Chapman & Hall Probability Series

PROBABILITY WITH A VIEW TOWARD STATISTICS

VOLUME I

J. Hoffmann-Jørgensen

PROBABILITY WITH A VIEW TOWARD STATISTICS

VOLUME I

Chapman & Hall Probability Series

Series Editors
Jean Bertoin, Université Paris VI
David S. Griffeath, University of Wisconsin
Marjorie G. Hahn, Tufts University
J. Michael Steele, University of Pennsylvania

Hoffmann-Jørgensen, J., *Probability with a View toward Statistics, Volumes I and II*
Stromberg, K., *Probability for Analysts*

PROBABILITY WITH A VIEW TOWARD STATISTICS

VOLUME I

J. Hoffmann-Jørgensen

University of Aarhus

Springer-Science+Business Media, B.V.

Copyright © 1994 by Springer Science+Business Media Dordrecht
Originally published by Chapman & Hall in 1994.
Softcover reprint of the hardcover 1st edition 1994

All rights reserved. No part of this book may be reprinted or utilized in any form or by any electronic, mechanical or other means, now known or hereafter invented, including photocopying and recording, or by an information storage or retrieval system, without permission in writing from the publishers.

Library of Congress Cataloging in Publication Data

Hoffmann-Jørgensen, J. (Jørgen) , 1929-

Probability with a view toward statistics / J. Hoffmann-Jørgensen
p. cm.
Includes index.
ISBN 978-0-412-05221-7 ISBN 978-1-4899-3019-4 (eBook)
DOI 10.1007/978-1-4899-3019-4

1. Probabilities. 2. Mathematical statistics.
I. Title.
QA273.H689 1994
519.2—dc20 94-18943
CIP

British Library Cataloguing in Publication Data available

Contents

Preface . . . . . *xiii*

Introduction . . . . . *xvi*

References to the History of Probability . . . . . *xxxix*

Chapter 1 : MEASURE THEORY *1*

1.1 Pavings and Set Functions 1
1.2 Stability of Pavings 3
1.3 Contents 4
1.4 Measures 6
1.5 Generation of Algebras and σ-Algebras 7
1.6 Sierpińsky's Stability Lemma 8
1.7 The Uniqueness Theorem I 10
1.8 The Borel σ-Algebra 11
1.9 Generation of the Borel σ-Algebra 12
1.10 The Borel σ-Algebra on $\mathbf{R}$ and $\mathbf{R}^k$ 13
1.11 Finite σ-Algebras 14
1.12 The σ-Algebra Generated by Singletons 15
1.13 The Counting Measure 16
1.14 Measures on $(\mathbf{N}, 2^{\mathbf{N}})$ 16
1.15 The Dirac Measure 17
1.16 Construction of Algebras 17
1.17 The Borel σ-Algebra Revisited 18
1.18 Construction of σ-Algebras 19
1.19 Null Sets and Almost Sure Properties 20
1.20 Operations on Measures 21
1.21 Set Functions 21
1.22 The μ-Measurable Sets 23
1.23 Carathéodory's Theorem 23
1.24 Additivity of Inner and Outer Measures 25
1.25 Tightness 27
1.26 Criteria for Inner and Outer Tightness 28
1.27 The Extension Theorem 29
1.28 Completion of Measures 30
1.29 Stability of Compact Pavings 31
1.30 The Compactness Lemma 32
1.31 Inner Tight Set Functions on Compact Pavings 34
1.32 Lebesgue-Stieltjes Measures on $\mathbf{R}$ 36
1.33 Lebesgue-Stieltjes Measures on $\mathbf{R}^k$ 37
1.34 Measures on Metric Spaces 38
1.35 Metric Outer Measures 41
1.36 Construction of Metric Outer Measures 41

1.37 The Uniqueness Theorem II 44
1.38 Measurable Functions 45
1.39 Borel Functions 46
1.40 Real-Valued Functions 47
1.41 The Measurable Limit Theorem 48
1.42 The σ-Algebra Generated by Functions 48
1.43 The Product σ-Algebra 49
1.44 Image Measures 51
1.45 Simple and Discrete Functions 52
1.46 Approximation of Measurable Functions 52
1.47 The Image Catastrophe 53
1.48 Functions on Product Spaces 54
1.49 The Functional Sierpińsky Lemma 54
1.50 The Projective Limit Theorem 58
1.51 References to Chapter 1 60
1.52 Exercises to Chapter 1 61

Chapter 2 : PROBABILITY MEASURES 100
2.1 Probability Spaces 100
2.2 Conditional Probabilities 101
2.3 Independence 102
2.4 Independence Criterion for σ-Algebras I 104
2.5 Independence Criterion for σ-Algebras II 106
2.6 Independence Criterion for σ-Algebras III 107
2.7 Random Variables, Vectors and Functions 108
2.8 Distribution Functions and Laws 109
2.9 Independent Random Functions 112
2.10 Independence Criterion for Random Functions 112
2.11 The Two Borel-Cantelli Lemmas 113
2.12 The 0–1 Law for Pavings 114
2.13 The 0–1 Law for Random Functions 114
2.14 Probability Mass Functions 116
2.15 Lottery 117
2.16 Pairwise Independence 118
2.17 Maximal Inequalities 119
2.18 Maximal Inequalities for Subadditive Schemes 120
2.19 A Factory Model 122
2.20 Fake Infection 123
2.21 Nuclear Safety 124
2.22 Waiting Times 125
2.23 Order Statistics 127
2.24 Uniform Distributions 129
2.25 Random Intervals 130
2.26 References to Chapter 2 131
2.27 Exercises to Chapter 2 132

Chapter 3 : INTEGRATION *152*
3.1 Integrals of Positive Functions 152
3.2 Integrals of Real Functions 156
3.3 The Calculus of Integrals 158
3.4 Monotone Convergence Theorem 159
3.5 Fatou's Lemma 160
3.6 Lebesgue's Dominated Convergence Theorem 161
3.7 Beppo Levi's Theorem 162
3.8 Integrals of Simple and Discrete Functions 163
3.9 Markov's Inequality 164
3.10 Jensen's Inequality 165
3.11 Hölder's Inequality 166
3.12 Minkowski's Inequality 167
3.13 Continuity of the Integral 168
3.14 The Standard Proof 170
3.15 The First Transformation Theorem 171
3.16 Semi-finite Measures 172
3.17 Indefinite Integrals 173
3.18 Radon-Nikodym-Lebesgue's Theorem 175
3.19 Radon-Nikodym Derivatives 178
3.20 Product Measures 181
3.21 Tonelli-Fubini's Theorem 183
3.22 Convergence in q-Mean 184
3.23 Dunford-Pettis' Theorem 189
3.24 Criteria for Uniform Integrability 191
3.25 Convergence Relations 194
3.26 The Lebesgue Integral on $\mathbf{R}$ and $\mathbf{R}^k$ 198
3.27 The Lebesgue-Stieltjes Integral 201
3.28 Integration by Parts 203
3.29 Substitution 204
3.30 Abstract Integrals as Ordinary Integrals 206
3.31 Continuity and Differentiability of Integrals 208
3.32 Integrals of Complex Functions 210
3.33 Examples of Convergence and Divergence 211
3.34 Failure of Tonelli-Fubini's Theorem 212
3.35 A Singular Integral 212
3.36 The Mellin Transform 214
3.37 The Gamma Function and Related Functions 216
3.38 Sums as Integrals 218
3.39 The Space $L^{(\infty)}(\mu, \mathbf{C})$ 218
3.40 The Uniqueness Theorem III 220
3.41 Measure Transforms 222
3.42 Integrals of Vector and Matrix Functions 225
3.43 References to Chapter 3 227
3.44 Exercises to Chapter 3 228

Chapter 4 : EXPECTATIONS AND MOMENTS *249*
4.1 Expectations 249
4.2 Computations of Expectations 250
4.3 Moments and Metrics 253
4.4 Variance and Covariance 256
4.5 Expectations of Independent Random Functions 258
4.6 Sums of Independent Random Vectors 261
4.7 Convergence of Partial Sums 262
4.8 The Birth of the Law of Large Numbers 263
4.9 Kronecker's and Cesàro's Lemmas 267
4.10 The Law of Large Numbers I 269
4.11 The Law of Large Numbers II 272
4.12 The Law of Large Numbers III 274
4.13 The Borel-Cantelli Lemma Revisited 275
4.14 Transforms of Random Functions 276
4.15 Inversion of Multiplicative Transforms 277
4.16 Distribution Functions 279
4.17 Probability Mass Functions 279
4.18 Characteristic Functions 280
4.19 Laplace Transforms 283
4.20 Moment Transforms 286
4.21 Generating Functions 288
4.22 The Normal Distribution 290
4.23 The Gamma Distribution 295
4.24 The Poisson Distribution 297
4.25 The log-Normal Distribution 297
4.26 The Hyper-Exponential Distribution 299
4.27 A Discontinuous Laplace Transform 301
4.28 The Mellin Transform and Moments 302
4.29 Subgaussian Random Variables 303
4.30 The Subvariance Inequalities 304
4.31 Sign-Symmetric Random Vectors 306
4.32 Khinchine's Inequality 307
4.33 The Law of Large Numbers IV 308
4.34 A Maximal Moment Inequality 308
4.35 A Tail Inequality 310
4.36 Convergence of Partial Sums Revisited 310
4.37 Alternative Expectations 313
4.38 Averages 316
4.39 The Petersburg Paradox 320
4.40 The Anomal Distribution 323
4.41 Bad Luck 326
4.42 References to Chapter 4 326
4.43 Exercises to Chapter 4 327

Chapter 5 : CONVERGENCE IN LAW *354*
5.1 Convergence in Law on Metric Spaces 354
5.2 The Portmanteau Lemma 358
5.3 A Criterion for Convergence in Law 360
5.4 The Calculus of Convergence in Law 364
5.5 Eventual Tightness 367
5.6 The Eventual Tightness Theorem 367
5.7 The Continuity Theorem 370
5.8 Criteria for Eventual Tightness on $\mathbf{R}^k$ 371
5.9 Convergence of Distribution Functions 372
5.10 Convergence of Characteristic Functions 373
5.11 Convergence of Probability Mass Functions 374
5.12 Convergence of Absolute Moments 375
5.13 Convergence of Moments 376
5.14 Convergence of Laplace Transforms 377
5.15 Convergence of Generating Functions 378
5.16 Preview of the Central Limit Theorem 379
5.17 The Lévy-Khinchine Distributions 382
5.18 The Stable Distribution $Stab(\mu, \sigma, p, \gamma)$ 385
5.19 The Upper and Lower Tail Functionals 387
5.20 The General Central Limit Theorem 389
5.21 The Normal Central Limit Theorem I 392
5.22 The Normal Central Limit Theorem II 394
5.23 The Normal Central Limit Theorem III 395
5.24 The Poisson Central Limit Theorem 396
5.25 The Stable Central Limit Theorem 397
5.26 A Review of the Central Limit Theorem 398
5.27 The Inversion Formulas 399
5.28 Infinitely Divisible Distributions 402
5.29 Stable Distributions 403
5.30 Characterization of Stable Distributions 404
5.31 The Mellin Transform of $\exp(-us - z\,s^p)$ 406
5.32 Stable Densities 409
5.33 Normal and Poisson Approximation of $Bi(n, p)$ 411
5.34 Normal Approximation of $Po(\lambda)$ 412
5.35 The Triangular Distribution 413
5.36 Fractiles 414
5.37 Convergence of Fractiles 415
5.38 Convergence of Empirical Fractiles 416
5.39 References to Chapter 5 418
5.40 Exercises to Chapter 5 419

Chapter 6 : CONDITIONAL EXPECTATIONS *440*
6.1 Information and σ-Algebras 440
6.2 Atoms and Information 441
6.3 The Calculus of Atoms 442
6.4 The Factorization Theorem 443
6.5 Separable and Blackwell σ-Algebras 445
6.6 Atomic Determination 446
6.7 Conditional Expectations 447
6.8 The Calculus of Conditional Expectations 451
6.9 Inequalities for Conditional Expectations 455
6.10 Convergence of Conditional Expectations 456
6.11 Conditional Densities 458
6.12 Conditionally Independent Pavings 460
6.13 The Weak Markov Property 461
6.14 The Markov Equivalence Theorem 462
6.15 Conditionally Independent Random Functions 463
6.16 Weak Markov Chains 464
6.17 The Permutation Invariant σ-Algebras 464
6.18 U-Statistics 466
6.19 Conditional Partial Sums 468
6.20 The Distribution of the Order Statistics 469
6.21 Conditional Maximal Inequalities 470
6.22 The Martingale Inequality 472
6.23 Kolmogorov's Inequality 472
6.24 Ottaviani's Inequality 472
6.25 Lévy's Inequality 473
6.26 The Exponential Inequality 473
6.27 Fish Auction 474
6.28 References to Chapter 6 476
6.29 Exercises to Chapter 6 477

Chapter 7 : MARTINGALES *484*
7.1 Filters and Stopping Times 484
7.2 Occurrence Times 487
7.3 The Calculus of Stopping Times 488
7.4 Composite Random Variables 490
7.5 Martingales 491
7.6 Examples of Martingales 494
7.7 Stopped Submartingales 497
7.8 The Martingale Inequalities 498
7.9 The Upcrossing Inequality 500
7.10 The Martingale Convergence Theorem 502
7.11 The Reversed Martingale Convergence Theorem 503
7.12 Lévy's Theorem 505

7.13 Optional Times 506
7.14 The Optional Sampling Theorem 508
7.15 Review of the Optional Sampling Theorem 509
7.16 Criteria for Optionality 509
7.17 Martingale Differences 511
7.18 The Converse Optional Sampling Theorem 513
7.19 Non-Negative Supermartingales 515
7.20 Convex Functions of Martingales 515
7.21 The Ballot Theorem 516
7.22 Rescaling of Scaled Martingales 518
7.23 The Most Favorable Opting Time 520
7.24 An Epidemic Model 523
7.25 Gaussian Martingales 529
7.26 A Water Reservoir Model 530
7.27 Convergence of U-Statistics 532
7.28 References to Chapter 7 535
7.29 Exercises to Chapter 7 536

***PREREQUISITES* 547**
P.1 Events and Subsets 547
P.2 The Calculus of Sets 549
P.3 Images and Co-images 550
P.4 The Real and Complex Numbers 551
P.5 Metric Spaces 555
P.6 Function Spaces 563
P.7 Relations and Orderings 564
P.8 Linear Algebra 565

***List of Symbols* 568**

***Author Index* 572**

***Subject Index* 574**

Preface

Guide to the teacher: This book is intended to serve as a graduate text for students majoring in probability and/or statistics. It doesn't require knowledge of any particular results from probability or statistics; however, a general knowledge of the elementary notions as supplied by any standard course in elementary probability theory or statistics will be helpful. It will be assumed that the students know elementary linear algebra and analysis including metric spaces, for instance, Rudin's book *Principles of Analysis*. For the convenience of the reader I have collected the prerequisites from linear algebra and analysis in the appendix "Prerequisites". The book is designed in such a way that it may be used for a variety of courses in probability and statistics or in pure probability theory; and it is divided in two volumes. The present volume treats the basics of modern probability theory including measure and integration theory, as described in more detail below. The second volume is aimed at applications to statistics and may be omitted in a pure probability course. As indicated, the book is a textbook in probability theory with emphasis on applications to statistics, which means that I have included parts of probability and measure theory which usually cannot be found in a textbook of pure probability.

