, 22, 25, 27, 50, 51, 59, 62, 64, 80, 86, 90, 104, 105, 109, 110, 133, 160, 217, 252, 257, 306, 307, 325, 378, 498, 522, 541
- Kolmogoroff, A. N., V, VI, 541
- Kolmogorov, A. N., 127, 295, 541, 542
- Kotz, S., 253, 264, 541
- Mayer, A., 133, 542
- Mayers, D. F., 374, 542
- McCullagh, P., 382, 388, 542
- Michel, H., 11, 542
- Mühlig, H., VIII, 541
- Musiol, G., VIII, 541
- Nelder, J. A., 382, 388, 542
- Rao, C. R., 370, 542
- Rasch, G., 392, 542
- Rose, N., 133, 542
- Rosen, K., VII, VIII, 542
- Semendyayev, K. A., VIII, 541
- Steyer, R., VI, 133, 152, 542
- Stocker, H., VIII, 273, 541
- Süli, E., 374, 542
- Thoemmes, F., 133, 542
- Tong, Y. L., 261, 542
- West, S. G., 133, 542
- Wise, G. L., 7, 23, 542

# Subject Index

- absolute continuity, 109
  - and  $P$ -equivalence, 164
  - and independence, 172
  - of a conditional-probability measure, 141
  - of marginal distributions, 181
- absolute value function, 63
- additivity of a measure
  - $\sigma$ -additivity, 17
  - finite additivity, 17
- adjusted conditional effect function, 444
- adjusted effect, 444
- adjusted logit effect, 448
- almost all, 161
- almost everywhere, 100
- average effect, 444
  
- Bayes' theorem
  - for densities, 516
  - for events, 139
- Bernoulli distribution, 246
- Bernoulli variable, 246
- binomial coefficient, 247
- binomial distribution, 247
  - approximation by Poisson distribution, 251
- bivariate normal distribution, 260
- Borel  $\sigma$ -algebra on  $\mathbb{R}$ , 12
- Borel  $\sigma$ -algebra on  $\mathbb{R}^n$ , 13, 15
- Borel set, 12
  
- cardinality, 8, 21
- Cartesian product, 3
- Cauchy distribution, 264
- Cauchy-Schwarz inequality, 223
- central  $F$ -distribution, 266
- central  $t$ -distribution, 264
  
- central limit theorem, 257
- central moment, 208
- $\chi^2$ -distribution, 262
- closed interval, 6
- CNS-uniqueness, 499
- codomain of a mapping, 41
- coefficient of determination
  - of a conditional expectation, 334
  - of a linear quasi-regression, 218, 232
- coefficient of variation, 211
- common null set uniqueness, 499
- complement of a set, 5
- composition of two mappings, 58
- conditional correlation
  - given a value of a random variable, 346
- conditional covariance
  - given a  $\sigma$ -algebra, 338
  - given a random variable, 339
  - given a value of a random variable, 340
  - rules of computation, 343
- conditional density, 515, 517
- conditional distribution, 492
  - and  $P$ -equivalence, 502
  - and conditional independence, 508
  - and conditional independence given a random variable, 507
  - and independence, 497
  - and mean independence, 509
  - existence, 498
  - of a random variable given a  $\sigma$ -algebra, 492
  - of a random variable given a random variable, 492

- of a random variable given a value
    - of a random variable, 494
- conditional distribution function, 515
- conditional effect function, 439
- conditional expectation
  - w.r.t. a conditional-probability measure, 406–408
  - and conditional densities, 516, 517
  - and joint distribution, 312
  - and mean-squared error, 307
  - coefficient of determination, 334
  - convergence, 305, 306
  - discrete, 283
  - given a  $\sigma$ -algebra, 296
  - given a random variable, 296
  - linear parametrization, 360
  - marginalization, 512
  - monotonicity, 305
  - rules of computation, 303, 304
  - uniqueness w.r.t. a probability measure, 298
- conditional expectation value
  - and conditional distribution, 511
  - given a value of a random variable, 278, 309
  - given an event, 278
  - of a composition, 280
  - rules of computation, 283, 284
- conditional independence
  - of random variables given an event, 174
- conditional independence given a  $\sigma$ -algebra
  - and conditional mean independence, 470
  - characterizations, 472
  - family of  $\cap$ -stable set systems, 474
  - family of events, 473
  - family of random variables, 475
  - family of set systems, 474
  - notation, 465
  - of two events, 460
  - of two random variables, 462
  - of two set systems, 461
  - properties, 469
- conditional independence given a random variable
  - and conditional distributions, 506
  - family of events, 473
  - family of random variables, 475
  - family of set systems, 474
  - of two events, 460
  - of two random variables, 462
  - of two set systems, 462
  - properties, 470
- conditional independence given an event
  - family of  $\cap$ -stable set systems, 148
  - family of set systems, 146
  - of two events, 145
- conditional intercept function, 439
- conditional logit effect function, 448
- conditional logit intercept function, 448
- conditional mean independence, 313
  - and conditional independence, 470
  - and independence, 314
  - from a  $\sigma$ -algebra, 313
  - from random variable, 313
- conditional probability
  - w.r.t. a conditional-probability measure, 407, 408
  - given a  $\sigma$ -algebra, 296
  - given a random variable, 296
  - given a value of a random variable, 309
  - given an event, 132
  - of an event given a value of a random variable, 279
  - of an event given an event, 279
- conditional probability density, *see* conditional density
- conditional standard deviation
  - given a  $\sigma$ -algebra, 339
  - given a random variable, 339
  - given a value of a random variable, 340
- conditional variance
  - given a  $\sigma$ -algebra, 339
  - given a random variable, 339
  - given a value of a random variable, 340
  - rules of computation, 344
- conditional-probability measure, 140, 503
- constant mapping, 48
- continuity of a measure from above, 25
- continuity of a measure from below, 25
- continuous random variable, 186

