Springer Texts in Statistics

Allan Gut

Probability: A Graduate Course

Springer

Allan Gut
Department of Mathematics
University of Uppsala
SE-751 06 Uppsala
Sweden

Library of Congress Cataloging-in-Publication Data
Gut, Allan, 1944–
Probability: a graduate course / Allan Gut.
p. cm. — (Springer texts in statistics)
Includes bibliographical references and index.
ISBN 0-387-22833-0 (alk. paper)
1. Probabilities. 2. Distribution (Probability theory) 3. Probabilistic number theory. I. Title. II. Series.
QA273.G869 2005
519.2—dc22 2004056469

ISBN 0-387-22833-0 Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 10978764

Typesetting: Pages created by the author using Springer's TEX macro package.

springeronline.com

Preface

Toss a symmetric coin twice. What is the probability that both tosses will yield a head?

This is a well-known problem that anyone can solve. Namely, the probability of a head in each toss is $1/2$, so the probability of two consecutive heads is $1/2 \cdot 1/2 = 1/4$.

BUT! What did we do? What is involved in the solution? What are the arguments behind our computations? Why did we multiply the two halves connected with each toss?

This is reminiscent of the centipede[1] who was asked by another animal how he walks; he who has so many legs, in which order does he move them as he is walking? The centipede contemplated the question for a while, but found no answer. However, from that moment on he could no longer walk.

This book is written with the hope that we are not centipedes.

There exist two kinds of probabilists. One of them is the mathematician who views probability theory as a purely mathematical discipline, like algebra, topology, differential equations, and so on. The other kind views probability theory as the *mathematical modeling of random phenomena*, that is with a view toward applications, and as a companion to statistics, which aims at finding methods, principles and criteria in order to analyze data emanating from experiments involving random phenomena and other observations from the real world, with the ultimate goal of making wise decisions. I would like to think of myself as both.

What kind of a random process describes the arrival of claims at an insurance company? Is it one process or should one rather think of different processes, such as one for claims concerning stolen bikes and one for houses that have burnt down? How well should the DNA sequences of an accused offender and a piece of evidence match each other in order for a conviction? A

[1] Cent is 100, so it means an animal with 100 legs. In Swedish the name of the animal is *tusenfoting*, where “tusen” means 1000 and “fot” is foot; thus an animal with 1000 legs or feet.

milder version is how to order different species in a phylogenetic tree. What are the arrival rates of customers to a grocery store? How long are the service times? How do the clapidemia cells split? Will they create a new epidemic or can we expect them to die out? A classical application has been the arrivals of telephone calls to a switchboard and the duration of calls. Recent research and model testing concerning the Internet traffic has shown that the classical models break down completely and new thinking has become necessary. And, last but (not?) least, there are many games and lotteries.

The aim of this book is to provide the reader with a fairly thorough treatment of the main body of basic and classical probability theory, preceded by an introduction to the mathematics which is necessary for a solid treatment of the material. This means that we begin with basics from measure theory, such as σ-algebras, set theory, measurability (random variables) and Lebesgue integration (expectation), after which we turn to the Borel-Cantelli lemmas, inequalities, transforms and the three classical limit theorems: the law of large numbers, the central limit theorem and the law of the iterated logarithm. A final chapter on martingales – one of the most efficient, important, and useful tools in probability theory – is preceded by a chapter on topics that could have been included with the hope that the reader will be tempted to look further into the literature. The reason that these topics did not get a chapter of their own is that beyond a certain number of pages a book becomes deterring rather than tempting (or, as somebody said with respect to an earlier book of mine: "It is a nice format for bedside reading").

One thing that is *not* included in this book is a philosophical discussion of whether or not chance exist, whether or not randomness exists. On the other hand, probabilistic modeling is a wonderful, realistic, and efficient way to model phenomena containing uncertainties and ambiguities, regardless of whether or not the answer to the philosophical question is yes or no.

I remember having read somewhere a sentence like "There exist already so many textbooks [of the current kind], so, why do I write another one?" This sentence could equally well serve as an opening for the present book.

Luckily, I can provide an answer to that question. The answer is the short version of the story of the mathematician who was asked how one realizes that the fact he presented in his lecture (because this was really a he) was trivial. After 2 minutes of complete silence he mumbled

I know it's trivial, but I have forgotten why.

I strongly dislike the arrogance and snobbism that encompasses mathematics and many mathematicians. Books and papers are filled with expressions such as "it is easily seen", "it is trivial", "routine computations yield", and so on. The last example is sometimes modified into "routine, but tedious, computations yield". And we all know that behind things that are easily seen there may be years of thinking and/or huge piles of scrap notes that lead nowhere, and one sheet where everything finally worked out nicely.

Clearly, things become routine after many years. Clearly, facts become, at least *intuitively*, obvious after some decades. But in writing papers and books we try to help those who do not know yet, those who want to learn. We wish to attract people to this fascinating part of the world. Unfortunately though, phrases like the above ones are repellent, rather than being attractive. If a reader understands immediately that's fine. However, it is more likely that he or she starts off with something that either results in a pile of scrap notes or in frustration. Or both. And nobody is made happier, certainly not the reader. I have therefore avoided, or, at least, tried to avoid, expressions like the above unless they are adequate.

The main aim of a book is to be helpful to the reader, to help her or him to understand, to inform, to educate, and to attract (and not for the author to prove himself to the world). It is therefore essential to keep the flow, not only in the writing, but also in the reading. In the writing it is therefore of great importance to be rather extensive and not to leave too much to the (interested) reader.

A related aspect concerns the style of writing. Most textbooks introduce the reader to a number of topics in such a way that further insights are gained through exercises and problems, some of which are not at all easy to solve, let alone trivial. We take a somewhat different approach in that several such "would have been" exercises are given, together with their solutions as part of the ordinary text – which, as a side effect, reduces the number of exercises and problems at the end of each chapter. We also provide, at times, results for which the proofs consist of variations of earlier ones, and therefore are left as an exercise, with the motivation that doing almost the same thing as somebody else has done provides a much better understanding than reading, nodding and agreeing. I also hope that this approach creates the atmosphere of a dialogue rather than of the more traditional monologue (or sermon).

The ultimate dream is, of course, that this book contains no errors, no slips, no misprints. Henrik Wanntorp has gone over a substantial part of the manuscript with a magnifying glass, thereby contributing immensely to making that dream come true. My heartfelt thanks, Henrik. I also wish to thank Raimundas Gaigalas for several perspicacious remarks and suggestions concerning his favorite sections, and a number of reviewers for their helpful comments and valuable advice. As always, I owe a lot to Svante Janson for being available for any question at all times, and, more particularly, for always providing me with an answer. John Kimmel of Springer-Verlag has seen me through the process with a unique combination of professionalism, efficiency, enthusiasm and care, for which I am most grateful.

Finally, my hope is that the reader who has digested this book is ready and capable to attack any other text, for which a solid probabilistic foundation is necessary or, at least, desirable.