Contents of Volume I: The introduction contains a brief account of the history of probability theory and statistics. In the text you will find more about the history of specific subjects, for instance, the normal distribution, the central limit theorem, averages and the ruin problem. I have always felt that the history of probability and statistics is important for a deeper understanding of probability, which has been neglected in many modern textbooks, and that it helps the students to get a better understanding of the important results.

Chapter 1 and 3 contains a complete course in measure and integration theory, which ever since Kolmogorov's book *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933) has been the indispensable foundation of a rigorous course in probability or statistics. The form and results are chosen to fit the needs of probability theory and statistics best possible. For instance, the Sierpińsky lemmas (1.7) and (1.49) and the associated uniqueness lemmas (1.7), (1.49), (3.40) and (4.15) are stated in a form which fits well for applications to independence criteria, the moment problem, the uniqueness theorem for transforms, Markov chains, conditional expectations etc. Similarly, the extension theorem (1.27) is stated in a form which facilities applications to Lebesgue-Stieltjes measure, Kolmogorov's consistency theorem, the continuity theorem, regular conditional distributions, etc. If your students already have taken a course in measure and integration theory, Chapters 1 and 3 may be omitted.

Chapter 2 is an introduction to rigorous probability with independence as the central theme. It contains various criteria for independence, the classical Borel-

Cantelli lemmas, the 0–1 law, and a maximal inequality needed in the proof of the law of large numbers.

Chapter 4 has two major themes: the law of large numbers and transforms of random variables (characteristic functions, Laplace transforms, moments, generating functions, Mellin transforms, etc.). The law of large numbers, see (4.5)–(4.12) and (4.33), is proved under various moment conditions with and without independence. The chapter also contain studies of the most common distributions with the normal distribution (one dimensional and multidimensional) as the leading example.

Chapter 5 contains a thorough study of the notion of convergence in law for random variables taking values in a metric space. It contains a general continuity theorem for transform; see (5.11) with the classical results for characteristic functions, Laplace transforms and moment transforms as special cases. The last part of the chapter is devoted to the study of the central limit theorem, stable and infinitely divisible distributions and the Lévy-Khinchine representation.

Chapter 6 has conditioning as its central theme. It is devoted to conditional expectations, and I have stressed the interpretation of the conditioning σ-algebra as the information available, and the interpretation of the conditional expectation as the "best" possible estimate of the random variable given the information. It contains a discussion of conditional independence with application to (weak) Markov chains, and it is shown how U-statistics can be described as conditional expectations.

Chapter 7 is devoted to the study of discrete parameter martingales. In (7.1)–(7.14), I pass through the basic theory of martingales with the martingale convergence theorem [see (7.10+11)] and optional sampling theorem [see (7.14)] as the major results. Sections (7.19)–(7.27) contain a series of applications theoretical (U-statistics, non-negative supermartingales and Gaussian martingales) as well as practical (ballot, epidemic, water reservoir and stock market).

The exercises contains a variety of problems ranging from simple exercises which teach the student to operate with the theoretical notions, to extensions of theory, often in a series of exercises aided by hints which teach the student how to derive new theorems from old ones, and which can be used as a basis for talks delivered by the students.

Contents of Volume II: Volume II is, as mentioned earlier, directly aimed at applications to statistics, for instance, the art of calculating densities of complicated transformations of random vectors (Chapter 8), exponential models (Chapter 12), consistency of maximum estimators (Chapter 13) and asymptotic normality of maximum estimators (Chapter 14). But it also contains parts of a pure probabilistic nature, such as stochastic processes (Chapter 9), regular conditional probabilities, (strong) Markov chains, random walks (Chapter 10), and optimal stopping strategies in random games (Chapter 11).

Guide to the student: You will notice that I have avoided the usual composition: Definition 4.1, Lemma 4.2, Theorem 4.3, Corollary 4.4, etc. Instead, lem-

mas, theorems and corollaries all carry names, such as "the uniqueness theorem," "Lebesgue's dominated convergence theorem," "the optional sampling theorem," etc. It is my experience that it is easier to remember names than numbers, and that a reference by name such as "the martingale convergence theorem" may save the work of looking for the result. However, to make it easier to find the correct reference all definitions, lemmas, theorems and corollaries carry two numbers, and all formulas carry three numbers. For instance, "the Sierpińsky lemma (1.7)" refers to Section 7 of Chapter 1 named "the Sierpińsky lemma," and "(6.8.2)" refers to Formula 2 of Section 8 of Chapter 6.

Acknowledgment: I would like to thank my colleagues and students for many valuable suggestions and assisting with the proofreading. In particular, I am indebted to S. Ladouceur, F. S. Larsen, G. Peskir and M.-C. Weber for their careful proofreading, and to K. Boddum and E. Yndgaard for their efficient typing of the first draft of the manuscript.

Introduction

The Origin of Probability

Probability owes its existence to gambling. It was game-theoretical problems, which was the main motivation for the pioneers of probability, and the notions of odds and numerical probabilities arose from the study of chance games. But let me begin the story at its beginning. The beginning seems to be a small bone, called *astragalus*, lying above the heel bone, called *talus*.[1] The astragalus from sheep has form and size, which makes it suitable for dice games. The astragali is reported to be much more numerous (10–50 times) than other kinds of bones in prehistory archeological excavations and, moreover, many of the prehistoric astragali show clear signs of being used for some purpose. Since there is clear evidence that the astragalus was commonly used as a die from 3500 B.C. until 1600 A.D.,[2] it is a fair guess that primitive man used the astragalus for some kind of dice game. The earliest evidence of gambling with astragali is an Egyptian tomb painting from 3500 B.C. showing a nobleman and his wife using an astragalus in a board game, and in the tomb of Reny-Soube at Thebe in Upper Egypt archeologists have found a complete board game[3] containing three astragali. From the antique Greek and Roman authors, we know that the astragali was diligently used in many different kinds of games.

The astragalus is asymmetric and contrary to a die it only has four sides on which it can rest. According to Greek and Roman authors the four possible sides of the astragalus was numbered: 1, 3, 4 and 6. Two of the sides (1 and 6) are narrower than the two others and have probability $\approx$ 10% of showing, and the two remaining sides (3 and 4) have probability $\approx$ 40% of showing. Four astragali was often used in a simple game called *rolling of the bones*. The worst throw (called *the dog*) in this game was four 1's and the best (called *Venus*) was 1, 3, 4, 6. In Table 1 you will find the probabilities of the different throws in the game "rolling the bones. Note that "the dog" has the smallest probability, and that the probability of getting "the Venus" is 384 times more probable than getting "the dog". Thus, the assignment of the values differs highly from the probabilities.[4] The Roman author Propertius (50 B.C. – 15 A.D.) writes: *I was seeking the Venus with favorable tali,*

---

1 The astragalus and the talus are two different bones, but in the literature they have been used indiscriminately to mean the astragalus bone.

2 The astragalus is still in use on the Faroe Islands for gambling.

3 Named "hounds and jackals" by the archeologists.

4 In the distant past as now people hated to lose and that may be reason that the highly improbable "dog" was claimed to be the worst possible throw.

but the dammed dog always leapt out. Because "the dog" comes approximately 1 out of 10,000 times, the good Propertius must have been very unlucky that day.

Table 1

Probabilities in the game "rolling of the bones"				
Throw				Probability
1111	6666			1×10^{-4}
3333	4444			256×10^{-4}
1113	1114	6663	6664	16×10^{-4}
3331	3336	4441	4446	256×10^{-4}
1116	6661			4×10^{-4}
3334	4443			1024×10^{-4}
1133	1144	6633	6644	96×10^{-4}
3344				1536×10^{-4}
1166				6×10^{-4}
1134	6634	3316	4416	192×10^{-4}
3314	3364	4413	4463	768×10^{-4}
1163	1164	6613	6614	48×10^{-4}
1346				384×10^{-4}

The astragalus was popular in the ancient Greece and Rome, and then as now the authorities tried to control gambling in various ways. The Roman Emperor Augustus (63 B.C. – 14 A.D.) banned gambling except for the Saturnalia (which is the month December). However, he himself was the first to break the ban. In a letter to the later Emperor Tiberius (42 B.C. – 37 A.D.) Augustus writes: *I dined dear Tiberius with the same company, . . . We gambled like old men during the meal both yesterday and today, for when the astragali were thrown whoever turned up the dog or the six set a denarius in the pool for each of the one of the astragali, and the whole was taken by anyone who threw the Venus.*

The most popular gambling tool of all times is the die. The first die might have been made by carving astragali to a more symmetric shape. The earliest known die dates back to $\approx$ 3500 B.C. The ancient dice seems to have been numbered 1–6

in a random way; however, from the 18th Dynasty (Egypt, 1370 B.C.), die makers settled on the three 2–partitions of 7, that is, 1 opposite 6, 2 opposite 5, and 3 opposite 4. There is no evidence that the Greek or the Romans made any attempt to compute odds or had a notion of quantitative or even qualitative probability, in spite of the fact that they had the necessary knowledge of mathematics to do so. Actually most of the ancient dice found are clearly skew, and it seems as if there were no attempts to make an "honest" die. There may be many reasons for this. First, due to the skewness of the astragalus it would be difficult to find regular patterns; second due to the number system of the Greeks and Romans the elementary arithmetic (in particular, multiplying and dividing numbers) was a difficult task only known to experts; the third and in my opinion the most probable reason is that there were no notion of randomness in the sense we have today. Serious gamblers then as now don't believe in randomness, but believe the die is governed by the Gods or Destiny or Luck. Actually in many religions the deity is supposed to reveal his (or hers) will in some holy random experiment, such as drawing lots or picking a white or black ball at random from an urn.

The medieval Europeans gambled eagerly with astragali, dice, and various board games, and the Church and State tried to ban or restrict gambling with the same negative result as today. For instance, the English crusaders in 1190 with rank below Knight were banned from gambling and knights were allowed to lose at most 20 shilling per day in gambling. When Bishop Wibold of Cambray ($\approx$ 960 A.D.) couldn't prevent the monks from gambling, he invented a moral game of dice. Three dice were the thrown, and to each of 56 possible outcomes he assigned a virtue, and the monk was supposed to exercise the assigned virtue for a certain length of time. The game never became popular, but it is interesting because of the fact that here we find the first recorded attempt at counting the total number of outcomes in a random experiment. In a Latin poem called *De Vetula*, which usually is ascribed to the Chancellor of the Amiens Cathedral, Richard de Fournival (1200 – 1250), we find the complete count of all possible outcomes in a random game. Again three dice are thrown, and the number of points is counted, in a table to "De Vetula" we find a lists of the number of ways that the sums 3, 4, 5, . . . , 18 can occur in one throw with three dice (called Cadentia, see Table 2), together with a list of the number of ways that the numbers 3, 4, 5, . . . , 18 can be written as a sum of three numbers between 1 and 6 (called Punctatura, see Table 2). It is uncertain why the table was made and there is no evidence, that the numbers "Cadentia" have been used for computing odds or probabilities. But from a manuscript, we know how the numbers were computed: All the 56 possible outcomes of one throw with 3 dice were listed, and a simple count of the appropriate numbers was made.[5] Note

[5] Pascal's triangle, that is, the binomial coefficients was known in Europe from $\approx$ 1300 A.D., and long time before that by the Arabs and Chinese. However, the relation between the binomial coefficients and combinations was first observed by B. Pascal in 1654, and what we nowadays call Pascal's triangle was called the arithmetic triangle.

that the unknown author of Table 2 has observed the symmetry between 3 and 18, 4 and 17 etc., and that the sum of the punctaturas equals 56, which is the number of possible outcomes discarding permutations, and the sum of the cadentias equals 216, which is the total number of outcomes including permutations.

Table 2
From the poem: *De Vetula*

3	18	Punctatura	1	Cadentia	1
4	17	Punctatura	1	Cadentia	3
5	16	Punctatura	2	Cadentia	6
6	15	Punctatura	3	Cadentia	10
7	14	Punctatura	4	Cadentia	15
8	13	Punctatura	5	Cadentia	21
9	12	Punctatura	6	Cadentia	25
10	11	Punctatura	6	Cadentia	27

Punctatura: The number of ways that a number can be written as a sum of 3 integers between 1 and 6.
Cadentia: The number of ways that a number can be obtained as the sum of points in one throw with 3 dice.