- without expectation, 202
- continuous uniform distribution, 130, 254
- convergence of conditional expectations, 305, 306
- convergence of integrals, 103
- correlation, 225
  - and slope of a linear quasi-regression, 226
  - invariance under linear transformations, 225
- countable intersection, 6
- countable union, 5
- countably generated  $\sigma$ -algebra, 11
- counting measure, 21
- covariance, 221
  - rules of computation, 222
- covariance matrix, 230
  - rules of computation, 231
- density, *see also* probability density 186
  - and probability function, 180
  - of the  $\chi^2$ -distribution, 263
  - of the  $F$ -distribution, 266
  - of the  $t$ -distribution, 264
  - of the bivariate normal distribution, 260
  - of the Cauchy distribution, 264
  - of the multivariate normal distribution, 259
  - of the standard normal distribution, 108
  - of the univariate normal distribution, 256
- density of a measure w.r. t. another measure, 106
- dichotomous function, 48
- dichotomous random variable, 158
- Dirac measure, 21
- discrete conditional expectation given a random variable, 283
- discrete conditional probability given a random variable, 284
- discrete distribution, 175
- discrete random variable, 175
  - without expectation, 201
- discrete regression, 285
- discrete uniform distribution, 245
- disjoint sets, 4
- distribution, 156
  - and distribution function, 182, 185
  - Bernoulli, 246
  - binomial, 247
  - $\chi^2$ , 262
  - Cauchy, 264
  - discrete uniform, 245
  - $F$ , 266
  - geometric, 252
  - multivariate normal, 259
  - of a composition, 157
  - of an indicator, 157
  - Poisson, 250
  - $t$ , 264
  - univariate normal, 256
  - univariate standard normal, 256
- distribution function, 182
  - and distribution, 182, 185
  - and independence, 185
  - joint and marginal distribution function, 185
  - of a binomial distribution, 247
  - of the geometric distribution, 253
  - of the Poisson distribution, 250
  - of the standard normal distribution, 256
  - of the univariate normal distribution, 256
- domain of a mapping, 41
- dominated convergence of integrals, 104
- Dynkin system, 16
- effect function, 439
- elementary event, 128
- elementary function, 83
- equivalence w.r. t. a measure, 65
  - necessary and sufficient conditions, 107
- equivalence w.r. t. a probability measure of random variables, 161
  - of two factorizations of a conditional expectation, 308
- equivalence class w.r. t. a measure, 66
- equivalence of probability densities w.r. t. a measure, 179
- equivalence relation, 65
- event, 127
- existence
  - of a conditional distribution, 498

- of a conditional expectation given a  $\sigma$ -algebra, 297
  - of an expectation, 199
- expectation
  - w.r.t.  $P$ , 199
  - w.r.t. a conditional-probability measure, 200, 279
  - of a distribution, 204
  - of a random matrix, 228
  - of a random variable, 199
  - of a random variable with a countable number of real values, 201
  - of a random variable with a finite number of real values, 200
  - of a random variable with density, 202
  - of a random vector, 228
  - of a sample mean, 206
  - of an indicator, 200
  - of the product of random variables under independence, 206
  - of the product of two random variables, 221
  - rules of computation, 207
- expectation of a random matrix
  - rules of computation, 229
- factorial of an integer, 247
- factorization
  - equivalence of two versions w.r.t. a probability measure, 308
  - of a conditional expectation given a random variable, 307
  - uniqueness, 308
- factorization of a composition, 59
- $F$ -distribution, 266
- filtration, 134
- final  $\sigma$ -algebra, 54
- finite additivity of a measure, 17
- finite intersection, 6
- finite measure, 23
- finite union, 5
- Fubini's theorem, 110
- $\Gamma$ -function, 262
- generating system
  - of a  $\sigma$ -algebra, 9
  - of a product  $\sigma$ -algebra, 16
- geometric distribution, 252
- distribution function, 253
- identically distributed random variables, 157
- identification
  - linear logit regression, 388
  - linear regression, 368
- identity
  - of mappings, 46
  - of random variables, 161
- identity mapping, 48
- i. i. d., 171
- image measure, 68
  - under a step function, 69
- image of a set under a mapping, 41, 43
- increasing sequence of nonnegative step functions, 91
- independence
  - and  $P$ -equivalence, 172
  - and absolute continuity, 172
  - and conditional mean independence, 314
  - and distribution function, 185
  - and probability densities, 187
  - and product measure, 172
  - conditional, *see* conditional independence
  - family of  $\sigma$ -algebras, 144
  - family of events, 143
  - family of random variables, 171
  - family of set systems, 144
  - measurable mappings, 174
  - of  $\cap$ -stable set systems, 145
  - of  $n$  random variables, 171
  - of a constant and a set of events, 173
  - of a random variable and a set system, 171
  - of an event and a set system, 144
  - of three events, 144
  - of two events, 143
  - of two random variables, 169
- independent and identically distributed, 171
- indicator, 20
- infinite measure, 23
- integrable, 94
- integral
  - w.r.t. a Dirac measure, 92