Uppsala *Allan Gut*
November 2004

Contents

Preface V

Outline of Contents XVII

Notation and Symbols XXI

1 Introductory Measure Theory 1
1 Probability Theory: An Introduction 1
2 Basics from Measure Theory 2
2.1 Sets 3
2.2 Collections of Sets 5
2.3 Generators 7
2.4 A Metatheorem and Some Consequences 9
3 The Probability Space 10
3.1 Limits and Completeness 11
3.2 An Approximation Lemma 13
3.3 The Borel Sets on $\mathbb{R}$ 14
3.4 The Borel Sets on $\mathbb{R}^n$ 16
4 Independence; Conditional Probabilities 16
4.1 The Law of Total Probability; Bayes' Formula 17
4.2 Independence of Collections of Events 18
4.3 Pair-wise Independence 19
5 The Kolmogorov Zero-one Law 20
6 Problems 22

2 Random Variables 25
1 Definition and Basic Properties 25
1.1 Functions of Random Variables 28
2 Distributions 30
2.1 Distribution Functions 30
2.2 Integration: A Preview 32

2.3 Decomposition of Distributions . . . 36
2.4 Some Standard Discrete Distributions . . . 39
2.5 Some Standard Absolutely Continuous Distributions . . . 40
2.6 The Cantor Distribution . . . 40
2.7 Two Perverse Examples . . . 42
3 Random Vectors; Random Elements . . . 43
3.1 Random Vectors . . . 43
3.2 Random Elements . . . 45
4 Expectation; Definitions and Basics . . . 46
4.1 Definitions . . . 46
4.2 Basic Properties . . . 48
5 Expectation; Convergence . . . 54
6 Indefinite Expectations . . . 58
7 A Change of Variables Formula . . . 60
8 Moments, Mean, Variance . . . 62
9 Product Spaces; Fubini's Theorem . . . 64
9.1 Finite-dimensional Product Measures . . . 64
9.2 Fubini's Theorem . . . 65
9.3 Partial Integration . . . 66
9.4 The Convolution Formula . . . 67
10 Independence . . . 68
10.1 Independence of Functions of Random Variables . . . 71
10.2 Independence of σ-Algebras . . . 71
10.3 Pair-wise Independence . . . 71
10.4 The Kolmogorov Zero-one Law Revisited . . . 72
11 The Cantor Distribution . . . 73
12 Tail Probabilities and Moments . . . 74
13 Conditional Distributions . . . 79
14 Distributions with Random Parameters . . . 81
15 Sums of a Random Number of Random Variables . . . 83
15.1 Applications . . . 85
16 Random Walks; Renewal Theory . . . 88
16.1 Random Walks . . . 88
16.2 Renewal Theory . . . 89
16.3 Renewal Theory for Random Walks . . . 90
16.4 The Likelihood Ratio Test . . . 91
16.5 Sequential Analysis . . . 91
16.6 Replacement Based on Age . . . 92
17 Extremes; Records . . . 93
17.1 Extremes . . . 93
17.2 Records . . . 93
18 Borel-Cantelli Lemmas . . . 96
18.1 The Borel-Cantelli Lemmas 1 and 2 . . . 96
18.2 Some (Very) Elementary Examples . . . 98
18.3 Records . . . 101

18.4 Recurrence and Transience of Simple Random Walks . . . 102
18.5 $\sum_{n=1}^{\infty} P(A_n) = \infty$ and $P(A_n \text{ i.o.}) = 0$ . . . 104
18.6 Pair-wise Independence . . . 104
18.7 Generalizations Without Independence . . . 105
18.8 Extremes . . . 107
18.9 Further Generalizations . . . 109
19 A Convolution Table . . . 113
20 Problems . . . 114

3 Inequalities . . . 119
1 Tail Probabilities Estimated via Moments . . . 119
2 Moment Inequalities . . . 127
3 Covariance; Correlation . . . 130
4 Interlude on L^p-spaces . . . 131
5 Convexity . . . 132
6 Symmetrization . . . 133
7 Probability Inequalities for Maxima . . . 138
8 The Marcinkiewics-Zygmund Inequalities . . . 146
9 Rosenthal's Inequality . . . 151
10 Problems . . . 153

4 Characteristic Functions . . . 157
1 Definition and Basics . . . 157
1.1 Uniqueness; Inversion . . . 159
1.2 Multiplication . . . 164
1.3 Some Further Results . . . 165
2 Some Special Examples . . . 166
2.1 The Cantor Distribution . . . 166
2.2 The Convolution Table Revisited . . . 168
2.3 The Cauchy Distribution . . . 170
2.4 Symmetric Stable Distributions . . . 171
2.5 Parseval's Relation . . . 172
3 Two Surprises . . . 173
4 Refinements . . . 175
5 Characteristic Functions of Random Vectors . . . 180
5.1 The Multivariate Normal Distribution . . . 180
5.2 The Mean and the Sample Variance Are Independent . . . 183
6 The Cumulant Generating Function . . . 184
7 The Probability Generating Function . . . 186
7.1 Random Vectors . . . 188
8 The Moment Generating Function . . . 189
8.1 Random Vectors . . . 191
8.2 Two Boundary Cases . . . 191
9 Sums of a Random Number of Random Variables . . . 192
10 The Moment Problem . . . 194

10.1 The Moment Problem for Random Sums . . . 196
11 Problems . . . 197

5 Convergence . . . 201
1 Definitions . . . 202
1.1 Continuity Points and Continuity Sets . . . 203
1.2 Measurability . . . 205
1.3 Some Examples . . . 206
2 Uniqueness . . . 207
3 Relations Between Convergence Concepts . . . 209
3.1 Converses . . . 212
4 Uniform Integrability . . . 214
5 Convergence of Moments . . . 218
5.1 Almost Sure Convergence . . . 218
5.2 Convergence in Probability . . . 220
5.3 Convergence in Distribution . . . 222
6 Distributional Convergence Revisited . . . 225
6.1 Scheffé's Lemma . . . 226
7 A Subsequence Principle . . . 229
8 Vague Convergence; Helly's Theorem . . . 230
8.1 Vague Convergence . . . 231
8.2 Helly's Selection Principle . . . 232
8.3 Vague Convergence and Tightness . . . 234
8.4 The Method of Moments . . . 237
9 Continuity Theorems . . . 238
9.1 The Characteristic Function . . . 238
9.2 The Cumulant Generating Function . . . 240
9.3 The (Probability) Generating Function . . . 241
9.4 The Moment Generating Function . . . 242
10 Convergence of Functions of Random Variables . . . 243
10.1 The Continuous Mapping Theorem . . . 245
11 Convergence of Sums of Sequences . . . 247
11.1 Applications . . . 249
11.2 Converses . . . 252
11.3 Symmetrization and Desymmetrization . . . 255
12 Cauchy Convergence . . . 256
13 Skorohod's Representation Theorem . . . 258
14 Problems . . . 260

6 The Law of Large Numbers . . . 265
1 Preliminaries . . . 266
1.1 Convergence Equivalence . . . 266
1.2 Distributional Equivalence . . . 267
1.3 Sums and Maxima . . . 268
1.4 Moments and Tails . . . 268