At the end of the 15th century some sort of notion of qualitative probability seemed to emerge. In a commentary to De Vetula, the unknown author claims that the sums 10 and 11 are more likely to show than the sums 3 and 18, but there is still no evidence of a quantitative notion of probabilities or odds. The first recorded probabilistic problem is found in the book *Summa de Arithmetica, Geometria, Proportioni e Proportionalitá* by Luca Pacioli (1445–1509). Luca was born in Toscana near the village Borgo. As young man he studied mathematics in Pérouse, Rome, Naples, Pisa and Venice, and later he became professor in mathematics in Milan and Firenze. He didn't contributed in any significant way to mathematics, but in his book *Summa de Arithmetica, Geometria, Proportioni e Proportionalitá* (1494) he summarized virtually all mathematics known at that time, and the book became a standard textbook of mathematics for a long period of time. The book contains a wealth of exercises, and many exercises of modern elementary mathematics textbooks are descendants of the exercises in *Summa de Arithmetica,*

Geometria, Proportioni e Proportionalitá. In a section entitled "Various Exercises" we find the following exercise, see also Exc.2.53:

> Two men A and B are playing "balla" (probably a medieval ball game). Both have set 10 gold coins at stake, and the one who first wins 6 games gets the 20 gold coins at stake. However, at a point where A has won 5 games and B has won 3 games, the play is interrupted. It supposed that the two men are equally skillful; how should they divide the stake?

Luca gave the answer 5:3; in 1556 N. Tartaglia (1499–1557) objected to the solution and proposed the answer 2:1; two years later G. F. Peverone claimed that the answer rightfully should be 6:1; and in 1654, B. Pascal gave the correct answer 7:1. Luca didn't considered the exercise as a problem of probability, but as a problem of the theory of proportions. It is striking that it should take 160 years for the foremost mathematicians to solve an exercise, which any student with minimal knowledge of probability nowadays can solve in a few minutes. The story shows that the notion of probability is a modern notion, which simply was not available – not even in an intuitive or informal form – in the days of Luca Pacioli.

The Emergence of Probabilities

Around the year 1500 some notion of qualitative probabilities had emerged and around 1550, Gerolamo Cardano wrote a little book *Liber de Ludo Alea* (The Book on Dice Games), in which he develops a theory of odds and probabilities. The two major results of the book are:

- The probability that one of two exclusive events occurs equals the sum of their probabilities
- The probability that two independent events occurs simultaneously equals the product of their probabilities

The book appears more as a diary than as a book. At the beginning of the book Cardano has several wrong forms of the two propositions above; for instance, he claims that since the probability of getting "1" in one throw with a die equals $\frac{1}{6}$, then the probability of getting at least one "1" in two throws with a die must be $\frac{2}{6}$. But later he realize the mistakes and gives the correct answers without removing the false statement. The book was never published, actually Cardano

kept it secret,[6] and the book was first discovered more than hundred years after his death. But at that time the theory of probabilities had been rediscovered by P. Fermat, B. Pascal and C. Huyghens. Cardano had a stormy life of which you will find a brief account of in the following.

Gerolamo Cardano (1501–1576): Gerolamo Cardanó was born the 24th of September 1501 in the Italian city Pavia. He was the illegitimate son of Facio Cardano and Chiara Michena.[7] Facio was a lawyer and he practiced medicine in Milan; he was also known as good geometrist,[8] and he taught geometry at the university of Milan. Chiara is described as a short, fat, healthy woman with a lively wit and a quick temper. Facio kept her as mistress in a separate house, but in 1506 he move into the house of Chiara, and he married her shortly before his death.

We have a most detailed description of the life of Gerolamo Cardano. Partly from his strange and extremely frank (and often self-contradictory) autobiography *De Vita Propia*. He was a man of great capacities and of great contradictions; he had a stormy life, and he was a true exponent of the new Renaissance man.

Facio instructed his son in mathematics, but Gerolamo insisted on studying medicine. Facio was opposed to Gerolamo's plans, but Chiara supported her son, and as usual the woman got her way. So after a violent quarrel with his father, Gerolamo went to the University of Pavia to study medicine in 1520. They were uneasy days; in 1521 King Francis I of France and Emperor Charles V of Spain started a war in Italy.[9] Due to the unrest, Cardano first moved to Milan and later to Padua. He became more known for his diligence at the gambling table than his diligence at the studies, and we know that he succeeded in losing the inheritance from his father at the gambling table in a short time. According to his autobiography he was elected Rector at the University of Padua in 1525 while still a student.[10]

After having failed twice he became a doctorate of medicine in 1526. His exam was so poor that he could not become a member of the medical societies of Milan, Pavia or Padua.[11] In those days membership of the Medical Society was necessary for practicing medicine, so in 1527 he went to a small village Sacco to practice there. Here he fell in love with Lucia Bandarina, the daughter of the local

---

6 Cardano was a passionate gambler, and since he was the sole person in the world having a notion of odds and the capacity of computing them, it gave him a significant advantage over his opponents, and a good reason to keep his findings secret.

7 Cardano hotly denies his illegitimacy, even though it was quite common and socially accepted.

8 According to some sources, Leonardi da Vinci occasionally sought the advice at Facio on geometric matters

9 France and Spain was the two great powers at the time of Cardano and, as today, the great powers preferred to fight their wars on other countries territory.

10 The claim seems unlikely, it was not unusual that a student was elected as Rector, but usually the Rector was a person of a high social rank, and according to the archives of the University of Padua no Rectors was elected in the period 1509–1527.

11 Cardano ascribes this to his illegitimacy, but again the claim seems unlikely.

innkeeper, and they married in 1531. In 1532 he tried again to enter the Medical Society of Milan, but again he was rejected, and he went to another poor village Gallerat to practice. He shared his spare time between the gambling table[12] and the study of mathematics. In 1534 he was ruined, and he moved into the poorhouse of Milan together with his wife and his newborn son, Giambatista (1534–1560). At this low point of his life his luck changed, as so many other times in the years to come. Through the influence of good friends of his father he obtained a lecturing position in geometry at the University of Milan. At this point, his mother Chiara pardoned his loss of the inheritance, and she moved into the house of Gerolamo. Soon after, Cardano wrote his first book on mathematics (1539) and in his lifetime he wrote 150–200 books on all kinds of subjects, for instance, medicine, law, mathematics, philosophy, pedagogics,[13] etc. Being extremely tenacious, Cardano finally was accepted as a member of the Medical Society of Milan. And later in his life he became known all over Europe partly for his mathematical and mainly for his medical skills.[14]

The time from 1547–1557 was a happy time for Cardano with prosperity and fame, but in 1557 luck changed again. His eldest son Giambatista married a prostitute of Milan. The marriage was unhappy; in 1560 Giambatista had a fierce quarrel with his wife, at which she told him, that he was not the father to any of their three children. After the quarrel Giambatista went down to the kitchen, where he made a bread stuffed with white arsenic; he served the bread to his wife and her family. The wife died and her family was close to join her. Giambatista was soon caught, and together with a servant, who assisted with the baking, he was accused of murder. He claimed himself not guilty, and his father Gerolamo, who also was practicing law, defended his son, but in vain. On a gloomy November day in 1560 Giambatista was executed for the murder of his wife. This was a serious blow for Gerolamo, and so much more because Giambatista was his favorite son. Gerolamo had three children, two sons Giambatista and Aldo, and one daughter Chiara. His younger son Aldo (1543–??) behaved very badly from a young age and became involved in a criminal gang at the age of about 13. At the age of 18, Aldo had already payed several visits to several Italian prisons. Shortly after the execution of Giambatista, Aldo's gang visited Gerolamo, locked him up in the cellar and robbed the house. In 1562 Gerolamo could no longer bear his reputation as the father of a murderer and a notorious outlaw, and he moved to Bologna, where good friends had offered him a professorship at the University of Bologna. But the suffering and

12 According to his autobiography he earned an essential part of his income by gambling. Since he was ruined many times in his life, he couldn't possible be as lucky as he claims in his autobiography.

13 Cardano wrote a book on how to raise you sons properly; however, then as now theoretical pedagogues seems to have difficulties with the practical raising of their own children. Of the two sons of Cardano one was executed for murder, and the other became a notorious outlaw and gang leader.

14 In 1552 he was called to Scotland to treat the Archbishop, John Hamilton, for astomia. Cardano cured the Archbishop telling him to avoid feathers of all kinds and to sleep with furs instead of feather linen. The cure worked, and the Archbishop became a loyal friend of Cardano for the rest of his life.

the depressions over his two sons had brought him to madness, and in his last years he was insanely paranoid, spending all his time fighting real but mostly imaginary enemies. His lectures at the university was avoided by the students because of his strange behavior. He was expelled from his position, accused of unnatural sexual behavior with minor boys. This might have been an excuse from the university to get rid of an old, paranoid, strange man, but it seems that his daughter Chiara believed the accusations telling him in a letter, that she was ashamed of his conduct. This brought Gerolamo still further into madness.

But fate had reserved a last blow for Gerolamo. On October 6, 1570 he was arrested by the Inquisition,[15] accused of heresy. The cause for the accusation has never been revealed[16], but probably it was an act attributed to the counterreformation movement led by the severe and orthodox Pope Pius V. In Gerolamo's books there were certainly many things which could displease a strict Inquisitor. Gerolamo Cardano was not a religious man, and he had engaged himself in may different kinds of divinations, for instance, astrology, chiromancy,[17] etc. After a few months in prison, Gerolamo was released, but he was banned from publishing and writing. His autobiography, which was concluded in secret after his release, contains no hints about the trial and the exact charges. He spend his last years in the house of his former student Rodolfo Silvestri. Finally, death released him from his sufferings in 1576.

In his lifetime he published 131 books, and after his death 111 more manuscripts were published, and it was said that just before his death he burned another 170 manuscripts. He engaged himself in all kind of sciences, and there is no doubt that he was a genius, but he had a bad reputation of plagiarizing others peoples' work or inventions without due reference. He probably did not invent the cardan shaft, but he had a horse carriage with a cardan shaft made as a present for the Spanish military governor Alfonso d'Avalos Machese; nor did he invent the suspension of the ship compass, which today bears his name, but he did invent a new way of raising a sunken vessel, a new kind of chimney improving the draft, a new method of distilling wine to make better brandy, and many other ingenious inventions. He was always filled with new ideas, but often the ideas were more theoretical than practical.

The Rise of Probability Theory

As mentioned in the previous section Cardano kept his findings about probabilities secret, and for a period of almost 100 years after the death of Cardano

15 The Inquisition was a religious court for the Catholic Church. It was responsible for the execution (usually by burning) of numerous persons who were found guilty in heresy.

16 The court trials at the Inquisition was usually kept secret.

17 Chiromancy is the art of telling the future of a person by reading the lines of the hand.

there were no attempt to make a general theory. In this period we only find few scattered investigations of particular probabilistic problems. The real breakthrough came in the summer of 1654, where Pierre de Fermat and Blaise Pascal laid down the foundation of elementary probability theory in a series of letters. All the letters except the first (from Pascal to Fermat) have been saved. The second letter is undated and starts as follows:

> *Monsieur*
> *Si j'entrepends de faire un point avec un seul dé en huit coups . . .*

(In English: *Sir, If I try to make a certain score with a single die in eight throws . . .*). The letter is an answer to a lost letter from Pascal to Fermat in the letter Fermat pointed out a mistake of Pascal, and in the response (dated July 29) Pascal acknowledged the mistake and reported a series of solutions to a variety of probabilistic problems. The correspondence went on all summer and fall until October 27, and in the letters Fermat laid down the theoretical foundation for elementary probability theory, which is the lottery model described in (2.15), and Pascal applied the theory of Fermat to solve a large variety of problems. For instance, in a letter dated August 24 Pascal solves the problem of Pacioli mentioned above, and he generalizes the result to an arbitrary number of plays interrupted at an arbitrary position. In November 1654 Pascal had a bad horse carriage accident, and the correspondence ended at this time. On July 25, 1660 Fermat wrote to Pascal and suggests a meeting, but Pascal responded that he is too ill to travel. It is uncertain what triggered Pascal's investigations into probability theory, he wasn't a gambler, but he knew a notorious gambler, Chevalier de Meré [see Exc.2.54], who may have fed Pascal with problems of chance games.

Pierre de Fermat (1601–1665): Fermat was born in the village Beaumont-de-Lomagne in Gascogne near Toulouse. He studied law at the University of Toulouse, and after his graduation in 1631 he was appointed as "Conseilleur de la Chambre des Requêtes du Parlement de Toulouse", that is, as counsellor at a special court taking care of civil lawsuits. He advanced slowly but steadily, and he finally, became the head of the court. In his spare time he studied mathematics and languages with great success. Considering that mathematics was only one of the hobbies of Fermat, he reached astonishing results, and he was one of the leading mathematicians of his time. Unlike mathematicians before and after, public recognition meant little or nothing to Fermat; most of his findings are contained in his huge correspondence with scientists in France, England, Italy, Germany and The Netherlands. He had only little personal contact with other scientists. He seems to be one of the rare persons who bloom in isolation, and his modesty made him believe that a rough sketch of a proof or the mere formulation of a theorem

were sufficient for everybody to fill in the details.[18] The most famous and difficult problem of mathematics goes under the name *Fermat's last problem.* The problem is the existence of solutions to *the Diophantic equation*: $x^n + y^n = z^n$, where $n \geq 2$ is a given and known integer, and x , y and z are unknown. A Diophantic equation is an equation where only integers are allowed as solutions; it is named after the Egyptian mathematician Diophantus ($\approx$ 250 A.D.) who wrote a book on the subject and found a formula for the integer solutions to $x^2 + y^2 = z^2$ [there are infinitely many solutions, the most well-known is $(3, 4, 5)$]. In a copy of Diophantus' book belonging to Fermat we find the following note in the margin:

> *I have found a wonderful proof that a cube cannot be divided into two cubes, and a biquadrate cannot be divided into two, etc. . . .*

Thus, Fermat claims that there are no integer solutions to the equation $x^n + y^n = z^n$, when $n = 3$ or 4 and possibly also for $n \geq 5$, but due to difficulties in reading the handwriting of Fermat that is, somewhat uncertain. For three centuries the foremost mathematicians have tried to prove the claim of Fermat. In 1770 the claim was proved for $n = 3, 4$, in 1808 the claim was proved for $n = 3, 4, 5, 6$, in 1910 the claim was proved for $3 \leq n \leq 31$, in 1954 the claim was proved for $3 \leq n \leq 619$, in 1982 the claim was proved for $3 \leq n \leq 253,747,889$. Thus, we are progressing, but there is a long way to infinity. The latest news (January 1994) is, that the theorem is claimed to be proved by A. Wiles, but there seems to be several gaps in the proof and at this time the proof is incomplete.