- w.r.t. a finite weighted sum of measures, 101
- w.r.t. a measure with density, 107
- w.r.t. a weighted sum of Dirac measures, 101
- w.r.t. a weighted sum of measures, 101
- w.r.t. an image measure, 102
- w.r.t. the Lebesgue measure, 94
- of  $\mu$ -equivalent functions, 99
- of a constant, 86
- of a function with a finite number of values, 103
- of a measurable function, 94
- of a nonnegative measurable function, 92
- of a nonnegative step function, 86
- of a positive measurable function, 98
- over a null set, 98
- over a subset, 86, 96
- over the union of two sets, 97
- intercept
  - of a simple linear quasi-regression, 220
  - of a simple linear regression, 362
- intercept function, 439
- intersection
  - of countably many sets  $A_1, A_2, \dots$ , 6
  - of finitely many sets  $A_1, \dots, A_n$ , 6
- invariance of regression coefficients, 375
- inverse image of a set under a mapping, 41
- joint distribution, 168
- joint distribution function, 184
- Lebesgue integral and Riemann integral, 104
- Lebesgue measure, 22
- linear combination of two functions, 62
- linear logistic regression, 388
- linear logit parametrization, 384
  - identification, 388
- linear logit regression, 388
- linear parametrization
  - differences between means as coefficients, 371
  - identification, 368
  - means as coefficients, 370
- linear parametrization of a conditional expectation, 360
- linear quasi-regression, 217
  - and linear regression, 366
  - and regression, 308
  - equivalent characterizations, 223
- linear regression, 369
  - and linear quasi-regression, 366
  - and normal distribution, 370
  - identification, 368
- linearity
  - of the integral, 96
- link function, 382
- log odds, 386
- log odds ratio, 386
- log-odds, 382
- logit effect function, 448
- logit intercept function, 448
- logit of a conditional probability, 382
- logit transformation, 382
- marginal density, 181
- marginal distribution, 168
- marginal distribution function, 515
- marginal probability density, 515
- marginalization of a conditional expectation, 512
- mean centered random variable, 209
- mean independence, 313
  - and conditional distributions, 509
  - and uncorrelatedness, 318
- mean squared error, 217, 232
- measurability w.r.t. a mapping, 56
- measurable mapping, 46
- measurable set, 5
- measurable space, 6
- measure, 17
- measure space, 17
- measure with density, 106
- mixture of probability measures, 132
- moment, 208
- moment of a numerical random variable, 208
- monotonicity
  - of a conditional expectation, 305
  - of a measure, 19
  - of an integral, 93, 100
- MSE*, mean squared error function, 217

- $\mu$ -almost everywhere, 100
- multiple correlation, 336
- multiple linear logistic regression, 389
- multiple linear quasi-regression, 232
- multiple linear regression, 369
- multiplication rule for probabilities, 135
- multivariate Bernoulli distribution, 247
- multivariate Bernoulli variable, 246
- multivariate mapping, 54
- multivariate normal distribution, 259
- multivariate random variable, 167
  
- negative part of a function, 62
- nonnegative step function, 83
  - and counting measure, 88
  - and Dirac measure, 87
  - normal representation, 84
- normal distribution, 256
  - and linear regression, 370
  - bivariate, 260
  - density, 256
  - multivariate, 259
  - standard, 256
  - univariate, 256
- normal representation of a nonnegative step function, 84
- null set, 28
  - integral over a null set, 98
- null-set equivalence of two measures, 109
  
- odds ratio, 387
- outcome of a random experiment, 128
  
- pairwise independence, 143
- $P$ -almost all, 161
- partial  $(X, Z=z)$ -conditional expectation, 412
- partial correlation, 345
- partition of a set, 10
- $P$ -equivalence, 161
  - and absolute continuity, 164
  - and conditional distributions, 502
  - and distributions, 163
  - and independence, 172
  - of compositions, 163
  - of random variables, 161
- $P^B$ -expectation, 200
- pointwise convergence, 91
- Poisson distribution, 250
  - approximation of the binomial distribution, 251
  - distribution function, 250
- positive part of a function, 62
- power set, 6
- probability density, *see also* density 179
  - of a continuous real-valued random variable, 186
  - of a probability measure, 179
  - of a random variable, 179
- probability function, 175
  - and density, 180
  - of a marginal distribution, 177
- probability measure, 127
  - with density, 178
- probability of an event, 127
- probability space, 127
- product  $\sigma$ -algebra, 14, 56
- product measure, 24
  - and independence, 172
- product set, 3
- projection mapping, 56
- $P$ -uniqueness, 298
  - of a conditional expectation, 298
  - of a conditional expectation w.r.t. conditional-probability measure, 421
- $P^B$ -uniqueness, 411
  
- quantile, 183
- quantile function, 183
  - and inverse distribution function, 183
- quasi-integrable, 94
- quasi-regression
  - linear, 217
  
- Radon-Nikodym derivative, 109
- Radon-Nikodym theorem, 109
  - and probability density, 179
- random sample, 171
- random variable, 155
  - continuous, 186
  - discrete, 175
  - discrete real-valued, 176
  - numerical, 156
  - real-valued, 156
- regressand, 285, 307
- regression, 307