2 A Weak Law for Partial Maxima . . . 269
3 The Weak Law of Large Numbers . . . 270
3.1 Two Applications . . . 276
4 A Weak Law Without Finite Mean . . . 278
4.1 The St. Petersburg Game . . . 283
5 Convergence of Series . . . 284
5.1 The Kolmogorov Convergence Criterion . . . 286
5.2 A Preliminary Strong Law . . . 288
5.3 The Kolmogorov Three-series Theorem . . . 289
5.4 Lévy's Theorem on the Convergence of Series . . . 292
6 The Strong Law of Large Numbers . . . 294
7 The Marcinkiewicz-Zygmund Strong Law . . . 298
8 Randomly Indexed Sequences . . . 301
9 Applications . . . 305
9.1 Normal Numbers . . . 305
9.2 The Glivenko-Cantelli Theorem . . . 306
9.3 Renewal Theory for Random Walks . . . 306
9.4 Records . . . 307
10 Uniform Integrability; Moment Convergence . . . 309
11 Complete Convergence . . . 311
11.1 The Hsu-Robbins-Erdős Strong Law . . . 312
11.2 Complete Convergence and the Strong Law . . . 314
12 Some Additional Results and Remarks . . . 315
12.1 Convergence Rates . . . 315
12.2 Counting Variables . . . 320
12.3 The Case $r = p$ Revisited . . . 321
12.4 Random Indices . . . 322
13 Problems . . . 323

7 The Central Limit Theorem . . . 329
1 The i.i.d. Case . . . 330
2 The Lindeberg-Lévy-Feller Theorem . . . 330
2.1 Lyapounov's Condition . . . 339
2.2 Remarks and Complements . . . 340
2.3 Pair-wise Independence . . . 343
2.4 The Central Limit Theorem for Arrays . . . 344
3 Anscombe's Theorem . . . 345
4 Applications . . . 348
4.1 The Delta Method . . . 349
4.2 Stirling's Formula . . . 350
4.3 Renewal Theory for Random Walks . . . 350
4.4 Records . . . 351
5 Uniform Integrability; Moment Convergence . . . 352
6 Remainder Term Estimates . . . 354
6.1 The Berry-Esseen Theorem . . . 355

6.2 Proof of the Berry-Esseen Theorem 6.2 ... 357
7 Some Additional Results and Remarks ... 362
7.1 Rates of Rates ... 362
7.2 Non-uniform Estimates ... 363
7.3 Renewal Theory ... 364
7.4 Records ... 364
7.5 Local Limit Theorems ... 365
7.6 Large Deviations ... 365
7.7 Convergence Rates ... 366
7.8 Precise Asymptotics ... 371
7.9 A Short Outlook on Extensions ... 374
8 Problems ... 376

8 The Law of the Iterated Logarithm ... 383
1 The Kolmogorov and Hartman-Wintner LILs ... 384
1.1 Outline of Proof ... 385
2 Exponential Bounds ... 385
3 Proof of the Hartman-Wintner Theorem ... 387
4 Proof of the Converse ... 396
5 The LIL for Subsequences ... 398
5.1 A Borel-Cantelli Sum for Subsequences ... 401
5.2 Proof of Theorem 5.2 ... 402
5.3 Examples ... 404
6 Cluster Sets ... 404
6.1 Proofs ... 406
7 Some Additional Results and Remarks ... 412
7.1 Hartman-Wintner via Berry-Esseen ... 412
7.2 Examples Not Covered by Theorems 5.2 and 5.1 ... 413
7.3 Further Remarks on Sparse Subsequences ... 414
7.4 An Anscombe LIL ... 416
7.5 Renewal Theory for Random Walks ... 417
7.6 Record Times ... 417
7.7 Convergence Rates ... 418
7.8 Precise Asymptotics ... 419
7.9 The Other LIL ... 419
8 Problems ... 420

9 Limit Theorems; Extensions and Generalizations ... 423
1 Stable Distributions ... 424
2 The Convergence to Types Theorem ... 427
3 Domains of Attraction ... 430
3.1 Sketch of Preliminary Steps ... 433
3.2 Proof of Theorems 3.2 and 3.3 ... 435
3.3 Two Examples ... 438
3.4 Two Variations ... 439

3.5 Additional Results . . . 440
4 Infinitely Divisible Distributions . . . 442
5 Sums of Dependent Random Variables . . . 448
6 Convergence of Extremes . . . 451
6.1 Max-stable and Extremal Distributions . . . 451
6.2 Domains of Attraction . . . 456
6.3 Record Values . . . 457
7 The Stein-Chen Method . . . 459
8 Problems . . . 464

10 Martingales . . . 467
1 Conditional Expectation . . . 468
1.1 Properties of Conditional Expectation . . . 471
1.2 Smoothing . . . 474
1.3 The Rao-Blackwell Theorem . . . 475
1.4 Conditional Moment Inequalities . . . 476
2 Martingale Definitions . . . 477
2.1 The Defining Relation . . . 479
2.2 Two Equivalent Definitions . . . 479
3 Examples . . . 481
4 Orthogonality . . . 487
5 Decompositions . . . 489
6 Stopping Times . . . 491
7 Doob's Optional Sampling Theorem . . . 495
8 Joining and Stopping Martingales . . . 497
9 Martingale Inequalities . . . 501
10 Convergence . . . 508
10.1 Garsia's Proof . . . 508
10.2 The Upcrossings Proof . . . 511
10.3 Some Remarks on Additional Proofs . . . 513
10.4 Some Questions . . . 514
10.5 A Non-convergent Martingale . . . 515
10.6 A Central Limit Theorem? . . . 515
11 The Martingale $\{E(Z \mid \mathcal{F}_n)\}$ . . . 515
12 Regular Martingales and Submartingales . . . 516
12.1 A Main Martingale Theorem . . . 517
12.2 A Main Submartingale Theorem . . . 518
12.3 Two Non-regular Martingales . . . 519
12.4 Regular Martingales Revisited . . . 519
13 The Kolmogorov Zero-one Law . . . 520
14 Stopped Random Walks . . . 521
14.1 Finiteness of Moments . . . 521
14.2 The Wald Equations . . . 522
14.3 Tossing a Coin Until Success . . . 525
14.4 The Gambler's Ruin Problem . . . 526

14.5 A Converse . . . 529
15 Regularity . . . 531
15.1 First Passage Times for Random Walks . . . 535
15.2 Complements . . . 537
15.3 The Wald Fundamental Identity . . . 538
16 Reversed Martingales and Submartingales . . . 541
16.1 The Law of Large Numbers . . . 544
16.2 *U*-statistics . . . 547
17 Problems . . . 548

A Some Useful Mathematics . . . 555
1 Taylor Expansion . . . 555
2 Mill's Ratio . . . 558
3 Sums and Integrals . . . 559
4 Sums and Products . . . 560
5 Convexity; Clarkson's Inequality . . . 561
6 Convergence of (Weighted) Averages . . . 564
7 Regularly and Slowly Varying Functions . . . 566
8 Cauchy's Functional Equation . . . 568
9 Functions and Dense Sets . . . 570

References . . . 577

Index . . . 589

Outline of Contents

In this extended list of contents, we provide a short expansion of the headings into a quick overview of the contents of the book.