Blaise Pascal (1623–1662): Pascal was a wonder kid with a fragile health. His father, Étienne Pascal, was judge in Clermont-Ferrand, and he is described as a competent mathematician. In 1631 the family moved to Paris to stimulate the education of Blaise; in 1639, only 16 years old Blaise wrote a remarkable paper on conical sections (*Essai pour le coniques*). Two years later he invented the first mechanical computer, which could add, subtract and multiply numbers.[19] In 1646 Blaise and his family converted to a Christian sect called Jansenism.[20] The Jansenists tried to reform the Catholic Church and were strong opponents of the Jesuits, but the sect was finally (1656) declared heretic by Pope Innocent X, and the members were persecuted. Jansen emphasized the "internal life" and he claimed natural sciences as a perversion against the will of God. Blaise was at first not a serious believer of Jansenism, and he continued his scientific work. But in the fall of 1654 he had a serious horse carriage accident, which he took as a sign

---

18 In his last letter of the 1654 correspondence he formulated a series of number theoretical theorems, and writes that he was sure that Pascal easily could fill in the proofs. Pascal couldn't prove any of theorems!

19 There still exists a few samples of Pascal's computer.

20 Named after the Bishop of Ypres, Cornelius Jansen (1585–1638).

of God, and went to the Jansenistic monastery in Port Royal outside Paris, where his sister was a nun. After that Pascal gave up his work with natural sciences.[21]

In 1655 a young Dutchman named Christiaan Huyghens (1629–1695) made a stop in Paris on his way to complete his studies in mathematics and physics at the University of Anger. At that time, the Parisian mathematicians met regularly in the house of Carcavi, and our young man carried a letter of introduction from his mathematics teacher Francis van Schooten to the Carcavi circle. He never met Fermat, who lived in Toulouse, or Pascal, who had withdrawn to the monastery of Port Royal, but he learned about their findings. However, he was only told about the results – not the proofs and arguments. He became intrigued by the problems of chance games, and he wrote a small book on probability theory. The original version was in Dutch, but it was soon translated into Latin, and was published in 1656 bearing the title *De Ratiociniis in Alea Ludo* (How to Reason in Dice Games). The book was warmly received by his contemporary fellow mathematicians, and for half a century it was the unique introductory text to probability theory, and it is fair to say that the book represents the real beginning of probability theory as a mathematical subject. *De Ratiociniis in Alea Ludo* consists of 14 problems with complete solutions. In the first three problems, Huyghens defines *the expectation*[22] of a random experiment [see (4.1)] and he uses the notion in a very effective and ingenious way to solve a series of nontrivial probabilistic problems. For instance, the 14th and last problem of *De Ratiociniis in Alea Ludo* goes as follows:

> If another and I throw in turn and turn about with 2 dice on the condition that I will have won when I have thrown 7 points, and he will win, when he has thrown 6; if I let him throw first find the ratio of my chance to his.

Thus, we have two persons, say A and B, throwing two dice in turn with the rule, that A begins and wins if he get 6 points before B obtains 7 points. Note that the probability of getting 6 and 7 points, respectively, in one throw with two dice equals $\frac{5}{36}$ and $\frac{1}{6}$, respectively. Thus, the probabilities are slightly disadvantageous to A, but A has the advantage of beginning. In the solution, Huyghens applies expectations in a very clever way, and his argument (in a slightly modernized language) goes as follows: Let a be the stake, and let x be expected winnings of B, then $a - x$ is the expected winnings of A. Moreover, let y be the expected winnings of B on the condition that B begins. If A is to throw, he has a probability of $\frac{5}{36}$ of winning immediately, and so $x = \frac{31}{36}y$, and if B is to throw, he has a probability of $\frac{1}{6}$ of winning immediately, and so

21 In 1658 he had a slight fall back. According to his sister he had toothache, and in order to forget the pain he wrote a short paper on the cycloid (that is, the curve described by a fixed point on a rolling wheel). The paper was published under a pseudonym.

22 In Huyghens notation, "the value of my chance".

$y = \frac{1}{6}a + \frac{5}{6}x$. Solving the two equations gives $x = \frac{31}{61}a$ and $a - x = \frac{30}{61}a$. Thus, if p and q are the probabilities that A and B respectively wins, then we find that $p = \frac{30}{61}$ and $q = \frac{31}{61}$, which shows that the advantage of beginning increases A's probability of winning.

Later Huyghens went into physics, where he laid his major scientific work, and he only returned to probability theory a few times in his life. He is best known for his theory of light, which became the theoretical background for the construction of microscopes and telescopes,[23] he also invented the pendulum clock and showed, that the time of one oscillation of a pendulum rolling on a cycloid doesn't depend on its position.[24] In 1669 Huyghens brother Ludovick wrote a letter to Christiaan telling him that he had made a table of life expectations to be used for computing the premiums of a certain kind of life insurance called "tontine". In his reply Christiaan pointed out the difference between the expected and probable lifetime, and that it is the latter which is needed for computing the premium of life insurance.[25]

The Origin of Statistics

The original meaning of the word "statistics" is "science of states", and in its early existence it was also called "political arithmetic". Like probability theory, statistics owe its existence to a small object – much smaller than the astragalus – but in spite of it smallness it is the most vicious creature who had ever existed on this earth. It is called *Yersina Pestis*, and it is the bacteria which causes the most infectious and lethal decease of all times – *the plague*. In 1346 A.D. the first plague epidemic hit Europe. It arrived at Sicily from the Middle East and it moved on to all of Europe (including remote parts like Iceland and Greenland), and when it finished in 1352 it had killed approximately one-third of the population of Europe (the estimated death toll is 50,000,000 persons). After that time the plague occurred regularly until about 1712. In Table 3 you find the years of the plague in Copenhagen and London. The last outburst of plague in Copenhagen occurred in 1711; in Table 7.1 you will find a table of the weekly mortality in Copenhagen. In all 22,535 persons out of a total population of approximately 60,000 died in Copenhagen that year, mostly due to the plague. Usually the plague would kill about one-third of the population. The bacteria *Yersina Pestis* has a complicated life cycle, which includes three hosts, rats, rat lice and humans. It is lethal to all three hosts and death is caused by prussic acid produced by the bacteria. Prussic

---

23 The microscope and telescope was invented before that time, but was improved largely by the methods of Huyghens.

24 The first pendulum clocks were modified for the "circular inaccuracy," but it was soon found that the correction was marginal to the other sources of inaccuracy.

25 This is still a common mistake, in particular, in the computations of risk probabilities for dangerous industrial plants such as nuclear power plants and chemical plants.

Table 3

Plague epidemics in Copenhagen and London	
Copenhagen	1349 ? 1405 ? 1536 1546 1583 1601 1625 1630 1637 1655 1665 ? 1711
London	1348 ? 1406 ? 1537 ? 1563 ? 1601 1625 ? 1665 ? ?

The question marks "?" indicate periods where I was unable to find data, but where I am certain that one or more plague epidemics has occurred.

acid is a highly poisonous chemical and causes a blue or dark coloring of the skin; for this reason the plague is also named *the black death.*[26] Around 1712 people recognized the connection between the plague and rats, and from that time we have tried by all means to exterminate the rats,[27] but in vain. However, we have pushed the rats down in the sewers, and thereby stopped the transition of the *Yersina Pestis* from rats to humans.

As mentioned earlier the original meaning of "statistics" is "the science of states". The science of states dates back to Aristotle (384 B.C. – 322 B.C.), whose book *Politeia* contains a description of 158 states. With the emergence of the new ideas of the Renaissance, scientists of states began to collect data such as population number, yearly production of various goods, etc. However, due to lack of real data the numbers was often purely hypothetical or quantitative. The most famous population count of all times took place on Christmas day in the year 1 B.C., when the Roman Emperor Augustus ordered all the inhabitants of the Roman Empire to go to their town of birth to be counted. Population counts even today are a difficult and expensive matter, and it requires a high administrative level, which was present in the Roman Empire, but was lost with its fall, and from 400–1800 there were no real population counts in Europe. Population counts are important for two reasons: They are the basic data needed for estimating the total tax revenue of the state and the size of army that the state can assemble at wartime. In the Middle Ages and the Renaissance, the states of Europe didn't have the administrative capacity to make population counts, but the Church was capable of counting the number of deaths (or rather burials) and the number of newborn babies (or rather baptized babies). The first regular collection of birth and death data goes back to 1538, when the British king Henry VIII introduced the so-called yearly bills of mortality. The year before, The British Isles was hit by the

26 Today the plague is no longer a serious threat to humans; most of the known antibiotics will kill the Yersina Pestis and cure the person.

27 It is ironic that the human race today are exterminating approximately one species per day without really trying, but the rat which we have tried so hard to exterminate is surviving without any sign of being an endangered species.

plague, and Thomas Cromwell, the real ruler of Britain at that time, ordered the Church to make a count of burials and baptized babies, probably with the purpose of estimating the future military and tax revenue potential. At the beginning these counts were irregular and disorganized, but from 1603 and onward we have regular and reasonably precise counts from the city of London. In Denmark such counts were performed from $\approx$ 1600, but the first reliable counts covering the whole country dates back to $\approx$ 1700 [see Table 7.1].

Political arithmetic: Statistics (or political arithmetic as it was named in its first years of existence) is the art of deducing estimates and properties of quantities which can not be observed directly. The inventor seems to be an English tradesman and haberdasher of small wares, John Graunt (1620–1674). In 1661 he published a book bearing the title *Natural and Political Observations upon the Bills of Mortality*, where he reported about his studies of the bills of mortality. He applied the bills to estimate the birth mortality, the number of inhabitants in London, the number of years to recover the former population level after a plague epidemic, life-tables (that is, a table of the probabilities that an x-year-old person will survive his or hers next birthday), etc. His methods are bold but dubious; however, many of his findings are in accordance with later and more reliable observations. For instance, he found that 36% of all children died before their sixth year birthday – a result which were confirmed by later and more precise computations. In 6 different ways he found that the population of London were $\approx$ 380,000. To give you an idea of his methods, I describe two of John Graunt's techniques to obtain the population number of London:

From the Bill of Mortality for London in the year 1660 John Graunt knew three facts. (1): 13,000 persons died in London in 1660; (2): 12,000 babies were born in London in 1660; (3): Of the 13,000 deaths in London that year, 3,200 occurred inside the walls, that is, in central London (at that time the Bills of Mortality also contained information about the number of burials in the individual parishes of London). Studying a map of central London gave him an estimate of 11,800 houses inside the walls. In his first method, John Graunt assumes that to each house there corresponds a family of 8 persons (man, wife, 3 children and 3 servants), and he thereby obtained the number 380,160 in the following way:

$$\text{Number of families inside the walls} = 11,800$$

$$\text{Total number of deaths} = 13.000$$

$$\text{Number of deaths inside the walls} = 3,200 \approx \tfrac{1}{4} \cdot 13,000$$

$$\text{One family} = 8 \text{ persons (man, wife, 3 children, 3 servants)}$$

$$\Downarrow$$

$$\text{Total population number} = 8 \times 4 \times 11,800 = 380,160$$

In his second method, John Graunt assumed that a woman in the age between 16 and 40 years old is pregnant approximately every second year, and that the number

of women between 16 and 76 years old is twice the number of the women 16 and 40. Moreover, assuming that to each woman between 16 and 76 there corresponds a family of 8 persons (man, wife, 3 children and 3 servants) he obtained the number 384,000 in the following way:

$$\begin{aligned}
&\text{Total number of newborns } = 12,000\\
&\text{One family } = 8 \text{ persons (man, wife, 3 children, 3 servants)}\\
&\qquad\qquad\qquad\qquad \Downarrow\\
&\text{Number of women aged 16-40 } = 2 \times 12,000 = 24,000\\
&\qquad\qquad\qquad\qquad \Downarrow\\
&\text{Number of women aged 16-76 } = 2 \times 24,000 = 48,000\\
&\qquad\qquad\qquad\qquad \Downarrow\\
&\text{Total population number } = 8 \times 48.000 = 384,000
\end{aligned}$$

No matter how shaky and dubious the methods of Graunt may seem retrospectively, his investigations were a breakthrough in statistics and the art of estimating quantities, which cannot be observed directly, and his work received due recognition. Soon the probabilists of the time (for instance, Abraham de Moivre, Daniel Bernoulli and Edmund Halley) took up the challenge to invent rigorous methods, particularly to compute reliable life-tables, which were an important tool for the emerging life insurance companies.