- and linear quasi-regression, 308
  - discrete, 285
- regression coefficients, 369
  - invariance, 375
- regressor, 285, 307
- residual
  - w.r.t. a conditional expectation, 330
  - w.r.t. a linear quasi-regression, 223, 232
- residual w.r.t. a conditional expectation
  - rules of computation, 332
- restriction of a measure, 23
- Riemann integral and Lebesgue integral, 104
- risk ratio, 387
- rules of computation
  - for a residual w.r.t. a conditional expectation, 332
  - for conditional covariances, 343
  - for conditional expectation values given a value of a random variable, 284
  - for conditional expectation values given an event, 283
  - for conditional expectations given a  $\sigma$ -algebra, 303
  - for conditional expectations given a random variable, 304
  - for conditional variances, 344
  - for covariance matrices, 231
  - for covariances, 222
  - for expectations of random matrices, 229
  - for measures, 19
  - for probabilities, 129
  - for the expectation of a random variable, 207
  - for variances, 209
- set system, 4
- $\sigma$ -additivity of a measure, 17
- $\sigma$ -algebra, 5
  - Borel, 12
  - countably generated, 11
  - final, 54
  - generated by a composition, 59
  - generated by a family of mappings, 55
  - generated by a mapping, 51
  - generated by a multivariate mapping, 54
  - generated by a set system, 9
  - generated by an indicator, 52
  - trivial, 28
- $\sigma$ -field, *see*  $\sigma$ -algebra
- $\sigma$ -finite measure, 23
- $\sigma$ -subadditivity, 19
- sign function, 64
- simple function, 83
- simple linear logistic regression, 389
- simple linear regression, 362, 369
  - identification of the intercept, 370
  - identification of the slope, 370
  - intercept, 362
  - slope, 362
- singleton, 6
- skewness, 212
- slope
  - of a linear quasi-regression and correlation, 226
  - of a simple linear quasi-regression, 220
  - of a simple linear regression, 362
- stability of a set system w.r.t. intersections,  $\cap$ -stability, 15
- standard deviation, 208
  - of the sample mean, 210
- standard error
  - of the sample mean, 210
- standard normal distribution, 256
- step function, 49
  - normal representation, 84
- strictly diagonally dominant matrix, 372
- $t$ -distribution, 264
- theorem
  - B. Levi, 103
  - Bayes' theorem for densities, 516
  - Bayes' theorem for events, 139
  - central limit, 257
  - Fubini, 110
  - of total probability, 137
  - Radon-Nikodym, 109
  - transformation theorem for a conditional expectation value, 280
  - transformation theorem for an integral, 102

- transformation theorem of an expectation, 203
- time order
  - between events, 134
  - between random variables, 160
- trace  $\sigma$ -algebra, 8
- trace of a set system, 8
- transformation theorem
  - for a conditional expectation value, 280
  - for an expectation, 203
  - for an integral, 102
  - for conditional distributions, 511
- triple-wise independence, 144
- trivial  $\sigma$ -algebra, 28
- trivial  $\sigma$ -algebra w.r.t. a measure, 28
  
- uncorrelated random variables, 221
- uncorrelatedness
  - mean independence, 318
- uncountable union, 6
- uniform distribution, 254
- union
  - of countably many sets, 5
  - of finitely many sets, 5
- uniqueness
  - of a conditional expectation w.r.t. a probability measure, 298
  - of a conditional expectation given a  $\sigma$ -algebra, 297
  - of a factorization, 308
- univariate normal distribution, 256
  - density, 256
  - distribution function, 256
  
- variance, 208
  - of an indicator, 210
  - of the sample mean, 210
  - rules of computation, 209
  
- Z-transformation, 211



## List of Symbols

$\neg$	Aa not, IX
$\vee$	Ab or, IX
$\wedge$	Ab and, IX
$\Rightarrow$	Aca implies, IX
$\Leftrightarrow$	Acb equivalent to, IX
$\exists$	Ada there is (synonymously, there exists), IX
$\forall$	Adb for all, IX
$\in$	Sacaaa element of, IX
$1_A$	SadaaaVmqaqa indicator (function) of the set $A$ , 20
$\cup$	Saeaaa union of sets, IX
$\cap$	Safaaa intersection of sets, IX
$\setminus$	Sahaaa set difference, IX
$A^c := \Omega \setminus A$	Sahcaa complement of a set $A$ with respect to a set $\Omega$ , IX
$\subset$	Sajaaa subset or equal, IX
$\times$	Sajbaa Cartesian product or product set, IX
$\otimes$	Sajbba product $\sigma$ -algebra, 14
$\otimes$	Sajbba product measure, 24
$\circ$	Sajbbb composition of two mappings, 58
$\odot$	SajbbbUfoaaa measure with density, 106
$\stackrel{\mu}{=}$	SajbbbVmaaaa $\mu$ -equivalence of mappings, 65
$\stackrel{\mu}{<}$	SajbbbVmhaaa smaller than except for a $\mu$ -null set, 67
$\stackrel{\mu}{>}$	SajbbbVmkaaa greater than except for a $\mu$ -null set, 67
$\stackrel{\mu}{\leq}$	SajbbbVmmeta smaller than or equal except for a $\mu$ -null set, 67
$\stackrel{\mu}{\geq}$	SajbbbVmoaaa greater than or equal except for a $\mu$ -null set, 67
$\stackrel{P}{=}$	SajbbbVmogaa $P$ -equivalence of random variables, 161
$\stackrel{\mu}{=} \mu\text{-a.a.}$	SajbbbVmsaaa equal for $\mu$ -almost all $\omega \in \Omega$ , 65
$\ll_{\mathcal{A}}$	SajbbbVpuaaa absolute continuity of a measure with respect to another measure, 109
$\approx_{\mathcal{A}}$	SajbbbVpvaaa null-set equivalence of two measures, 109