Chapter 1. Introductory Measure Theory

The mathematical foundation of probability theory is measure theory and the theory of Lebesgue integration. The bulk of the introductory chapter is devoted to measure theory: sets, measurability, σ-algebras, and so on. We do not aim at a full course in measure theory, rather to provide enough background for a solid treatment of what follows.

Chapter 2. Random Variables

Having set the scene, the first thing to do is to forget probability spaces (!). More precisely, for modeling random experiments one is interested in certain specific quantities, called *random variables*, rather than in the underlying probability space itself. In Chapter 2 we introduce random variables and present the basic concepts, as well as concrete applications and examples of probability models. In particular, Lebesgue integration is developed in terms of expectation of random variables.

Chapter 3. Inequalities

Some of the most useful tools in probability theory and mathematics for proving finiteness or convergence of sums and integrals are *inequalities*. There exist many useful ones spread out in books and papers. In Chapter 3 we make an attempt to present a sizable amount of the most important inequalities.

Chapter 4. Characteristic Functions

Just as there are i.a. Fourier transforms that transform convolution of functions into multiplication of their corresponding transforms, there exist probabilistic transforms that "map" addition of independent random variables into multiplication of their transforms, the most prominent one being the characteristic function.

Chapter 5. Convergence

Once we know how to add random variables the natural problem is to investigate asymptotics. We begin by introducing some convergence concepts, prove uniqueness, after which we investigate how and when they imply each other. Other important problems are when, and to what extent, limits and expectations (limits and integrals) can be interchanged, and when, and to what extent, functions of convergent sequences converge to the function of the limit.

Chapter 6. The Law of Large Numbers

The law of large numbers states that (the distribution of) the arithmetic mean of a sequence of independent trials stabilizes around the center of gravity of the underlying distribution (under suitable conditions). There exist *weak* and *strong* laws and several variations and extensions of them. We shall meet some of them as well as some applications.

Chapter 7. The Central Limit Theorem

The central limit theorem, which (in its simplest form) states that if the variance is finite, then the arithmetic mean, properly rescaled, of a sequence of independent trials approaches a normal distribution as the number of observations increases. There exist many variations and generalizations, of the theorem, the central one being the Lindeberg-Lévy-Feller theorem. We also prove results on moment convergence, and rate results, the foremost one being the celebrated Berry-Esseen theorem.

Chapter 8. The Law of the Iterated Logarithm

This is a special, rather delicate and technical, and very beautiful, result, which provides precise bounds on the oscillations of sums of the above kind. The name obviously stems from the iterated logarithm that appears in the expression of the parabolic bound.

Chapter 9. Limit Theorems; Extensions and Generalizations

There are a number of additional topics that would fit well into a text like the present one, but for which there is no room. In this chapter we shall meet a number of them – stable distributions, domain of attraction, infinite divisibility, sums of dependent random variables, extreme value theory, the Stein-Chen method – in a somewhat more sketchy or introductory style. The reader who gets hooked on such a topic will be advised to some relevant literature (more can be found via the Internet).

Chapter 10. Martingales

This final chapter is devoted to one of the most central topics, not only in probability theory, but also in more traditional mathematics. Following some introductory material on conditional expectations and the definition of a martingale, we present several examples, convergence results, results for stopped martingales, regular martingales, uniformly integrable martingales, stopped random walks, and reversed martingales.

In Addition

A list of notation and symbols precedes the main body of text, and an appendix with some mathematical tools and facts, a bibliography, and an index conclude the book. References are provided for more recent results, for more nontraditional material, and to some of the historic sources, but in general not to the more traditional material. In addition to cited material, the list of references contains references to papers and books that are relevant without having been specifically cited.

Suggestions for a Course Curriculum

One aim with the book is that it should serve as a graduate probability course – as the title suggests. In the same way as the sections in Chapter 9 contain materials that no doubt would have deserved chapters of their own, Chapters 6,7, and 8 contain sections entitled “Some Additional Results and Remarks”, in which a number of additional results and remarks are presented, results that are not as central and basic as earlier ones in those chapters.

An adequate course would, in my opinion, consist of Chapters 1-8, and 10, except for the sections “Some Additional Results and Remarks”, plus a skimming through Chapter 9 at the level of the instructor’s preferences.

Notation and Symbols

Ω	the sample space
ω	an elementary event
$\mathcal{F}$	the σ-algebra of events
x^+	$\max\{x, 0\}$
x^-	$-\min\{x, 0\}$
$[x]$	the integer part of x
$\log^+ x$	$\max\{1, \log x\}$
$\sim$	the ratio of the quantities on either side tends to 1
$\mathbb{N}$	the (positive) natural numbers
$\mathbb{Z}$	the integers
$\mathbb{R}$	the real numbers
$\mathcal{R}$	the Borel σ-algebra on $\mathbb{R}$
$\lambda(\,\cdot\,)$	Lebesgue measure
$\mathbb{Q}$	the rational numbers
$\mathbb{C}$	the complex numbers
C	the continuous functions
C_0	the functions in C tending to 0 at $\pm\infty$
C_B	the bounded continuous functions
$C[a, b]$	the functions in C with support on the interval $[a, b]$
D	the right-continuous, functions with left-hand limits
$D[a, b]$	the functions in D with support on the interval $[a, b]$
D^+	the non-decreasing functions in D
$\mathbb{J}_G$	the discontinuities of $G \in D$
$I\{A\}$	indicator function of (the set) A
$\#\{A\}$	number of elements in (cardinality of) A
$\|A\|$	number of elements in (cardinality of) A
A^c	complement of A
∂A	the boundary of A
$P(A)$	probability of A

$X, Y, Z, \ldots$	random variables
$F(x),\ F_X(x)$	distribution function (of X)
$X \in F$	X has distribution (function) F
$C(F_X)$	the continuity set of F_X
$p(x),\ p_X(x)$	probability function (of X)
$f(x),\ f_X(x)$	density (function) (of X)
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots$	random (column) vectors
$\mathbf{X}', \mathbf{Y}', \mathbf{Z}', \ldots$	the transpose of the vectors
$F_{X,Y}(x,y)$	joint distribution function (of X and Y)
$p_{X,Y}(x,y)$	joint probability function (of X and Y)
$f_{X,Y}(x,y)$	joint density (function) (of X and Y)
$E,\ E\,X$	expectation (mean), expected value of X
$\mathrm{Var}\,,\ \mathrm{Var}\,X$	variance, variance of X
$\mathrm{Cov}\,(X,Y)$	covariance of X and Y
$\rho,\ \rho_{X,Y}$	correlation coefficient (between X and Y)
$\mathrm{med}\,(X)$	median of X
$g(t),\ g_X(t)$	(probability) generating function (of X)
$\psi(t),\ \psi_X(t)$	moment generating function (of X)
$\varphi(t),\ \varphi_X(t)$	characteristic function (of X)
$X \sim Y$	X and Y are equivalent random variables
$X \overset{a.s.}{=} Y$	X and Y are equal (point-wise) almost surely
$X \overset{d}{=} Y$	X and Y are equidistributed
$X_n \overset{a.s.}{\to} X$	X_n converges almost surely (a.s.) to X
$X_n \overset{p}{\to} \mathrm{X}$	X_n converges in probability to X
$X_n \overset{r}{\to} \mathrm{X}$	X_n converges in r-mean (L^r) to X
$X_n \overset{d}{\to} X$	X_n converges in distribution to X
$X_n \overset{a.s.}{\not\to}$	X_n does not converge almost surely
$X_n \overset{p}{\not\to}$	X_n does not converge in probability
$X_n \overset{r}{\not\to}$	X_n does not converge in r-mean (L^r)
$X_n \overset{d}{\not\to}$	X_n does not converge in distribution
$\Phi(x)$	standard normal distribution function
$\phi(x)$	standard normal density (function)
$F \in \mathcal{D}(G)$	F belongs to the domain of attraction of G
$g \in \mathcal{RV}\,(\rho)$	g varies regularly at infinity with exponent ρ
$g \in \mathcal{SV}$	g varies slowly at infinity