Insurance: In 1492 Christopher Columbus discovered the continent of America[28] and, soon after, the sail routes south of Africa and America to India, China and Japan were discovered. In the Middle Ages trade with India, China and Japan followed long and dangerous caravan routes through Asia to the ports of the Middle East, where the goods were shipped to the Italian city-states (Venice, Firenze, Genova etc.). The new sail routes cut the cost of transport immensely, and they were very profitable but also very risky. The high risk called for insurance, and the first marine insurance was established in the beginning of 16th century in Northern Italy. Originally, insurance was an ordinary bet; for instance, a merchant sending a ship on an intercontinental voyage could make a bet with his fellow merchants that his ship would not return in safety. If the ship was lost he would thereby get some compensation, and if the ship returned in safety, he would have to share some of his profit with the bet holders. But soon insurance was organized by companies. In 1601, the British Queen Elisabeth I passed the first law of insurance, and in 1688 Edward Lloyd opened a coffee house on Tower street, London. There as in other coffee houses, merchants, seafarers and insurance brokers assembled

[28] Or rather the sail route from Europe to America, the continent was, of course, discovered centuries before by the Indians.

to drink coffee and to make business. From its informal beginning, Lloyd's coffee house grew in repute and influence, and from 1700 on he published *Lloyd's News*, giving information of ship movements and other information of interest to marine insurance brokers. As time went by, the publication grew into what nowadays is entitled *Lloyd's Register of Shipping*. Today, *Lloyd's* is a insurance market with about 5,000 *underwriters of Lloyd* (that is, insurance companies associated to Lloyd's) and about 2,000 *insurance brokers* (that is, companies which on the behalf of the policy holders seeks the best possible insurance contract). Life insurance came into existence at the same time as marine insurance. At the beginning, life insurance like marine insurance was ordinary bets, but soon it was organized in life-companies. In the 17th and 18th centuries *the tontine*[29] was a popular form of life insurance. The tontine policy provided a yearly refund at the end of a definite period such as 15 or 20 years until the death of the policy holder, and usually the premium was paid in total at the signing of the policy contract. It was frequently used by cities to raise money for some definite projects. In the middle of the 18th century fire insurance was added to the options of insurance.

It is of vital interest for an insurance company to be able to compute reliable risk probabilities, such as "the probability that a cargo ship bound for China will return in safety" or "the probability that a man aged 35 years will survive his 70th birthday". Inspired by the works of John Graunt, scientists of the time tried to construct more rigorous methods of computing reliable life-tables. Since 1584, regular and precise recordings of funerals (including the age of the deceased) and births was maintained by the Silesian city Breslau. The clergyman and scientist Kaspar Neuman of Breslau was attracted by the materials and communicated his findings to the Royal Society in London. The society forwarded the material to the English astronomer Edmund Halley (1656–1742) who published a remarkable paper in 1693 with the title *An Estimate of the Degrees of Mortality of Mankind Drawn from the Curious Tables of Births and Funerals of the City of Breslaw.* In the paper Halley invented a rigorous and precise method of computing life-tables. The method of Halley goes as follows: Let a_x denote the (unknown) number of living aged x years, let $a(x,y)$ denote the (unknown) number of living of age $\geq x$ and $< y$, and let d_x denote the (known) number of deceased in the previous year aged x years. Assuming that $a_{101} = 0$ and that the survival probabilities are constant from year to year we get $a_{x+1} = a_x + d_x$ from which we obtain the following formulas for the unknown quantities a_x and $a(x,y)$:

$$a_x - a_y = \sum_{j=x}^{y-1} d_j \quad , \quad a_x = \sum_{j=x}^{100} d_j$$

$$a(x,y) = (y-x)\sum_{j=y}^{100} d_j + \sum_{j=x}^{y-1} (j-x+1)\, d_j$$

29 Named after Lorenzo Tonti.

Table 4

The probabilites of human Life, according to Dr. Halley From Abraham de Moivre: "Doctrine of Chances"											
Age	Living	Age	Living	Age	Living	Age	Living	Age	Living	Age	Living
1	1000	16	622	31	523	46	387	61	232	76	78
2	855	17	616	32	515	47	377	62	222	77	68
3	798	18	610	33	507	48	367	63	212	78	58
4	760	19	604	34	499	49	357	64	202	79	49
5	732	20	598	35	490	50	346	65	192	80	41
6	710	21	592	36	481	51	335	66	182	81	34
7	692	22	586	37	472	52	324	67	172	82	28
8	680	23	580	38	463	53	313	68	162	83	23
9	670	24	574	39	454	54	302	69	152	84	19
10	661	25	567	40	445	55	292	70	142	*	*
11	653	26	560	41	436	56	282	71	131		
12	646	27	553	42	427	57	272	72	120		
13	640	28	546	43	417	58	262	73	109		
14	634	29	539	44	407	59	252	74	98		
15	628	30	531	45	397	60	242	75	88		

Having computed the numbers a_x it is an easy matter to compute the probability that a person aged x years will survive his next birthday. In de Moivre's book *Doctrine of Chances* we find Halley's life-table [see Table 4] which shows the expected evolution of a group of 1000, 1-year-old babies. I will leave the history of statistics at this point and turn to the progress of probability; in the text you find more historical notes, for instance, in (4.38) you will find a brief account of the history of taking averages, and in (4.39) you will find Daniel Bernoulli's solution to the so-called insurance paradox.

The Progress of Probability theory

In the beginning of the 17th century Nicholas Bernoulli moved from the city of Antwerp in The Netherlands due to the severe government of the Duke of Alba.

After a while he settled down with his family in the Swiss city of Basel, and he soon established himself as a merchant banker. On Christmas Day 1654, his oldest son James Bernoulli was born. James Bernoulli (1654–1705) became the first of a series of eminent mathematicians of the Bernoulli family. From 1676 to 1682 he travelled between the European universities learning science and mathematics. On his journeys he learned about Huyghens book *De Rationiciis in Alea Ludo*, and he became attracted by the subject. From 1690 until his death in 1705 he worked on a book on probability theory entitled *Ars Conjectandi* (The Art of Conjecturing). The book was left unfinished at his death, and the publication was delayed for eight years by his brother John Bernoulli under the pretence that his nephew Nicholas Bernoulli was to complete the book. John became an eminent mathematician like his elder brother James, but he never ceased to envy him.[30] The mathematical community awaited the publication of *Ars Conjectandi* with excitement and finally, John had to publish the unfinished manuscript in 1713. The book contained a wealth of new results, in particular, the theorem which we now call "the law of large numbers" and which James named "the golden theorem" [see (4.8)–(4.12)]. The law of large numbers is, together with the central limit theorem which was discovered by de Moivre in 1733, unquestionably the most important results of probability theory. In the last and unfinished part of the book, James tried to apply the law of large numbers to make a solid foundation of daily-life probabilities, such as "the probability that a man aged 35 years will survive his 70th birthday," which indeed not were covered by the simple lottery model of Fermat and Pascal; but as so many of his successors without success – probability theory had to wait until 1933 before a firm foundation was found.

Several years later Abraham de Moivre (1667–1754) took up the tread in a most ingenious way. De Moivre was born in the French village Vitry; he studied mathematics at Sedan and later at the Sorbonne University in Paris. Together with his family he belonged to the Protestant sect called the Huguenots. On October 18, 1685 the Edict of Nantes[31] was evoked, and the 18-year-old Abraham was imprisoned together with many other Huguenots. Three years later he escaped the prison and fled to England. There he arrived 21 years of age, with no money and no friends, but with a solid knowledge of mathematics. He set up a small school of mathematics on Fleet Street in London, and all his life he had to walk from house to house giving mathematics lessons to children of wealthy people. Besides his work as schoolmaster he also earned money advising gentlemen about the proper odds in betting, but he doesn't seem to have been engaged in gambling otherwise.[32] In 1718, he wrote a marvellous book *Doctrine of Chances* (revised in

---

30 James was indeed a better mathematician than John.

31 The Edict of Nantes was a cease-fire treaty between the French King and the Huguenots allowing them to exercise their religion in certain cities of France.

32 Cardano is the only probabilist of the past, who was seriously engaged in gambling, and for a good reason: In the dedication to "Doctrine of Chances" de Moivre writes: *There are many people in*

1738 and 1756). In 1733 de Moivre discovered the second pearl of probability[33] – the central limit theorem – and the result was included (without proof) in the second and third edition of *Doctrine of Chances*; in (4.22), (5.1) and (5.16) you will learn more about de Moivre's work on the central limit theorem. The book also contains a wealth of solutions to nontrivial problems (even measured with today standards) among these the so-called "ruin problem", which you can find in Sections (11.12+13) of Volume II.

De Moivre never found the normal density function, but he found a series expansion of the normal distribution function. Probability had to wait $\approx$ 60 years before de Moivre's theorem was conceived as we do today. In 1808 the American surveyor Robert Adrain (1775–1843) published a remarkable paper, in which he suggested the normal density as the proper distribution of errors of measurements. One year later, Karl Friederich Gauss (1777–1855) published a paper with the same conclusion. At that time United States was far from the leading mathematical circles, and the paper of Adrain passed unnoticed. Because Gauss was, without competition, the leading mathematician of his time, the normal distribution is also called the Gauss distribution [see (4.22)]. In the period 1770–1820, Pierre-Simon de Laplace concluded the works of J. Bernoulli, A. de Moivre and K. F. Gauss, and the result is included in his famous book *Théorie Analytique des Probabilités* (1812, 1814, 1820). In the book, Laplace extends the central limit theorem of de Moivre and demonstrates the connection to the newly found normal distribution. Moreover, he essentially exhausted the possibilities of the classical Fermat-Pascal model of probability theory. For the rest of the century, probability theory was on the move to find a new foundation which would include the real-life probabilities of the statisticians and insurance companies.

The Search for a New Foundation

After the publication of Laplace' book *Théorie Analytique des Probabilités*, probabilists set out to extend probability beyond the classical Fermat-Pascal model, but with the problem of lacking a firm mathematical foundation. The trail was found by Lvovich Pafnufty Chebyshev (1821–1884), who introduced for the first time the random variables on the stage of probability. He defined a random variable as "a real variable which can assume different values with different probabilities", and he proved his famous inequality [see (4.4.4)], and the first version of the central limit theorem for general random variables. The original proof of his

the world who are prepossessed with an opinion that the doctrine of chances has a tendency to promote play; but they will soon be undecieved . . . it is rather a guard against it, by setting in a clear light, the advantages and disadvantages of those games wherein chance is concerned.

33 The three pearls of probability are the law of large numbers, the central limit theorem and the law of the iterated logarithm.

inequality was extremely long and complicated but was simplified to the three-line proof we know today by his student Andrei Andreiwich Markov (1856–1922), who together with Alexander Mikhailovich Lyapounov (1857–1918) (likewise a student of Chebyshev) corrected and extended the central limit theorem of Chebyshev [see (5.16)]. Around the turn of this century the status of probability theory was such that it was considered as a part of physics instead of mathematics. In the year 1900 a large conference on mathematics was held in Paris, and the leading mathematician of that time, David Hilbert (1862–1943), was invited to give his opinion of the agenda of mathematics in the 20th century. Hilbert did that with an astonishing prevision in the form of 23 problems,[34] which he predicted would be the central mathematical themes of this century. In his sixth problem Hilbert ask for axiomatizations (with as few and simple axioms as possible) of those parts of physics in which mathematics is used, and he explicitly mentions mechanics and probability theory as two physical subjects to be axiomatized. After the lecture Émile Borel suggested that one could use the unit interval with the Lebesgue measure, which Borel had invented the year before, as a model for "continuous" probabilities. A few years later Rademacher constructed a sequence of independent functions on the unit interval taking the values ± 1 with probability $\frac{1}{2}$, and Steinhaus succeeded in constructing a sequence of independent normally distributed functions on the unit interval. However, the unit interval with the Lebesgue measure imposes too few degrees of freedom,[35] in particular, when it comes to complicated objects like Brownian motions.

Brownian motions: In 1827 the English botanist Robert Brown studied the irregular movements, which later was named Brownian motions, of small particles in water. These motions were discovered long time before Brown was born, and the general accepted theory (the so-called vitalism) at that time was that the particles were alive. Brown rejected the theory and showed that small particles of stone or glass performed the same kind of movements; however, he was not able to find a satisfactory theory, and the problem of Brownian motions was discussed among physicists for the rest of the 19th century until Albert Einstein in 1905 suggested the probabilistic model, which we use today [see (9.8) and (9.10)]. Actually, Einstein was surpassed by the economist Bachelier, who in 1900 suggested exactly the same model as Einstein, but as a model of the fluctuations on the French stock market.[36] Einstein's model was later confirmed in a series of experiments performed by Jean Pierre Perrin, who received the Nobel Prize in 1926 for his work with Brownian motions. It was soon found that a stochastic process satisfying the requirements

---

34 Most of the mathematics of this century is somehow related to one or more of the 23 problems of Hilbert.

35 Actually, it can be shown that essentially all stochastic processes can be defined on the unit interval, but the proof requires modern measure theory.

36 Today the Brownian motion is widely used in economics, biology, ecology etc., as a model for practical everything except original Brownian motions.

of Einstein would have continuous sample paths, which was to be expected, but it was also shown that the sample paths would have to be nowhere differentiable, which was shocking as it meant that a Brownian particle travels with infinite speed. This fact was so discomforting that probabilists began to doubt that there existed a consistent theory covering the model of Einstein. However, the eminent mathematician Norbert Wiener set out to prove the existence of Einstein's model for Brownian motions, and after a "tour de force" he finally, succeeded to do so at the end of the 1920's.