$\perp\!\!\!\perp_P$	SajbbbVtvaaa independence with respect to the probability measure $P$ , 143
$[a, b]$	Sajbca closed interval between real numbers $a$ and $b$ , 6
$]a, b]$	Sajbcc half-open interval including $b$ but not $a$ , 12
$\{x\}$	Sajbcg singleton, i. e., the set that contains $x$ as the only element, 6
$\emptyset$	Sajbcj empty set, IX
$(A_i, i \in I)$	Sajhcg family of sets $A_i, i \in I$ , where the index set $I$ may be finite, countable, or uncountable, IX
$\bigcup_{i \in I} A_i$	Sakaaa union of the sets $A_i, i \in I$ , IX
$\bigcap_{i \in I} A_i$	Samaaa intersection of the sets $A_i, i \in I$ , IX
$\bigcup_{i=1}^n A_i$	Samcaa union of finitely many sets $A_1, \dots, A_n$ , 5
$\bigcup_{i=1}^{\infty} A_i$	Sameaa union of countably many sets $A_1, A_2, \dots$ , 5
$\bigcap_{i=1}^n A_i$	Samfba intersection of finitely many sets $A_1, \dots, A_n$ , 6
$\bigcap_{i=1}^{\infty} A_i$	Samfca intersection of countably many sets $A_1, A_2, \dots$ , 6
$\prod_{i=1}^n A_i$	Samfga Cartesian product or product set of $n$ sets $A_i$ , IX
$\lim_{n \rightarrow \infty} a_n$	Sbaaaa limit of a sequence $a_1, a_2, \dots$ of real numbers, IX
$\sum_{i=1}^n a_i$	Scaaaa sum of the numbers $a_1, \dots, a_n$ , IX
$\sum_{i=1}^{\infty} a_i$	Sccaaaa $\lim_{n \rightarrow \infty} \sum_{i=1}^n a_i$ , IX
$\prod_{i=1}^n a_i$	Sdaaaa product of the real numbers $a_1, \dots, a_n$ , IX
$\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$	Sdchaa product $\sigma$ -algebra of the $\mathcal{A}_1, \dots, \mathcal{A}_n$ , 14
$\bigotimes_{i=1}^n \mathcal{A}_i$	Seaaaa product $\sigma$ -algebra of the $\mathcal{A}_1, \dots, \mathcal{A}_n$ , 14
$\int f d\mu$	Siaaaa integral of a measurable function $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ with respect to the measure $\mu$ , 94
$\int_A f d\mu$	Sigaaa integral of a measurable function $f: (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$ with respect to the measure $\mu$ over a subset $A$ of $\Omega$ , 96
$\int_a^b f(x) dx$	Simaaa Riemann integral of the function $f$ from $a$ to $b$ , 105
$\mathcal{A}  _{\Omega_0}$	Svcaaa trace of the set system $\mathcal{A}$ in the set $\Omega_0$ , 8
$\mathcal{C}'_f$	Svgaaa final $\sigma$ -algebra of $\mathcal{C}$ under $f$ , 54
$(\Omega, \mathcal{A})$	Tcaaaa measurable space, 6
$(\Omega, \mathcal{A}, \mu)$	Tcfaaa measure space, 17
$(\Omega, \mathcal{A}, P)$	Tcjaaa probability space, 127
$f: A \rightarrow B$	Uaaaa mapping $f$ assigning to each $a \in A$ one and only one element $b \in B$ , IX
$f(A)$	Udaaaa image of the set $A$ under $f$ , 41
$f^{-1}(A')$	Ugaaaa inverse image of the set $A'$ under $f$ , 41
$\{f \in A'\}$	Uhaaaa $:= f^{-1}(A')$ , 41
$\{f = \omega'\}$	Ujaaaa $:= f^{-1}(\{\omega'\})$ , 42
$(f_i, i \in I)$	Umaaaa family of mappings, 55
$f^{-1}(\mathcal{E}')$	Umaaaa set of the inverse images of all sets $A' \in \mathcal{E}'$ . If $\mathcal{E}'$ is a $\sigma$ -algebra, then $f^{-1}(\mathcal{E}')$ is the $\sigma$ -algebra generated by $f$ , 49
$g \circ f, g(f)$	Umgaaa composition of $f$ and $g$ , 58