$\text{Be}(p)$	Bernoulli distribution
$\beta(r, s)$	beta distribution
$\text{Bin}(n, p)$	binomial distribution
$C(m, a)$	Cauchy distribution
$\chi^2(n)$	chi-square distribution
$\delta(a)$	one-point distribution
$\text{Exp}(a)$	exponential distribution
$F(m, n)$	(Fisher's) F-distribution
$\text{Fs}(p)$	first success distribution
$\Gamma(p, a)$	gamma distribution
$\text{Ge}(p)$	geometric distribution
$H(N, n, p)$	hypergeometric distribution
$L(a)$	Laplace distribution
$\text{LN}(\mu, \sigma^2)$	log-normal distribution
$N(\mu, \sigma^2)$	normal distribution
$N(0, 1)$	standard normal distribution
$\text{NBin}(n, p)$	negative binomial distribution
$\text{Pa}(k, \alpha)$	Pareto distribution
$\text{Po}(m)$	Poisson distribution
$\text{Ra}(\alpha)$	Rayleigh distribution
$t(n)$	(Student's) t-distribution
$\text{Tri}(a, b)$	triangular distribution on (a, b)
$U(a, b)$	uniform or rectangular distribution on (a, b)
$W(a, b)$	Weibull distribution
$X \in P(\theta)$	X has a P-distribution with parameter θ
$X \in P(\alpha, \beta)$	X has a P-distribution with parameters α and β

a.e.	almost everywhere
a.s.	almost surely
cf.	*confer*, compare, take counsel
i.a.	*inter alia*, among other things, such as
i.e.	*id est*, that is
i.o.	infinitely often
iff	if and only if
i.i.d.	independent, identically distributed
viz.	*videlicet*, in which
w.l.o.g.	without loss of generality
♠	hint for solving a problem
♣	bonus remark in connection with a problem
□	end of proof, definitions, exercises, remarks, etc.

1

Introductory Measure Theory

1 Probability Theory: An Introduction

The object of *probability theory* is to describe and investigate mathematical models of random phenomena, primarily from a theoretical point of view. Closely related to probability theory is *statistics*, which is concerned with creating principles, methods, and criteria in order to treat data pertaining to such (random) phenomena or data from experiments and other observations of the real world, by using, for example, the theories and knowledge available from the theory of probability.

Probability models thus aim at describing *random experiments*, that is, experiments that can be repeated (indefinitely) and where future outcomes cannot be exactly predicted – due to randomness – even if the experimental situation can be fully controlled.

The basis of probability theory is the *probability space*. The key idea behind probability spaces is the *stabilization of the relative frequencies*. Suppose that we perform "independent" repetitions of a random experiment and that we record each time if some "event" A occurs or not (although we have not yet mathematically defined what we mean by independence or by an event). Let $f_n(A)$ denote the number of occurrences of A in the first n trials, and $r_n(A)$ the relative frequency, $r_n(A) = f_n(A)/n$. Since the dawn of history one has observed the stabilization of the relative frequencies, that is, one has observed that (it seems that)

$$r_n(A) \quad \text{converges to some real number as} \quad n \to \infty.$$

The intuitive interpretation of the probability concept is that if the probability of some event A is 0.6, one should expect that by performing the random experiment "many times" the relative frequency of occurrences of A should be approximately 0.6.

The next step is to axiomatize the theory, to make it mathematically rigorous. Although games of chance have been performed for thousands of years, a mathematically rigorous treatment of the theory of probability only came about in the 1930's by the Soviet/Russian mathematician A.N. Kolmogorov (1903–1987) in his fundamental monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung* [163], which appeared in 1933.

The first observation is that a number of rules that hold for relative frequencies should also hold for probabilities. This immediately calls for the question "which is the minimal set of rules?"

In order to answer this question one introduces the *probability space* or *probability triple* $(\Omega, \mathcal{F}, P)$, where

- Ω is the *sample space*;
- $\mathcal{F}$ is the collection of *events*;
- P is a probability measure.

The fact that P is a probability measure means that it satisfies the three *Kolmogorov axioms* (to be specified ahead).

In a first course in probability theory one learns that

"the collection of events = the subsets of Ω",

maybe with an additional remark that this is not quite true, but true enough for the purpose of that course.

To clarify the situation we need some definitions and facts from measure theory in order to answer questions such as

"What does it mean for a set to be measurable?"

After this we shall return to a proper definition of the probability space.

2 Basics from Measure Theory

In addition, to straighten out the problems raised by such questions, we need rules for how to operate on what we shall define as events. More precisely, a problem may consist of finding the probability of one or the other of two things happening, or for something *not to happen*, and so on. We thus need rules and conventions for how to handle events, how we can combine them or not combine them. This means i.a. that we need to define collections of sets with a certain structure. For example, a collection such that the intersection of two events is an event, or the collection of sets such that the complement of an event is an event, and also rules for how various collections connect. This means that we have to confront ourselves with some notions from measure theory. Since this is rather a tool than a central theme of this book we confine ourselves to an overview of the most important parts of the topic, leaving some of the "routine but tedious calculations" as exercises.

2.1 Sets

Definitions; Notation

A set is a collection of "objects", concrete or abstract, called *elements*. A set is *finite* if the number of elements is finite, and it is *countable* if the number of elements is countable, that is, if one can label them by the positive integers in such a manner that no element remains unlabeled. Sets are usually denoted by capitals from the early part of the alphabet, A, B, C, and so on. If several sets are related, "of the same kind", it is convenient to use the same letter for them, but to add indices; $A_1, A_2, A_3, \ldots$.

The set $A = \{1, 2, \ldots, n\}$ is finite; $|A| = n$. The natural numbers, $\mathbb{N}$ constitute a countable set, and so do the set of rationals, $\mathbb{Q}$, whereas the set of irrationals and the set of reals, $\mathbb{R}$, are uncountable, as is commonly verified by Cantor's diagonal method. Although the natural numbers and the reals belong to different collections of sets they are both infinite in the sense that the number of elements is infinite in both sets, but the infinities are different. The same distinction holds true for the rationals and the irrationals. Infinities are distinguished with the aid of the *cardinal numbers*, which came about after Cantor's proof of the fact that the infinity of the reals is larger than that of the natural numbers, that there are "more" reals than natural numbers. Cardinal numbers are denoted by the Hebrew letter *alef*, where the successively larger cardinal numbers have increasing indices. The first cardinal number is $\aleph_0 = |\mathbb{N}|$, the *cardinality* of $\mathbb{N}$. Moreover, $|\mathbb{R}| = 2^{\aleph_0}$.