The method of Norbert Wiener was tailor-made for Brownian motions and could hardly be used for proving the existence of other complicated processes such as stable processes or the Ohrenstein-Uhlenbeck process. But in 1933 the ingenious mathematician Andrei Nikolaevich Kolmogorov (1903–1989) published *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Foundations of Probability Theory), which was to set probability on its modern course. In *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Kolmogorov applied abstract measure theory to provide a new foundations for probability theory: A *probability space* is defined to be a triple $(\Omega, \mathcal{F}, P)$ satisfying:

(Pr.1) Ω is a non-empty set, $\mathcal{F}$ is a non-empty set of subsets of Ω, and P is a function from $\mathcal{F}$ into the non-negative real numbers $\mathbf{R}_+$

(Pr.2) $F, F_1, F_2, \ldots \in \mathcal{F} \quad \Rightarrow \quad F^c \in \mathcal{F}$ and $\bigcup_{n=1}^{\infty} F_n \in \mathcal{F}$

(Pr.3) $P(\emptyset) = 0$ and $P(\Omega) = 1$

(Pr.4) If $F_1, F_2, \ldots \in \mathcal{F}$ are disjoint, then $P\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} P(F_n)$

A non-empty collection $\mathcal{F}$ of sets satisfying (Pr.2) is called a *σ-algebra*, and a non-negative function P satisfying (Pr.3) and (Pr.4) is called a *probability measure* [see (1.1)]. With this notion at hand we define a *random variable* to be a function from Ω into the real line $\mathbf{R}$ such that the set $\{\omega \mid X(\omega) \leq a\}$ belong to $\mathcal{F}$ for every $a \in \mathbf{R}$ [see (2.7)]. In *Grundbegriffe der Wahrscheinlichkeitsrechnung* Kolmogorov proves his famous *consistency theorem* which, loosely speaking, states that every random phenomenon can be modeled within the axioms (Pr.1)–(Pr.4) unless the specifications contains evident inconsistencies [see Section (9.4) of Volume II, where you will find the precise statement and proof]. The consistency theorem solves in one stroke all the problems of existence of rigorous models for random systems such as the Brownian motion, the Poisson process, the stable processes, the Ohrenstein-Uhlenbeck process, etc., and the book was received by the contemporary probabilists with almost euphoria. Suddenly, all the mess of the house of probability was cleaned up and probability could enter mathematics as a

rigorous and firmly founded discipline. Since 1933 the vast majority of probabilists have been working within the framework of Kolmogorov. However, a few years before (1931), von Mises found another rigorous and firm foundation of probability theory, taking frequencies and the law of large numbers as the basic object, but the work of von Mises was completely overshadowed by Kolmogorov's model, and only a few – very few[37] – probabilists have worked within the framework of von Mises.

Kolmogorov also consider the weaker notion of a *finitely additive probability space*, which is defined to be a triple $(\Omega, \mathcal{F}, P)$ satisfying:

(Fpr.1) Ω is a non-empty set, $\mathcal{F}$ is a non-empty set of subsets of Ω, and P is a function from $\mathcal{F}$ into the non-negative real numbers $\mathbf{R}_+$

(Fpr.2) $F, G \in \mathcal{F} \quad \Rightarrow \quad F^c \in \mathcal{F}$ and $F \cup G \in \mathcal{F}$

(Fpr.3) $P(\emptyset) = 0$ and $P(\Omega) = 1$

(Fpr.4) If $F, G \in \mathcal{F}$ are disjoint, then $P(F \cup G) = P(F) + P(G)$

A non-empty collection $\mathcal{F}$ of sets satisfying (Fpr.2) is called an *algebra*, and a non-negative function P satisfying (Fpr.3) and (Fpr.4) is called a *probability content* [see (1.1)]. The notion has had little impact on the course of probability, and today, the study of *pure probability contents* [that is, a probability contents, which doesn't satisfy (Pr.4)] is closer to logic and the foundation of set theory than to probability theory. For instance, if $\mathcal{F}$ is a σ-algebra, then it can be shown that it is not possible to ***construct*** a pure probability content, and that the existence of a pure probability content is equivalent to a certain form of the axiom of choice.

What Is a Probability?

After this brief account of the history of probabilities it is appropriate to ask: "What is a probability really? And what does history teaches us about probabilities?" The history demonstrates that the notions of randomness and probabilities are difficult – very difficult – to apprehend, and that the notions are fairly new and alien to humans.[38] Second, the history and a wealth of examples, many of which can be found in this text, demonstrate that our intuition about probability is poor and often takes a wrong track. Thus, when you evaluate probabilities trust your computations and doubt your intuition. For this reason the interpretation of probabilities presents

[37] For instance, Kolmogorov in the 1970's applied the ideas of von Mises to define independence of the digits in a given sequence of 0's and 1's.

[38] I don't think that mankind ever was meant to apprehend randomness and probabilities.

a difficult and fundamental problem, and the past – from the emergence of probabilities in 1550 until today – contains numerous examples of misinterpretations and miscalculations. There are three (at least) ways of interpreting probabilities:

A priori probabilities: If someone shows a die and claims that the six possible outcomes have probability $\frac{1}{6}$, you would certainly agree without hesitation, unless you know the particular die and have observed some kind of skewness. Probabilities assigned in this way are called *a priori probabilities*, that is, probabilities that all sensible persons can agree on. A priori probabilities usually stem from an assumption of uniformity, that is, the assumption that certain outcomes are equally probable, and they are often based on the uniform distributions [see (2.15) and (2.24)]. For instance, the classical model of Fermat and Pascal is a typical example of an a priori assignment of probabilities. In statistics the idea is closely related to the so-called *Bayesian school of statisticians*. In Chapters 1 and 2, you will see how to model a priori probabilities within the axioms of Kolmogorov. The main disadvantage of the notion are the many cases where there is no consensus about the probability assignment, and the less numerous cases where the ostensible obvious probability assignment disagrees with observations; see for instance, the anomal distribution described in Section (4.40).

Frequential probabilities: The most common interpretation of a probability is that if you repeat the event independently a certain number of times, then the probability approximately equals the frequency of the occurrence of the event. Such an interpretation is called a *frequential probability*. For instance, claiming that the probability of getting one head in one toss with a coin is $\frac{1}{2}$ is usually interpreted as follows: if you throw the coin 100 times, then you will get approximately 50 heads and 50 tails. According the recordings of one of largest Danish fire insurance companies, the probability that my house will burn down completely within the next year is $\frac{1}{20{,}000}$. In this case, the next year can, of course, not be repeated, but in the files of the insurance company we find records of fires in houses of the same kind and age as my house, which can be used to find a frequency and, thus, a frequential probability. It is important to observe that a frequential probability is a collective notion, which says little or nothing about the individual. For instance, if a doctor tells a cancer patient that he has a 60% chance of surviving the next year, then the statement has little meaning to the patient – either he survives or he don't – but to the doctor the statement has perfectly valid meaning: Of a group of say 100 patients with the same kind cancer the doctor expects 60 to survive and 40 to die within the next year. The model of Kolmogorov gives the basic rules of the calculus of probability, but it isn't a model of frequential probabilities and it doesn't say anything about the probability of a given event; nevertheless, the frequential interpretation of probabilities is the main link between reality and probability theory. Thus, within the framework of Kolmogorov we have to ***prove*** the validity of the frequential interpretation, that is,

to prove the law of large numbers.[39] This will be done in Sections (4.8)–(4.12) and (4.32). The main advantage of a frequential probability is its robustness; that is, new information is unlikely to alter the probability in any significant way (provided the probability is based on many reliable observations). The main disadvantage is that doesn't applies to individual events and requires that the event can be repeated independently many times.

Subjective probabilities: At present the Danish State is building the longest bridge in the world over the highly trafficked waters between the islands of Funen and Zealand. In the summer of 1991, I was asked by a member of the Danish Parliament to give my opinion of the evaluation of the probability that the bridge would by destroyed by a collision with a ship within the next 100 years, a projection made by a consulting firm. The firm had computed the probability to be 5%.[40] In this case the probability is neither an a priori probability – I doubt that anyone would be able to assign a probability without thorough investigations – nor is it a frequential probability – the bridge is not yet finished, and is only built to last 100 years; moreover, there are no comparable bridges in the world.[41] It is a so-called *subjective probability*. A subjective probability is interpreted as the degree of belief in the person stating the probability. Many daily-life probabilities are subjective; for instance, the bookmakers odds are of this type, and often the exchange rates on the currency market are of this type.[42] Kolmogorov's model is in reality a model of subjective probabilities: The axioms don't tell what the probabilities are, but when you have fixed your subjective probabilities they tell you how to calculate other probabilities. The main advantage of a subjective probability is its broadness of application, and its major disadvantages is its subjectivity – you may disagree with practically everybody about the probability assignment – and its sensibility to new information. Before the Chernobyl accident the probability of a serious accident on a nuclear power plant was (subjectively) considered as almost negligible, but after the accident the probability was drastically increased.

References to the History of Probability

The history of probability and statistics serves as entrance for a better understanding of the subjects. There is a number of books covering the history of probability up to about 1820 (for instance, [2], [3], [4] and [5]) but only few

---

39 In many ways it would be more natural to follow the ideas of von Mises, and take the law of large numbers as an axiom and not as a theorem to be proved.

40 This is a very high risk probability; for instance, the probability that my house will burn down completely within the next 100 years is 0.5%.

41 The consulting firm derived the probability by simulations on a computer.

42 Often the exchange rates on the currency market doesn't reflect reality but the speculators' expectation of the future market.

books covering the history of statistics (for instance, [3] and [6]) or the history of probability after 1820 (for instance, [1]).

[1] Adams, J. (1974): *The Life and Times of the Central Limit Theorem*, Kaedman Publishing Co., New York

[2] David, F. N. (1962): *Games, Gods and Gambling*, Charles Griffin & Co., London

[3] Hald, A. (1990): *A History of Probability and Statistics and Their Applications before 1750*, John Wiley & Sons, New York

[4] Maistrov, E. (1974): *Probability – a Historical Sketch*, Academic Press, New York and London

[5] Ore, O. (1953): *Cardano – the Gambling Scholar*, Princeton University Press, Princeton

[6] Westergaard, H. (1932): *Contribution to the History of Statistics*, P. S. King & Son Ltd., London

Chapter 1

MEASURE THEORY

1.1 Pavings and Set Functions

Modern probability theory takes its foundation in measure theory, and in this chapter we shall dip into the elementary part of this theory. Measure theory consists of two ingredients: sets of subsets, called pavings, and functions from sets of subsets into the real line $\mathbf{R}$ or the extended real line $\overline{\mathbf{R}}$ called set functions. In the Introduction you have already encountered two kinds of pavings: algebras and σ-algebras. The characteristic of algebras and σ-algebras is that they are stable under certain set operations, such as complementation and finite or countable unions. We shall now formalize this stability notion.

Let Ω be a set; then 2^{Ω} (*the power set*) denotes the set of all subsets of Ω . A *paving* on Ω is a non-empty subset of 2^{Ω} ; that is, if $\mathcal{F}$ is a paving on Ω , then $\mathcal{F}$ is a set of subsets of Ω containing at least one set. We shall consider 11 *set operations*:

$$\cup f\ ,\ \cup c\ ,\ \uparrow\ ,\ \sum f\ ,\ \sum c\ ,\ \cap f\ ,\ \cap c\ ,\ \downarrow\ ,\ C\ ,\ \setminus\ ,- \tag{1.1.1}$$

called *finite union, countable union, increasing union, disjoint finite union, disjoint countable union, finite intersection, countable intersection, decreasing intersection, complementation, difference,* and *proper difference* respectively. Note that "f" stands for *finite* and "c" stands for *countable*. Let $\mathcal{F}$ be a paving on Ω , then we say that $\mathcal{F}$ is

- $(\cup f)$-*stable* if $F_1 \cup F_2 \in \mathcal{F}\ \ \forall F_1, F_2 \in \mathcal{F}$

- $(\cup c)$-*stable* if $\cup_{n=1}^{\infty} F_n \in \mathcal{F} \quad \forall F_1, F_2, \cdots \in \mathcal{F}$
- $(\uparrow)$-*stable* if $F \in \mathcal{F} \quad \forall F_1, F_2, \cdots \in \mathcal{F}$ with $F_n \uparrow F$ [see (P.1.8)]
- $(\sum f)$-*stable* if $F_1 \cup F_2 \in \mathcal{F} \quad \forall F_1, F_2 \in \mathcal{F}$ with $F_1 \cap F_2 = \emptyset$
- $(\sum c)$-*stable* if $\cup_{n=1}^{\infty} F_n \in \mathcal{F} \quad \forall F_1, F_2 \ldots \in \mathcal{F}$ with $F_n \cap F_m = \emptyset \quad \forall n \neq m$
- $(\cap f)$-*stable* if $F_1 \cap F_2 \in \mathcal{F} \quad \forall F_1, F_2 \in \mathcal{F}$
- $(\cap c)$-*stable* if $\cap_{n=1}^{\infty} F_n \in \mathcal{F} \quad \forall F_1, F_2, \cdots \in \mathcal{F}$
- $(\downarrow)$-*stable* if $F \in \mathcal{F} \quad \forall F_1, F_2, \cdots \in \mathcal{F}$ with $F_n \downarrow F$ [see (P.1.9)]
- (C)-*stable* if $F^c \in \mathcal{F} \quad \forall F \in \mathcal{F}$
- $(\setminus)$-*stable* if $F_1 \setminus F_2 \in \mathcal{F} \quad \forall F_1, F_2 \in \mathcal{F}$
- $(-)$-*stable* if $F_1 \setminus F_2 \in \mathcal{F} \quad \forall F_1, F_2 \in \mathcal{F}$ with $F_2 \subseteq F_1$

We shall also allow combinations of the 11 symbols in (1.1.1), such as $(\cap f, \cup f, \sum c)$-stable, meaning that the paving in question is stable under all the operations in the list. In the Introduction, I introduced two types of pavings, algebras and σ-algebras. With our new formalism above we may define an *algebra* respectively a *σ-algebra* on Ω, to be a $(\cup f, C)$-stable respectively a $(\cup c, C)$-paving on Ω.

The second ingredient of measure theory is the notion of set functions. Let $\mathcal{F}$ be a paving on the set Ω. Then a function μ from $\mathcal{F}$ into the extended positive line $\overline{\mathbf{R}}_+ = [0, \infty]$ is called a *set function* if $\emptyset \in \mathcal{F}$ and $\mu(\emptyset) = 0$. In the Introduction we have met two kind of set functions, probability measures and probability contents, and we shall now extend these notions.