$ f $	Umiaaa absolute value function of $f$ , 63
$f^+$	Umkaaa positive part of the function $f$ , 62
$f^-$	Umlaaa negative part of the function $f$ , 62
$f_n \uparrow f$	Umoaaa the sequence $f_1, f_2, \dots$ of functions converges point-wise and monotonically from below to $f$ , 91
$f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$	Vaaaaa $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ are measurable spaces and the mapping $f: \Omega \rightarrow \Omega'$ is $(\mathcal{A}, \mathcal{A}')$ -measurable, 47
$f: (\Omega, \mathcal{A}, \mu) \rightarrow \Omega'$	Veaaaa $(\Omega, \mathcal{A}, \mu)$ is a measure space and $f: \Omega \rightarrow \Omega'$ is a mapping, 64
$f: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$	Vhaaaa $f: \Omega \rightarrow \Omega'$ is an $(\mathcal{A}, \mathcal{A}')$ -measurable mapping and $\mu$ is a measure on $(\Omega, \mathcal{A})$ , 64
$f \underset{\mu}{=} g$	Vmaaaa the mappings $f$ and $g$ are $\mu$ -equivalent, 65
$f \underset{\mu}{<} g$	Vmhaaa $f$ is smaller than $g$ except for a set $A$ of arguments with $\mu(A) = 0$ , 67
$f \underset{\mu}{>} g$	Vmkaaa $f$ is greater than $g$ except for a set $A$ of arguments with $\mu(A) = 0$ , 67
$f \underset{\mu}{\leq} g$	Vmmaaa $f$ is smaller than or equal to $g$ except for a set $A$ of arguments with $\mu(A) = 0$ , 67
$f \underset{\mu}{\geq} g$	Vmoaaa $f$ is greater than or equal to $g$ except for a set $A$ of arguments with $\mu(A) = 0$ , 67
$f(\omega) \underset{\mu\text{-a.a.}}{=} g(\omega)$	Vmsaaa $f(\omega) = g(\omega)$ , for $\mu$ -almost all $\omega \in \Omega$ , 65
$\mu_f$	Vptaaa image measure of $\mu$ under $f$ , 68
$\nu \ll_{\mathcal{A}} \mu$	Vpuaaa the measure $\nu$ is absolutely continuous with respect to the measure $\mu$ , 109
$\nu \underset{\mathcal{A}}{\approx} \mu$	Vpvaaa the measures $\mu$ and $\nu$ are null-set equivalent, i. e., they are absolutely continuous with respect to each other, 109
$A \underset{P}{\perp} B$	Vtvaaa the events $A$ and $B$ are independent with respect to the probability measure $P$ , 143
$\underset{P}{\perp} (A_i, i \in I)$	Vtvea a family of events $A_i$ that are independent with respect to the probability measure $P$ , 143
$A \underset{P}{\perp} C   B$	Vtvdaa the events $A$ and $C$ are $B$ -conditionally independent with respect to the probability measure $P$ , 145
$\mathcal{E}_1 \underset{P}{\perp} \mathcal{E}_2$	Vtvgaa the set systems $\mathcal{E}_1$ and $\mathcal{E}_2$ are independent with respect to the probability measure $P$ , 144
$\underset{P}{\perp} (\mathcal{E}_i, i \in I)$	Vtvhaa a family of set systems $\mathcal{E}_i$ that are independent with respect to the probability measure $P$ , 144
$A \underset{P}{\perp} \mathcal{E}$	Vtvoaa the event $A$ and the set system $\mathcal{E}$ are independent with respect to the probability measure $P$ , 144
$\underset{P}{\perp} (X_i, i \in I)$	Vtvtha a family of random variables $X_i$ that are independent with respect to the probability measure $P$ , 171

$\perp_P X_1, \dots, X_n$	Vtvta the random variables $X_1, \dots, X_n$ are independent with respect to the probability measure $P$ , 171
$\mathcal{E} \perp_P X$	Vtvtoa the set system $\mathcal{E}$ and the random variable $X$ are independent with respect to the probability measure $P$ , 171
$X \perp_P (Y_i, i \in I)$	Vtvtua the random variable $X$ and the ( $\sigma$ -algebra generated by the) family of random variables are independent with respect to the probability measure $P$ , 171
$\perp_P (\mathcal{E}_i, i \in I)   B$	Vtwaa a family of set systems $\mathcal{E}_i$ that are $B$ -conditionally independent with respect to the probability measure $P$ , 146
$\bar{Y}$	Xtyaaa arithmetic mean (sample mean) of the real-valued random variables $Y_1, \dots, Y_n$ , 206
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}$	Zababe set systems, sometimes $\sigma$ -algebras, 4
$\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$	Zaeabe product $\sigma$ -algebra of the $\mathcal{A}_1, \dots, \mathcal{A}_n$ , 14
$\mathcal{A}  _{\Omega_0}$	Zahabe trace of the set system $\mathcal{A}$ in the set $\Omega_0$ , 8
$\mathcal{B}$	Zbcaaa Borel $\sigma$ -algebra on $\mathbb{R}$ , 12
$\mathcal{B}_n$	Zbccaa Borel $\sigma$ -algebra on $\mathbb{R}^n$ , 13
$\bar{\mathcal{B}}$	Zbcfaa Borel $\sigma$ -algebra on $\bar{\mathbb{R}}$ , 13
$\bar{\mathcal{B}}_n$	Zbgaaa Borel $\sigma$ -algebra on $\bar{\mathbb{R}}^n$ , 13
$\mathcal{B}_{n,p}$	Zbvaaa binomial distribution with parameters $n$ and $p$ , 247
$b_{n,p}$	Zbvaaab probability function of the binomial distribution with parameters $n$ and $p$ , 247
$Cov(X, Y)$	Zcmeta covariance of the random variables $X$ and $Y$ , 221
$Corr(X, Y)$	Zcgmeta correlation of the random variables $X$ and $Y$ , 225
$\mathcal{C}'_f$	Zchmeta final $\sigma$ -algebra of $\mathcal{C}$ under $f$ , 54
$\chi_n^2$	Zcvmeta central $\chi^2$ -distribution with $n$ degrees of freedom, 263
$\frac{d\nu}{d\mu}$	Zdmeta Radon-Nikodym density (also Radon-Nikodym derivative) of $\nu$ with respect to $\mu$ , 109
$\delta_\omega$	Zdmeta Dirac measure at (point) $\omega$ , 21
$E(Y)$	Zemeta expectation of the random variable $Y$ , 199
$E(\mathbf{x})$	Zefmeta column vector of expectations, 228
$E^B(Y)$	Zehmeta expectation of the random variable $Y$ with respect to the probability measure $P^B$ , 200
$E(Y B)$	Zeimeta conditional expectation value of $Y$ given the event $B$ , 278
$E(Y X=x)$	Zeifmeta conditional expectation value of $Y$ given the event $\{X=x\}$ , also denoted by $E(Y \{X=x\})$ , 278
$E(\mathbf{X})$	Zejmeta matrix of expectations, 228
$E_Y(g)$	Zejmeta expectation of the random variable $g$ with respect to the distribution of the random variable $Y$ , 203
$E_Y^{X=x}(g)$	Zelmeta expectation of $g$ with respect to the distribution $P_Y^{X=x}$ , 280
$E^{X=x}(Y)$	Zemeta expectation of $Y$ with respect to the conditional-probability measure $P^{X=x}$ , 279