Let us mention in passing that a long-standing question, in fact, one of Hilbert's famous problems, has been whether or not there exist infinities between $|\mathbb{N}|$ and $|\mathbb{R}|$. The famous *continuum hypothesis* states that this is not the case, a claim that can be formulated as $\aleph_1 = 2^{\aleph_0}$. The interesting fact is that it has been proved that this claim can neither be proved nor disproved within the usual axiomatic framework of mathematics. Moreover, one may assume it to be true or false, and neither assumption will lead to any contradictory results. The continuum hypothesis is said to be *undecidable*. For more, see [51].

Set Operations

Just as real (or complex) numbers can be added or multiplied, there exist *operations* on sets. Let $A, A_1, A_2, \ldots$ and $B, B_1, B_2, \ldots$ be sets.

- Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$;
- Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$;
- Complement: $A^c = \{x : x \notin A\}$;
- Difference: $A \smallsetminus B = A \cap B^c$;
- Symmetric difference: $A \,\triangle\, B = (A \smallsetminus B) \cup (B \smallsetminus A)$.

We also use standard notations such as $\cup_{k=1}^{n} A_k$ and $\cap_{j=1}^{\infty} B_j$ for unions and intersections of finitely or countably many sets.

Exercise 2.1. Check to what extent the associative and distributive rules for these operations are valid. □

Some additional terminology:

- the empty set: $\emptyset$;
- subset: A is a *subset* of B, $A \subset B$, if $x \in A \Longrightarrow x \in B$;
- disjoint: A and B are *disjoint* if $A \cap B = \emptyset$;
- power set: $\mathfrak{P}(\Omega) = \{A : A \subset \Omega\}$;
- $\{A_n,\, n \geq 1\}$ is *non-decreasing*, $A_n \nearrow$, if $A_1 \subset A_2 \subset \cdots$;
- $\{A_n,\, n \geq 1\}$ is *non-increasing*, $A_n \searrow$, if $A_1 \supset A_2 \supset \cdots$.

The *de Morgan formulas*,

$$\Big(\bigcup_{k=1}^{n} A_k\Big)^c = \bigcap_{k=1}^{n} A_k^c \quad \text{and} \quad \Big(\bigcap_{k=1}^{n} A_k\Big)^c = \bigcup_{k=1}^{n} A_k^c\,, \tag{2.1}$$

can be verified by picking $\omega \in \Omega$ belonging to the set made up by the left-hand side and then show that it also belongs to the right-hand side, after which one does the same the other way around (please do that!). Alternatively one realizes that both members express the same fact. In the first case, this is the fact that an element that does not belong to any A_k whatsoever belongs to all complements, and therefore to their intersection. In the second case this is the fact that an element that does not belong to every A_k belongs to at least one of the complements.

Limits of Sets

It is also possible to define limits of sets. However, not every sequence of sets has a limit.

Definition 2.1. *Let $\{A_n,\, n \geq 1\}$ be a sequence of subsets of Ω. We define*

$$A_* = \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m,$$

$$A^* = \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

If the sets A_* and A^* agree, then

$$A = A_* = A^* = \lim_{n\to\infty} A_n.$$ □

One instance when a limit exists is when the sequence of sets is monotone.

Proposition 2.1. *Let $\{A_n,\, n \geq 1\}$ be a sequence of subsets of Ω.*

(i) *If* $A_1 \subset A_2 \subset A_3 \cdots$, *then*

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

(ii) *If* $A_1 \supset A_2 \supset A_3 \cdots$, *then*

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Exercise 2.2. Prove the proposition. □

2.2 Collections of Sets

Collections of sets, are defined according to a setup of rules. Different rules yield different collections. Certain collections are more easy to deal with than others depending on the property or theorem to prove. We now present a number of rules and collections, as well as results on how they connect. Since much of this is more or less well known to a mathematics student we leave essentially all proofs, which consist of longer or shorter, sometimes routine but tedious, manipulations, as exercises.

Let $\mathcal{A}$ be a non-empty collection of subsets of Ω, and consider the following set relations:

(a) $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$;
(b) $A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A}$;
(c) $A, B \in \mathcal{A} \Longrightarrow A \cap B \in \mathcal{A}$;
(d) $A, B \in \mathcal{A},\ B \subset A \Longrightarrow A \smallsetminus B \in \mathcal{A}$;
(e) $A_n \in \mathcal{A},\, n \geq 1, \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$;
(f) $A_n \in \mathcal{A},\, n \geq 1,\, A_i \cap A_j = \emptyset,\, i \neq j \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$;
(g) $A_n \in \mathcal{A},\, n \geq 1, \Longrightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$;
(h) $A_n \in \mathcal{A},\, n \geq 1,\, A_n \nearrow \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$;
(j) $A_n \in \mathcal{A},\, n \geq 1,\, A_n \searrow \Longrightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.

A number of relations among these rules and extensions of them can be established. For example (a) and one of (b) and (c), together with the de Morgan formulas, yield the other; (a) and one of (e) and (g), together with the de Morgan formulas, yield the other; (b) and induction shows that (b) can be extended to any finite union of sets; (c) and induction shows that (c) can be extended to any finite intersection of sets, and so on.

Exercise 2.3. Check these statements, and verify some more. □

Here are now definitions of some collections of sets.

Definition 2.2. *Let $\mathcal{A}$ be a collection of subsets of Ω.*

- *$\mathcal{A}$ is an* algebra *or a* field *if $\Omega \in \mathcal{A}$ and properties* (a) *and* (b) *hold;*
- *$\mathcal{A}$ is a* σ-algebra *or a* σ-field *if $\Omega \in \mathcal{A}$ and properties* (a) *and* (e) *hold;*
- *$\mathcal{A}$ is a* monotone class *if properties* (h) *and* (j) *hold;*
- *$\mathcal{A}$ is a* π-system *if property* (c) *holds;*
- *$\mathcal{A}$ is a* Dynkin system *if $\Omega \in \mathcal{A}$, and properties* (d) *and* (h) *hold.* □

Remark 2.1. Dynkin systems are also called λ-*systems*.

Remark 2.2. The definition of a Dynkin system varies. One alternative, in addition to the assumption that $\Omega \in \mathcal{A}$, is that (a) and (f) hold. □

Exercise 2.4. The obvious exercise is to show that the two definitions of a Dynkin system are equivalent. □

The definitions of the different collections of sets are obviously based on minimal requirements. By manipulating the different properties (a)–(j), for example together with the de Morgan formulas, other properties can be derived. The following relations between different collections of sets are obtained by such manipulations.