Let $\mathcal{D}$ and $\mathcal{F}$ be pavings on Ω, and let μ be a function from $\mathcal{D}$ into $\overline{\mathbf{R}}$. Then we say that μ is a *content* on $\mathcal{F}$ if $\mathcal{F} \subseteq \mathcal{D}$, $\mathcal{F}$ is an algebra and μ satisfies

$$\mu(\emptyset) = 0 \quad \text{and} \quad \mu(F) \geq 0 \qquad \forall F \in \mathcal{F} \tag{1.1.2}$$

$$\mu(F_1 \cup F_2) = \mu(F_1) + \mu(F_2) \quad \text{if } F_1, F_2 \in \mathcal{F} \text{ and } F_1 \cap F_2 = \emptyset \tag{1.1.3}$$

Similarly, we say that μ is a *measure* on $\mathcal{F}$ if $\mathcal{F} \subseteq \mathcal{D}$, $\mathcal{F}$ is a σ-algebra and μ satisfies

$$\mu(\emptyset) = 0 \quad \text{and} \quad \mu(F) \geq 0 \qquad \forall F \in \mathcal{F} \tag{1.1.4}$$

$$\mu\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mu(F_n) \quad \text{if } F_n \in \mathcal{F} \text{ and } F_n \cap F_m = \emptyset \quad \forall n \neq m \tag{1.1.5}$$

1.2 Stability of Pavings

Let $\mathcal{F}$ be a paving on the set Ω and let $F_1, \ldots, F_n \in \mathcal{F}$ where $n \geq 1$ is an integer. Then we have the following:

(1.2.1) $\mathcal{F}$ *is* $(\cup f)$*-stable* $\Rightarrow$ $F_1 \cup \cdots \cup F_n \in \mathcal{F}$

(1.2.2) $\mathcal{F}$ *is* $(\sum f)$*-stable and* $F_i \cap F_j = \emptyset \;\; \forall i \neq j \;\Rightarrow\; F_1 \cup \cdots \cup F_n \in \mathcal{F}$

(1.2.3) $\mathcal{F}$ *is* $(\cap f)$*-stable* $\Rightarrow$ $F_1 \cap \cdots \cap F_n \in \mathcal{F}$

(1.2.4) $\mathcal{F}$ *is* $(\setminus)$*-stable* $\Leftrightarrow$ $\mathcal{F}$ *is* $(\cap f, -)$*-stable*

(1.2.5) $\mathcal{F}$ *is* $(\setminus, \sum f)$*-stable* $\Rightarrow$ $\mathcal{F}$ *is* $(\cup f)$*-stable*

(1.2.6) $\mathcal{F}$ *is* $(\setminus, \sum c)$*-stable* $\Rightarrow$ $\mathcal{F}$ *is* $(\cup c)$*-stable*

(1.2.7) $\mathcal{F}$ *is* $(\cup c, \setminus)$*-stable and* $\Omega \in \mathcal{F}$ $\Leftrightarrow$ $\mathcal{F}$ *is a* σ*-algebra*

(1.2.8) $\mathcal{F}$ *is* $(\cup f, \setminus)$*-stable and* $\Omega \in \mathcal{F}$ $\Leftrightarrow$ $\mathcal{F}$ *is an algebra*

(1.2.9) $\mathcal{F}$ *is an algebra* $\Rightarrow$ $\mathcal{F}$ *is* $(\cup f, \cap f, \sum f, C, \setminus, -)$*-stable*

(1.2.10) *If* $\mathcal{F}$ *is a* σ*-algebra, then* $\mathcal{F}$ *is stable under all operations in (1.1.1)*

Remark: Note that (1.2.9) says that an algebra is stable under all the usual set operations applied at most a finite number of times; and (1.2.10) says that a σ-algebra is stable under all the usual set operations applied at most a countable number of times.

Proof: (1.2.1), (1.2.2) and (1.2.3) follow by a simple induction argument. Moreover, (1.2.4), (1.2.5), (1.2.7) and (1.2.8) follow easily from the formulas

$$A \cap B = A \setminus (A \setminus B) \quad , \quad A \setminus B = A \setminus (A \cap B)$$

$$A \cup B = A \cup (B \setminus A) \quad , \quad A \cap (B \setminus A) = \emptyset \quad , \quad A^c = \Omega \setminus A$$

(1.2.6) requires a little more work. Let $A_1, A_2, \ldots \in \mathcal{F}$ be subsets of Ω, then we define *the disjungation* of the sequence $\{A_n\}$ as

$$B_1 := A_1 \quad , \quad B_{n+1} := A_{n+1} \setminus \bigcup_{j=1}^{n} A_j = \bigcap_{j=1}^{n} (A_{n+1} \setminus A_j) \tag{1.2.11}$$

Then I claim that $\{B_n\}$ satisfies

(1.2.12) $\quad B_n \cap B_m = \emptyset \qquad \forall n \neq m$

(1.2.13) $\quad \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ and $\bigcup_{n=1}^{k} B_n = \bigcup_{n=1}^{k} A_n \qquad \forall k \geq 1$

To see this, let $\omega \in \Omega$, and let $N(\omega)$ be the least integer j such that $\omega \in A_j$ if such an integer exists, and set $N(\omega) = \infty$ if $\omega \notin A_j$ for all $j \geq 1$; then $B_k = \{\omega \mid N(\omega) = k\}$ and

$$\bigcup_{n=1}^{k} A_n = \{\omega \mid N(\omega) \leq k\} \quad , \quad \bigcup_{n=1}^{\infty} A_n = \{\omega \mid N(\omega) < \infty\}$$

Hence, (1.2.12) and (1.2.13) follow.

(1.2.6): Suppose that $\mathcal{F}$ is $(\setminus, \sum c)$-stable, then $\mathcal{F}$ is $(\cap f, \setminus)$-stable by (1.2.4), and so $B_n \in \mathcal{F}$ by (1.2.3). Hence by (1.2.12), (1.2.13) and $(\sum c)$-stability of $\mathcal{F}$ we have

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$$

(1.2.9): Since $\emptyset = A^c \cap A$ and $\Omega = A^c \cup A$, we see that $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$ whenever $\mathcal{F}$ is an algebra. Since $A \setminus B = (A^c \cup B)^c$, we see that $\mathcal{F}$ is $(\setminus)$-stable. Hence, (1.2.9) follows from (1.2.4).

(1.2.10): Let $\mathcal{F}$ be a σ-algebra, then $\mathcal{F}$ is an algebra and so $\emptyset \in \mathcal{F}$, $\Omega \in \mathcal{F}$ and $\mathcal{F}$ is $(\cup f, \cap f, \sum f, C, \setminus, -)$-stable by (1.2.9). Moreover, since $\mathcal{F}$ is $(\cup c, C)$-stable and

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c$$

we see that $\mathcal{F}$ is $(\cap c)$-stable. Hence, (1.2.10) follows. □

1.3 Contents

Suppose that $\mu : \mathcal{D} \to \overline{\mathbf{R}}$ *is a content on* $\mathcal{F}$ *and let* $F, F_0, F_1, F_2, \cdots$ *be subsets of* Ω *all belonging to* $\mathcal{F}$ *. Then we have*

(1.3.1) $\quad \mu(F_1) = \mu(F_1 \cap F_0) + \mu(F_1 \setminus F_0)$

(1.3.2) $\mu(F_0) \le \mu(F_1)$ *if* $F_0 \subseteq F_1$

(1.3.3) $\mu(F_1 \setminus F_0) = \mu(F_1) - \mu(F_0)$ *if* $F_0 \subseteq F_1$ *and* $\mu(F_0) < \infty$

(1.3.4) $\mu(F_0) + \mu(F_1) = \mu(F_0 \cup F_1) + \mu(F_0 \cap F_1)$

(1.3.5) $\mu\left(\bigcup_{j=1}^{n} F_j\right) = \sum_{j=1}^{n} \mu(F_j)$ *if* $F_1, \cdots, F_n$ *are disjoint*

(1.3.6) $\mu(F) \le \sum_{j=1}^{n} \mu(F_j)$ *if* $F \subseteq F_1 \cup \cdots \cup F_n$

(1.3.7) $\sum_{j=1}^{n} \mu(F_j) \le \mu(F)$ *if* $\bigcup_{j=1}^{n} F_j \subseteq F$ *and* $F_1, \ldots, F_n$ *are disjoint*

(1.3.8) $\sum_{j=1}^{\infty} \mu(F_j) \le \mu(F)$ *if* $\bigcup_{n=1}^{\infty} F_j \subseteq F$ *and* $F_1, F_2, \ldots$ *are disjoint*

Proof: (1.3.1) follows from (1.1.3) since $F_1 \cap F_0$ and $F_1 \setminus F_0$ are disjoint with union equal to F_1 . If $F_0 \subseteq F_1$, then by (1.3.1) we have

$$\mu(F_1) = \mu(F_0) + \mu(F_1 \setminus F_0)$$

Hence, (1.3.2) follows from (1.1.2), and (1.3.3) follows by moving $\mu(F_0)$ to the other side of the equality sign [note that if $\mu(F_0) = \infty$, then $\mu(F_1) = \infty$ and so the right-hand side of (1.3.3) takes the form: $\infty - \infty$, which is not defined].

(1.3.4) follows easily from (1.3.1) and (1.3.5) follows easily from (1.1.3) and a simple induction in n.

(1.3.6): Let $G_1, \ldots, G_n$ be the disjungation of $F_1, \ldots, F_n$ [see (1.2.11)]:

$$G_1 = F_1 \quad , \quad G_{j+1} = F_{j+1} \setminus \bigcup_{i=1}^{j} F_i \quad \text{for } 1 \le j \le n-1$$

Then $G_j \subseteq F_j$ and by (1.2.12+13) we have that $G_1, \cdots, G_n$ are disjoint and $G_1 \cup \cdots \cup G_n = F_1 \cup \cdots \cup F_n$ by (1.2.13). Hence, by (1.3.5) and (1.3.2) we have

$$\mu(F) \le \mu\left(\bigcup_{j=1}^{n} F_j\right) = \mu\left(\bigcup_{j=1}^{n} G_j\right) = \sum_{j=1}^{n} \mu(G_j) \le \sum_{j=1}^{n} \mu(F_j)$$

(1.3.7+8): Let $n \in \mathbf{N}$, then by (1.3.2) and (1.3.5) we have

$$\sum_{j=1}^{n} \mu(F_j) = \mu\left(\bigcup_{j=1}^{n} F_j\right) \le \mu(F)$$

Hence, (1.3.7) holds, and letting $n \to \infty$, we see that (1.3.8) follows. □

1.4 Measures

Let $\mu : \mathcal{D} \to \overline{\mathbf{R}}$ *be a measure on* $\mathcal{F}$ *. Then* μ *is a content on* $\mathcal{F}$*, and if* $F, F_1, F_2, \cdots \in \mathcal{F}$ *, then we have*

$$\mu(F) \leq \sum_{j=1}^{\infty} \mu(F_j) \quad \text{if } F \subseteq \bigcup_{n=1}^{\infty} F_n \tag{1.4.1}$$

$$\mu(F) = \lim_{n\to\infty} \mu(F_n) \quad \text{if } F_n \uparrow F \tag{1.4.2}$$

$$\mu(F) = \lim_{n\to\infty} \mu(F_n) \quad \text{if } F_n \downarrow F \text{ and } \mu(F_1) < \infty \tag{1.4.3}$$

$$\mu(\liminf F_n) \leq \liminf_{n\to\infty} \mu(F_n) \tag{1.4.4}$$

$$\mu(\limsup F_n) \geq \limsup_{n\to\infty} \mu(F_n) \quad \text{if } \mu\left(\bigcup_{n=1}^{\infty} F_n\right) < \infty \tag{1.4.5}$$

$$\mu(\limsup F_n) = 0 \quad \text{if } \sum_{n=1}^{\infty} \mu(F_n) < \infty \tag{1.4.6}$$

Remark: (1.4.4) and (1.4.5) are special cases of the so-called *Fatou lemma* [see (3.5)] and (1.4.6) goes under the name: *the first Borel-Cantelli lemma.* We shall later meet *the second Borel-Cantelli lemma* [see (2.11.2)].

Proof: (1.4.1) is proved in exactly the same way as (1.3.6).

(1.4.2): Suppose that $F_n \uparrow F$. Then $F_1 \subseteq F_2 \subseteq \cdots$ and $F = \cup_1^\infty F_n$. Set $G_1 = F_1$ and $G_n = F_n \setminus F_{n-1}$ for $n = 2, 3, \cdots$. Then $G_1, G_2, \cdots$ are disjoint and $F = \cup_1^\infty G_n$. Hence, by (1.1.5) we have that $\mu(F) = \sum_{n=1}^\infty \mu(G_n)$. If $\mu(F_m) = \infty$ for some $m \geq 1$, then $\mu(F_n) = \infty = \mu(F)$ for all $n \geq m$ and so $\mu(F_n) \to \mu(F)$. So suppose that $\mu(F_n) < \infty$ for all $n \geq 1$. Then by (1.3.3) we have that $\mu(G_n) = \mu(F_n) - \mu(F_{n-1})$ for all $n \geq 2$. Hence we find

$$\begin{aligned}
\mu(F) &= \sum_{n=1}^{\infty} \mu(G_n) = \lim_{k\to\infty} \sum_{n=1}^{k} \mu(G_n) \\
&= \lim_{k\to\infty} \left\{ \mu(F_1) + (\mu(F_2) - \mu(F_1)) + \cdots + (\mu(F_k) - \mu(F_{k-1})) \right\} \\
&= \lim_{k\to\infty} \mu(F_k)
\end{aligned}$$

(1.4.3): Suppose that $F_n \downarrow F$ and $\mu(F_1) < \infty$. Then $F_1 \supseteq F_2 \supseteq \cdots$ and $F = \cap_1^\infty F_n$. Hence $\mu(F_n) < \infty$ for all $n = 0, 1, 2, \cdots$ by (1.3.2). Set $G_n = F_n \setminus F_{n+1}$; then $G_1, G_2, \ldots$ are disjoint and $F_1 \setminus F = \cup_1^\infty G_n$. Hence,

by (1.3.3) and (1.1.5) we have

$$\begin{aligned}\mu(F_1) - \mu(F) = \mu(F_1 \setminus F) &= \sum_{n=1}^{\infty} \mu(G_n) = \lim_{k\to\infty} \sum_{n=1}^{k-1} \mu(G_n) \\ &= \lim_{k\to\infty} \left\{ (\mu(F_1) - \mu(F_2)) + \cdots (\mu(F_{k-1}) - \mu(F_k)) \right\} \\ &= \mu(F_1) - \lim_{k\to\infty} \mu(F_k)\end{aligned}$$

and so (1.4.3) holds.