$E^{\{X=x\}}(Y)$	Zemaaa expectation of $Y$ with respect to the conditional-probability measure $P^{X=x}$ , 279
$\epsilon$	Zepbaa residual with respect to a (multiple) linear quasi-regression, 232
$F_X$	Zfcaaa distribution function of a real-valued random variable $X$ , 182
$F_{X_1, \dots, X_n}$	Zfcgaa joint distribution function of $X_1, \dots, X_n$ , 184
$F(x) \Big _a^b$	Zfgaaa $F(b) - F(a)$ , 105
$f_X$	Zfgbaa probability density of a continuous real-valued random variable $X$ , 186
$ f $	Zfgjaa absolute value function of $f$ , 63
$f^+$	Zfhaaa positive part of the function $f$ , 62
$f^-$	Zfiaaa negative part of the function $f$ , 62
$(f_i, i \in I)$	Zflaaa family of mappings, 55
$f \circ \mu$	Zfoaaa measure with density $f$ with respect to $\mu$ , 106
$F_{m,n}$	Zfvaaa $F$ -distribution with $m$ and $n$ degrees of freedom, 266
$g \circ f, g(f)$	Zgfaaa composition of $f$ and $g$ , 58
$\Gamma$	Zggaaa gamma function, 262
$\mathcal{G}_p$	Zgvaaa geometric distribution with parameter $p$ , 252
$1_A$	Ziaaaa indicator (function) of the set $A$ , 20
$1_{X \in A'}$	Ziahaa indicator of the event $\{X \in A'\}$ , i. e., $1_{X \in A'} = 1_{X^{-1}(A')}$ , 158
$1_{X=x}, 1_{\{X=x\}}$	Zialaa indicator variable of the event $\{X=x\}$ , 279
$\mathcal{I}_1$	Zicaaa set system of all half-open intervals in $\mathbb{R}$ , 12
$id$	Zidaaa identity mapping, 48
$\mathcal{I}_n$	Zigaaa set system of all half-open cuboids in $\mathbb{R}^n$ , 13
$\lambda, \lambda_n$	Zlraaa Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}_n)$ , where $\lambda := \lambda_1$ , 22
$\mu, \nu$	Zmaaaa general symbols for measures, 17
$\mu_{\#}$	Zmaaaa counting measure, 21
$\mu_f$	Zmfaaa image measure of $\mu$ under $f$ , 68
$\mu_1 \otimes \dots \otimes \mu_n$	Zmgaaa product measure of $\mu_1, \dots, \mu_n$ , 24
$\nu \underset{\mathcal{A}}{\ll} \mu$	Zncabe the measure $\nu$ is absolutely continuous with respect to the measure $\mu$ , 109
$\nu \underset{\mathcal{A}}{\approx} \mu$	Zngabe the measures $\mu$ and $\nu$ are null-set equivalent, i. e., they are absolutely continuous with respect to each other, 109
$\mathbb{N}$	Znhabe set of all positive integers without zero, i. e., $\mathbb{N} = \{1, 2, \dots\}$ , 5
$\mathbb{N}_0$	Znhcbe set of all nonnegative integers including zero, i. e., $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ , 4
$\mathcal{N}_{\mu, \sigma^2}$	Znvaaa univariate normal distribution with parameters $\mu$ and $\sigma^2$ , 256
$\mathcal{N}_{\mu, \Sigma}$	Znvcaa multivariate normal distribution with parameters $\mu$ and $\Sigma$ , 259