Theorem 2.1. *The following connections hold:*

1. *Every algebra is a π-system.*
2. *Every σ-algebra is an algebra.*
3. *An algebra is a σ-algebra if and only if it is a monotone class.*
4. *Every σ-algebra is a Dynkin system.*
5. *A Dynkin system is a σ-algebra if and only if it is π-system.*
6. *Every Dynkin system is a monotone class.*
7. *Every σ-algebra is a monotone class.*
8. *The power set of any subset of Ω is a σ-algebra on that subset.*
9. *The intersection of any number of σ-algebras, countable or uncountable, is, again, a σ-algebra.*
10. *The countable union of a non-decreasing sequence of σ-algebras is an algebra, but not necessarily a σ-algebra.*
11. *If $\mathcal{A}$ is a σ-algebra, and $B \subset \Omega$, then $B \cap \mathcal{A} = \{B \cap A : A \in \mathcal{A}\}$ is a σ-algebra on B.*
12. *If Ω and Ω' are sets, $\mathcal{A}'$ a σ-algebra in Ω' and $T : \Omega \to \Omega'$ a mapping, then $T^{-1}(\mathcal{A}') = \{T^{-1}(A') : A' \in \mathcal{A}'\}$ is a σ-algebra on Ω.*

Exercise 2.5. (a) Prove the above statements.

(b) Find two σ-algebras, the union of which is not an algebra (only very few elements in each suffice).
(c) Prove that if, for the infinite set Ω, $\mathcal{A}$ consists of all $A \subset \Omega$, such that either A or A^c is finite, then $\mathcal{A}$ is an algebra, but not a σ-algebra. □

2.3 Generators

Let $\mathcal{A}$ be a collection of subsets of Ω. Since the power set, $\mathfrak{P}(\Omega)$, is a σ-algebra, it follows that there exists at least one σ-algebra containing $\mathcal{A}$. Since, moreover, the intersection of any number of σ-algebras is, again, a σ-algebra, there exists a *smallest* σ-algebra containing $\mathcal{A}$. In fact, let

$$\mathcal{F}^* = \{\sigma\text{-algebras } \supset \mathcal{A}\}.$$

The smallest σ-algebra containing $\mathcal{A}$ equals

$$\bigcap_{\mathcal{G}\in\mathcal{F}^*} \mathcal{G},$$

and is unique since we have intersected *all* σ-algebras containing $\mathcal{A}$.

Definition 2.3. *Let $\mathcal{A}$ be a collection of subsets of Ω. The smallest σ-algebra containing $\mathcal{A}$, $\sigma\{\mathcal{A}\}$, is called the* σ-algebra generated by $\mathcal{A}$. *Similarly, the smallest Dynkin system containing $\mathcal{A}$, $\mathfrak{D}\{\mathcal{A}\}$, is called the* Dynkin system generated *by $\mathcal{A}$, and the smallest monotone class containing $\mathcal{A}$, $\mathfrak{M}\{\mathcal{A}\}$, is called the* monotone class generated *by $\mathcal{A}$. In each case $\mathcal{A}$ is called the* generator *of the actual collection.* □

Remark 2.3. The σ-algebra generated by $\mathcal{A}$ is also called "the minimal σ-algebra containing $\mathcal{A}$". Similarly for the other collections.

Remark 2.4. Let $\{\mathcal{A}_n,\ n \geq 1\}$ be σ-algebras. Even though the union need not be a σ-algebra, $\sigma\{\bigcup_{n=1}^{\infty} \mathcal{A}_n\}$, that is, the σ-algebra generated by $\{\mathcal{A}_n,\ n \geq 1\}$, always exists. □

Exercise 2.6. Prove that

(i) If $\mathcal{A} = A$, a single set, then $\sigma\{\mathcal{A}\} = \sigma\{A\} = \{\emptyset, A, A^c, \Omega\}$.
(ii) If $\mathcal{A}$ is a σ-algebra, then $\sigma\{\mathcal{A}\} = \mathcal{A}$. □

The importance and usefulness of generators is demonstrated by the following two results.

Theorem 2.2. *Let $\mathcal{A}$ be an algebra. Then*

$$\mathfrak{M}\{\mathcal{A}\} = \sigma\{\mathcal{A}\}.$$

Proof. Since every σ-algebra is a monotone class (Theorem 2.1) and $\mathfrak{M}\{\mathcal{A}\}$ is the minimal monotone class containing $\mathcal{A}$, we know from the outset that $\mathfrak{M}\{\mathcal{A}\} \subset \sigma\{\mathcal{A}\}$. To prove the opposite inclusion we must, due to the minimality of $\sigma\{\mathcal{A}\}$, prove that $\mathfrak{M}\{\mathcal{A}\}$ is a σ-algebra, for which it is sufficient to prove that $\mathfrak{M}\{\mathcal{A}\}$ is an algebra (Theorem 2.1 once more). This means that we have to verify that properties (a) and (b) hold;

$$\begin{cases} B \in \mathfrak{M}\{\mathcal{A}\} & \Longrightarrow \quad B^c \in \mathfrak{M}\{\mathcal{A}\}, \quad \text{and} \\ B, C \in \mathfrak{M}\{\mathcal{A}\} & \Longrightarrow \quad B \cup C \in \mathfrak{M}\{\mathcal{A}\}. \end{cases} \tag{2.2}$$

Toward this end, let

$$\mathcal{E}_1 = \big\{B \in \mathfrak{M}\{\mathcal{A}\} : B \cup C \in \mathfrak{M}\{\mathcal{A}\} \text{ for all } C \in \mathcal{A}\big\},$$
$$\mathcal{E}_2 = \big\{B \in \mathfrak{M}\{\mathcal{A}\} : B^c \in \mathfrak{M}\{\mathcal{A}\}\big\}.$$

We first note that $\mathcal{E}_1$ is a monotone class via the identities

$$\Big(\bigcap_{k=1}^{\infty} B_k\Big) \cup C = \bigcap_{k=1}^{\infty} (B_k \cup C) \quad \text{and} \quad \Big(\bigcup_{k=1}^{\infty} B_k\Big) \cup C = \bigcup_{k=1}^{\infty} (B_k \cup C), \tag{2.3}$$

and that $\mathcal{E}_2$ is a monotone class via the de Morgan formulas, (2.1).

Secondly, by definition, $\mathcal{A} \subset \mathfrak{M}\{\mathcal{A}\}$, and by construction,

$$\mathcal{A} \subset \mathcal{E}_k \subset \mathfrak{M}\{\mathcal{A}\}, \quad k = 1, 2,$$

so that, in view of minimality of $\mathfrak{M}\{\mathcal{A}\}$,

$$\mathfrak{M}\{\mathcal{A}\} = \mathcal{E}_1 = \mathcal{E}_2.$$

To finish off, let

$$\mathcal{E}_3 = \big\{B \in \mathfrak{M}\{\mathcal{A}\} : B \cup C \in \mathfrak{M}\{\mathcal{A}\} \text{ for all } C \in \mathfrak{M}\{\mathcal{A}\}\big\}.$$

Looking at $\mathcal{E}_1 = \mathfrak{M}\{\mathcal{A}\}$ from another angle, we have shown that for every $B \in \mathfrak{M}\{\mathcal{A}\}$ we know that if $C \in \mathcal{A}$, then $B \cup C \in \mathfrak{M}\{\mathcal{A}\}$, which means that

$$\mathcal{A} \subset \mathcal{E}_3.$$

Moreover, $\mathcal{E}_3$ is a monotone class via (2.3), so that, by minimality again, we must have $\mathfrak{M}\{\mathcal{A}\} = \mathcal{E}_3$.