$(\Omega, \mathcal{A}, P)$	Zohaaa probability space, 127
$\mathcal{P}(\Omega)$	Zpbhaa power set of $\Omega$ , 6
$P$	Zpcaaa probability measure, 127
$P(A)$	Zpceaa probability of the event $A$ , 127
$P(X=x)$	Zpcgaa probability of the event $\{X=x\} = X^{-1}(\{x\})$ , 156
$P(X \in A')$	Zpcjaa probability of the event $\{X \in A'\} = X^{-1}(A')$ , 156
$P(A B)$	Zpcjaa conditional probability of $A$ given $B$ (with respect to the probability measure $P$ ), 132
$P(X_1 \in A', X_2 \in B')$	Zpcjca probability of the event $\{X_1 \in A'\} \cap \{X_2 \in B'\}$ , 169
$P(A X=x)$	Zpcnca conditional probability of the event $A$ given the event $\{X=x\}$ with $P(\{X=x\}) > 0$ , also denoted by $P(A \{X=x\})$ , 279
$P(Y=y \{X=x\})$	Zpcncfa conditional probability of the event $\{Y=y\}$ given the event $\{X=x\}$ with $P(\{X=x\}) > 0$ , also denoted by $P(Y=y X=x)$ , 280
$P(A X)$	Zpcnia conditional probability of the event $A$ (with respect to $P$ ) given the discrete random variable $X$ , 284
$P^{X=x}$	Zpcpaa the $(X=x)$ -conditional probability measure on $(\Omega, \mathcal{A})$ , 164
$P_X$	Zpcpca distribution of the random variable $X$ (with respect to $P$ ), 156
$P_{X_1, \dots, X_n}$	Zpcpha joint distribution of the random variables $X_1, \dots, X_n$ , the distribution of the multivariate random variable $X = (X_1, \dots, X_n)$ , 168
$(P_X)_g$	Zpcqaa image measure of $P_X$ under $g$ , 157
$\Phi$	Zpdcaa distribution function of the standard normal distribution, 256
$P^B$	Zpcnaa $B$ -conditional-probability measure, 140
$P_X^B$	Zpfgaa distribution of $X$ with respect to the conditional-probability measure $P^B$ , 164
$p_X$	Zpiaaa probability function of a discrete random variable $X$ , 175
$p_{X_1, X_2}$	Zpifaa probability function of the bivariate random variable $X = (X_1, X_2)$ , 177
$\pi_j$	Zpkaaa $j$ th projection mapping, 56
$\mathcal{P}_\lambda$	Zpvaaa Poisson distribution with parameter $\lambda$ , 250
$Q_X$	Zqfaaa quantile function of a real-valued random variable $X$ , 183
$\mathbb{Q}$	Zqhaaa set of all rational numbers, 13
$Q_{lin}(Y X)$	Zqlaaa the composition of $X$ and the linear quasi-regression (or linear least-squares regression of $Y$ on $X$ ), 218
$Q_{lin}(Y X_1, \dots, X_n)$	Zqmaaa linear quasi-regression of $Y$ on $X_1, \dots, X_n$ , 232
$\mathbb{R}$	Zrjcbe set of all real numbers, 3
$\mathbb{R}^2$	Zrjfaa Cartesian product $\mathbb{R} \times \mathbb{R}$ , 3
$\mathbb{R}^n$	Zrjfaaa $n$ -fold Cartesian product $\mathbb{R} \times \dots \times \mathbb{R}$ , 3
$\bar{\mathbb{R}}$	Zrjjaa extended set of all real numbers, i. e., $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ , 13
$SD(Y)$	Zscaaa standard deviation of the random variable $Y$ , 208

$SE(\bar{Y})$	Zscdaa standard error of the sample mean of the random variables $Y_1, \dots, Y_n$ , 210
$\text{sgn}(f)$	Zscdha sign function of $f$ , 64
$\sigma(\mathcal{E})$	Zsdhaa $\sigma$ -algebra generated by the set system $\mathcal{E}$ , 9
$\sigma(f)$	Zsdhda $\sigma$ -algebra generated by the mapping $f$ , 51
$\sigma(f_1, \dots, f_n)$	Zsdhfa $\sigma$ -algebra generated by the mappings $f_1, \dots, f_n$ , 54
$\sigma(X)$	Zsdhja $\sigma$ -algebra generated by the random variable $X$ , 169
$\sigma(f, \mathcal{A})$	Zsdmca the $\sigma$ -algebra generated by the union of $\sigma(f)$ and $\mathcal{A}$ , 52
$\Sigma_{xy}$	Zsiaaa covariance matrix of $\mathbf{x}$ and $\mathbf{y}$ , 230
$\Sigma_{xx}$	Zsmaaa variance-covariance matrix of $\mathbf{x}$ , 230
$t_n$	Ztvaaa $t$ -distribution with $n$ degrees of freedom, 264
$\mathcal{U}_B$	Zuvaaa continuous uniform distribution on the set $B$ , 254
$\text{Var}(Y)$	Zvaaaa variance of the random variable $Y$ , 208
$\mathbf{x}$	Zxcaaa column vector of numerical random variables, 227
$\mathbf{X}$	Zxdaaa matrix of numerical random variables, 228
$X^{-1}(\mathcal{A}'_X)$	Zxdhjb $\sigma$ -algebra generated by the random variable $X$ , 169
$X \sim \mathcal{L}$	Zxgaaa the random variable $X$ has the distribution $\mathcal{L}$ . Examples for $\mathcal{L}$ are $\mathcal{B}_{n,p}$ , $\mathcal{P}_\lambda$ , $\mathcal{N}_{0,1}$ , or $F_{m,n}$ , 247
$X \stackrel{P}{=} Y$	Zxhaaa $X$ and $Y$ are $P$ -equivalent, 161
$X(\omega) \stackrel{P\text{-a.a.}}{=} Y(\omega)$	Zxhjaa $X$ and $Y$ are identical for $P$ -almost all $\omega \in \Omega$ , 161
$\bar{X}, \bar{Y}$	Zyjaaa arithmetic mean (sample mean) of the random variables $X_1, \dots, X_n$ or $Y_1, \dots, Y_n$ , respectively, 206
$Z_Y$	Zzaaaa $Z$ -transformation of the random variable $Y$ , 211