We have thus shown that $\mathfrak{M}\{\mathcal{A}\}$ obeys properties (a) and (b). □

By suppressing the minimality of the monotone class, the following corollary emerges (because an arbitrary monotone class contains the minimal one).

Corollary 2.1. *If $\mathcal{A}$ is an algebra and $\mathcal{G}$ a monotone class containing $\mathcal{A}$, then*

$$\mathcal{G} \supset \sigma\{\mathcal{A}\}.$$

A related theorem, *the monotone class theorem*, concerns the equality between the Dynkin system and the σ-algebra generated by the same π-system.

Theorem 2.3. (The monotone class theorem)
If $\mathcal{A}$ is a π-system on Ω, then

$$\mathfrak{D}\{\mathcal{A}\} = \sigma\{\mathcal{A}\}.$$

Proof. The proof runs along the same lines as the previous one. Namely, one first observes that $\mathfrak{D}\{\mathcal{A}\} \subset \sigma\{\mathcal{A}\}$, since every σ-algebra is a Dynkin system (Theorem 2.1) and $\mathfrak{D}\{\mathcal{A}\}$ is the minimal Dynkin system containing $\mathcal{A}$.

For the converse we must show that $\mathfrak{D}\{\mathcal{A}\}$ is a π-system (Theorem 2.1). In order to achieve this, let

$$\mathcal{D}_C = \big\{B \subset \Omega : B \cap C \in \mathfrak{D}\{\mathcal{A}\}\big\} \quad \text{for} \quad C \in \mathfrak{D}\{\mathcal{A}\}.$$

We claim that $\mathcal{D}_C$ is a Dynkin system.

To prove this we check the requirements for a collection of sets to constitute a Dynkin system. In this case we use the following alternative (recall Remark 2.2), namely, we show that $\Omega \in \mathcal{D}_C$, and that (a) and (f) hold.

Let $C \in \mathfrak{D}\{\mathcal{A}\}$.

- Since $\Omega \cap C = C$, it follows that $\Omega \in \mathcal{D}_C$.
- If $B \in \mathcal{D}_C$, then

$$B^c \cap C = (\Omega \smallsetminus B) \cap C = (\Omega \cap C) \smallsetminus (B \cap C),$$

which shows that $B^c \in \mathcal{D}_C$.

- Finally, if $\{B_n,\, n \geq 1\}$ are disjoint sets in $\mathcal{D}_C$, then

$$\Big(\bigcup_{n=1}^{\infty} B_n\Big) \cap C = \bigcup_{n=1}^{\infty} (B_n \cap C),$$

which proves that $\bigcup_{n=1}^{\infty} B_n \in \mathcal{D}_C$.

The requirements for $\mathcal{D}_C$ to be a Dynkin system are thus fulfilled. And, since C was arbitrarily chosen, this is true for any $C \in \mathfrak{D}\{\mathcal{A}\}$.

Now, since, by definition, $\mathcal{A} \subset \mathcal{D}_A$ for every $A \in \mathcal{A}$, it follows that

$$\mathfrak{D}\{\mathcal{A}\} \subset \mathcal{D}_A \quad \text{for every} \quad A \in \mathcal{A}.$$

For $C \in \mathfrak{D}\{\mathcal{A}\}$ we now have $C \cap A \in \mathfrak{D}\{\mathcal{A}\}$ for every $A \in \mathcal{A}$, which implies that $\mathcal{A} \subset \mathcal{D}_C$, and, hence, that $\mathfrak{D}\{\mathcal{A}\} \subset \mathcal{D}_C$ for every $C \in \mathfrak{D}\{\mathcal{A}\}$. Consequently,

$$B, C \in \mathfrak{D}\{\mathcal{A}\} \quad \Longrightarrow \quad B \cap C \in \mathfrak{D}\{\mathcal{A}\},$$

that is, $\mathfrak{D}\{\mathcal{A}\}$ is a π-system. □

By combining Theorems 2.2 and 2.3 (and the exercise preceding the former) the following result emerges.

Corollary 2.2. *If $\mathcal{A}$ is a σ-algebra, then*

$$\mathfrak{M}\{\mathcal{A}\} = \mathfrak{D}\{\mathcal{A}\} = \sigma\{\mathcal{A}\} = \mathcal{A}.$$

2.4 A Metatheorem and Some Consequences

A frequent proof technique is to establish some kind of reduction from an infinite setting to a finite one; simple functions, rectangles, and so on. Such proofs can often be identified in that they open by statements such as

"it suffices to check rectangles",

"it suffices to check step functions".

The basic idea behind such statements is that there either exists some approximation theorem that "takes care of the rest", or that some convenient part of Theorem 2.1 can be exploited for the remaining part of the proof. Our next result puts this device into a more stringent form, although in a somewhat metaphoric sense.

Theorem 2.4. (A Metatheorem)
(i) *Suppose that some property holds for some monotone class $\mathcal{E}$ of subsets. If $\mathcal{A}$ is an algebra that generates the σ-algebra $\mathcal{G}$ and $\mathcal{A} \subset \mathcal{E}$, then $\mathcal{E} \supset \mathcal{G}$.*
(ii) *Suppose that some property holds for some Dynkin system $\mathcal{E}$ of subsets. If $\mathcal{A}$ is a π-system that generates the σ-algebra $\mathcal{G}$ and $\mathcal{A} \subset \mathcal{E}$, then $\mathcal{E} \supset \mathcal{G}$.*

Proof. Let

$$\mathcal{E} = \{E : \text{the property is satisfied}\}.$$

(i): It follows from the assumptions and Theorem 2.2, respectively, that

$$\mathcal{E} \supset \mathfrak{M}\{\mathcal{A}\} = \sigma\{\mathcal{A}\} = \mathcal{G}.$$

(ii): Apply Theorem 2.3 to obtain

$$\mathcal{E} \supset \mathfrak{D}\{\mathcal{A}\} = \sigma\{\mathcal{A}\} = \mathcal{G}.$$ □

Remark 2.5. As the reader may have discovered, the proofs of Theorems 2.2 and 2.3 are of this kind.

Remark 2.6. The second half of the theorem is called *Dynkin's π-λ theorem.* □

3 The Probability Space

We now have sufficient mathematics at our disposal for a formal definition of the *probability space* or *probability triple*, $(\Omega, \mathcal{F}, P)$.

Definition 3.1. *The triple $(\Omega, \mathcal{F}, P)$ is a* probability (measure) space *if*

- *Ω is the* sample space*, that is, some (possibly abstract) set;*
- *$\mathcal{F}$ is a σ-algebra of sets (events) – the* measurable *subsets of Ω. The "atoms", $\{\omega\}$, of Ω, are called* elementary events*;*
- *P is a probability measure,*

that is, P satisfies the following Kolmogorov axioms*:*

1. *For any $A \in \mathcal{F}$, there exists a number $P(A) \geq 0$; the probability of A.*
2. *$P(\Omega) = 1$.*
3. *Let $\{A_n, n \geq 1\}$ be disjoint. Then*

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$ □