(1.4.4): Let $G_n = \cap_{j=n}^{\infty} F_j$; then $G_n \subseteq F_j$ for all $j \geq n$ and $G_n \uparrow \liminf F_n$ by (P.1.6), and so by (1.4.2) we have

$$\mu(\liminf F_n) = \lim_{n\to\infty} \mu(G_n) \leq \lim_{n\to\infty} \left\{ \inf_{j\geq n} \mu(F_j) \right\} = \liminf_{n\to\infty} \mu(F_n)$$

(1.4.5): Let $H_n = \cup_{j=n}^{\infty} F_j$; then $H_n \downarrow \limsup F_n$ by (P.1.7), and $\mu(H_1) < \infty$ by assumption. Hence, (1.4.5) follows as above from (1.4.3).

(1.4.6): Let H_n be defined as above; then $H_n \downarrow \limsup F_n$ and by (1.4.1) we have

$$0 \leq \mu(H_n) \leq \sum_{j=n}^{\infty} \mu(F_j)$$

and since $\sum_1^{\infty} \mu(F_j) < \infty$, we see that $\mu(H_n) \to 0$ and so (1.4.6) follows from (1.4.3). □

1.5 Generation of Algebras and σ-Algebras

In this section I shall describe a method of constructing new pavings from old ones, a method which is of fundamental importance to measure theory. Let $(\alpha, \beta, \cdots)$ be a list containing some of the 11 symbols from (1.1.1). Let $\{\mathcal{F}_j \mid j \in J\}$ be a family of pavings which all are $(\alpha, \beta, \cdots)$-stable and set

$$\hat{\mathcal{F}} := \bigcap_{j\in J} \mathcal{F}_j = \{F \subseteq \Omega \mid F \in \mathcal{F}_j \quad \forall j \in J\}$$

Warning: The intersection above is **not** the intersection in Ω , but the intersection in the power set 2^{Ω} .

Then I claim that $\hat{\mathcal{F}}$ is $(\alpha, \beta, \cdots)$-stable. For instance, if $\alpha = (\uparrow)$ and $F_n \in \hat{\mathcal{F}}$ for all $n \geq 1$ such that $F_n \uparrow F$, then $F_n \in \mathcal{F}_j$ for all $n \geq 1$ and all

$j \in J$, by the very definition of the intersection. Hence, $F \in \mathcal{F}_j$ for all $j \in J$ since $\mathcal{F}_j$ is $(\uparrow)$-stable, and so $F \in \hat{\mathcal{F}}$. Thus, $\hat{\mathcal{F}}$ is $(\uparrow)$-stable, and the same argument works for the other 10 operations in (1.1.1).

Let $\mathcal{F}$ be a given paving on Ω , and let $(\alpha, \beta, \cdots)$ be a list containing some of the symbols from (1.1.1). Consider the set

$$\Gamma = \{\mathcal{G} \mid \mathcal{G} \text{ is an } (\alpha, \beta, \cdots)\text{-stable paving and } \mathcal{G} \supseteq \mathcal{F}\}$$

that is, Γ is a set of pavings. Note that $2^{\Omega} \in \Gamma$, and by the argument above we have that $\hat{\mathcal{F}} = \cap_{\mathcal{G} \in \Gamma} \mathcal{G}$ is an $(\alpha, \beta, \cdots)$-stable paving, and $\mathcal{F} \subseteq \hat{\mathcal{F}}$ since $\mathcal{F} \subseteq \mathcal{G}$ for all $\mathcal{G} \in \Gamma$. Thus, $\hat{\mathcal{F}} \in \Gamma$, and clearly we have

(1.5.1) $\hat{\mathcal{F}}$ is the smallest $(\alpha, \beta, \cdots)$-stable paving on Ω , which contains $\mathcal{F}$

In particular, we may define

(1.5.2) $\alpha(\mathcal{F}) :=$ the smallest algebra which contains $\mathcal{F}$

(1.5.3) $\sigma(\mathcal{F}) :=$ the smallest σ-algebra which contains $\mathcal{F}$

whenever $\mathcal{F}$ is a paving on Ω . We say that $\alpha(\mathcal{F})$ respectively $\sigma(\mathcal{F})$ is *the algebra* respectively *the σ-algebra generated by* $\mathcal{F}$. In this way we may construct lots of algebras and σ -algebras: Take any paving $\mathcal{F}$ whatsoever, then $\alpha(\mathcal{F})$ is an algebra and $\sigma(\mathcal{F})$ is a σ-algebra.

If γ is one of the operations in (1.1.1) with $\gamma \neq (C)$, then we define its *complementary operation* γ^c as follows:

$$(\cup f)^c = (\cap f) \quad , \quad (\cup c)^c = (\cap c) \quad , \quad (\uparrow)^c = (\downarrow) \quad , \quad (\sum f)^c = (-)$$

$$(\sum c)^c = (-) \quad , \quad (\cap f)^c = (\cup f) \quad , \quad (\cap c)^c = (\cup c) \quad , \quad (\downarrow)^c = (\uparrow)$$

$$(\setminus)^c = (\cup f) \quad , \quad (-)^c = (\sum f)$$

Note that the definition of $\alpha(\mathcal{F})$ and $\sigma(\mathcal{F})$ is descriptive, but not constructive. The definition does not tell exactly which subsets of Ω belong to $\alpha(\mathcal{F})$ or $\sigma(\mathcal{F})$, but it describes $\alpha(\mathcal{F})$ and $\sigma(\mathcal{F})$ as the smallest algebra respectively σ-algebra which contains $\mathcal{F}$. Our next lemma explores how to use this description.

1.6 Sierpińsky's Stability Lemma

Let $\mathcal{D}$ and $\mathcal{G}$ be two pavings on Ω such that $\mathcal{G} \subseteq \mathcal{D}$ and $\mathcal{D}$ is $(\alpha, \beta, \cdots)$-stable, where $(\alpha, \beta, \cdots)$ is a list of some of the operations from (1.1.1) which do

not include complementation (C). *Let* $\mathcal{D}_0$ *be the smallest* $(\alpha, \beta, \cdots)$*-stable paving containing* $\mathcal{G}$. *Then* $\mathcal{G} \subseteq \mathcal{D}_0 \subseteq \mathcal{D}$ *and we have*

(1.6.1) *Suppose that* $G \cap G' \in \mathcal{D}_0$ *for all* $G, G' \in \mathcal{G}$. *Then* $\mathcal{D}_0$ *is* $(\cap f, \alpha, \beta, \ldots)$*-stable*

(1.6.2) *Suppose that for every operation* γ *in the list* $(\alpha, \beta, \cdots)$, *we have that the complementary operation* γ^c *also occurs in the list and that* $G^c \in \mathcal{D}_0$ *for all* $G \in \mathcal{G}$. *Then* $\mathcal{D}_0$ *is* $(C, \alpha, \beta, \ldots)$*-stable. In particular, if* $(\alpha, \beta) = (\cup c, \cap c)$, *then* $\mathcal{D}_0 = \sigma(\mathcal{G})$, *and if* $(\alpha, \beta) = (\cup f, \cap f)$, *then* $\mathcal{D}_0 = \alpha(\mathcal{G})$.

(1.6.3) *Suppose that for every operation* γ *in the list* $(\alpha, \beta, \cdots)$, *we have that the complementary operation* γ^c *also occurs in the list and that* $G^c \in \mathcal{D}_0$ *and* $G \cap G' \in \mathcal{D}_0$ *for all* $G, G' \in \mathcal{G}$. *Then* $\mathcal{D}_0$ *is* $(C, \cap f, \alpha, \beta, \ldots)$*-stable and* $\alpha(\mathcal{G}) \subseteq \mathcal{D}_0$. *In particular, if* $(\alpha, \beta) = (\downarrow, \uparrow)$, *then* $\mathcal{D}_0 = \sigma(\mathcal{G})$

Remark: Note that if $\mathcal{G}$ is $(\cap f)$-stable, then $G \cap G' \in \mathcal{G} \subseteq \mathcal{D}_0$ for all $G, G' \in \mathcal{G}$, and if $\mathcal{G}$ is (C)-stable, then $G^c \in \mathcal{G} \subseteq \mathcal{D}_0$ for all $G \in \mathcal{G}$.

Proof: By definition of $\mathcal{D}_0$, we have $\mathcal{G} \subseteq \mathcal{D}_0$, and since $\mathcal{G} \subseteq \mathcal{D}$ and $\mathcal{D}$ is $(\alpha, \beta, \cdots)$-stable, then $\mathcal{D}_0 \subseteq \mathcal{D}$ by minimality of $\mathcal{D}_0$. Thus, $\mathcal{G} \subseteq \mathcal{D}_0 \subseteq \mathcal{D}$.

(1.6.1): Let us define

$$\mathcal{G}_A = \{B \mid A \cap B \in \mathcal{D}_0\} \qquad \forall A \subseteq \Omega$$

By the distributive laws (P.2.5)–(P.2.8) it follows that $\mathcal{G}_A$ is $(\alpha, \beta, \cdots)$-stable. For instance, if $(\uparrow)$ belongs to the list $(\alpha, \beta, \cdots)$ and $B_n \in \mathcal{G}_A$ so that $B_n \uparrow B$, then $A \cap B_n \in \mathcal{D}_0$ and $A \cap B_n \uparrow A \cap B$; thus, $A \cap B \in \mathcal{D}_0$ by $(\uparrow)$-stability of $\mathcal{D}_0$ and so $B \in \mathcal{G}_A$, meaning that $\mathcal{G}_A$ is $(\uparrow)$-stable. A similar argument works for all other operations from (1.1.1) except complementation, but we have assumed that (C) does not belong to the list $(\alpha, \beta, \cdots)$. Thus, we conclude that $\mathcal{G}_A$ is $(\alpha, \beta, \cdots)$-stable for all $A \subseteq \Omega$.

By assumption we have that $\mathcal{G} \subseteq \mathcal{G}_A$ for all $A \in \mathcal{G}$, and $\mathcal{G}_A$ is $(\alpha, \beta, \cdots)$-stable. Hence, by minimality of $\mathcal{D}_0$ we have that $\mathcal{D}_0 \subseteq \mathcal{G}_A$ for all $A \in \mathcal{G}$. Note that $B \in \mathcal{G}_A$ if and only if $A \in \mathcal{G}_B$. Hence, $\mathcal{G} \subseteq \mathcal{G}_B$ for all $B \in \mathcal{D}_0$, and since $\mathcal{G}_B$ is $(\alpha, \beta, \cdots)$- stable, then by minimality of $\mathcal{D}_0$ we have that $\mathcal{D}_0 \subseteq \mathcal{G}_B$ for all $B \in \mathcal{D}_0$. Thus, we conclude that $A \cap B \in \mathcal{D}_0$ for all $A, B \in \mathcal{D}_0$, and so $\mathcal{D}_0$ is $(\cap f, \alpha, \beta, \cdots)$-stable.

(1.6.2): In this case we set

$$\mathcal{D}_1 = \{A \subseteq \Omega \mid A \in \mathcal{D}_0 \text{ and } A^c \in \mathcal{D}_0\}$$

Then $\mathcal{D}_1$ is clearly (C)-stable and $\mathcal{D}_1 \subseteq \mathcal{D}_0$, and I claim that $\mathcal{D}_1$ is $(\alpha, \beta, \cdots)$-stable. For instance, if $(\sum c)$ belongs to the list $(\alpha, \beta, \cdots)$ and $\{A_n\}$ are disjoint sets in $\mathcal{D}_1$ with $A = \cup_{n=1}^{\infty} A_n$, then $A \in \mathcal{D}_0$ by $(\sum c)$-stability of $\mathcal{D}_0$. Moreover,

$$A^c = A_1^c \setminus B \qquad \text{where} \qquad B = \bigcup_{n=2}^{\infty} A_n$$

and since $A_1 \cap B = \emptyset$, we have $A_1^c \supseteq B$ and $A_1^c, B \in \mathcal{D}_0$. The complementary operation to $(\sum c)$ equals $(-)$, which by assumption belongs to the list $(\alpha, \beta, \cdots)$; thus, $\mathcal{D}_0$ is $(-)$-stable and so $A^c \in \mathcal{D}_0$, meaning that $A \in \mathcal{D}_1$. Hence, $\mathcal{D}_1$ is $(\sum c)$-stable. A similar argument works for all the other operations from (1.1.1) and so $\mathcal{D}_1$ is $(\alpha, \beta, \cdots)$-stable.

By assumption we have that $\mathcal{G} \subseteq \mathcal{D}_1$, and $\mathcal{D}_1$ is $(\alpha, \beta, \cdots)$-stable. Then $\mathcal{D}_0 \subseteq \mathcal{D}_1$ by minimality of $\mathcal{D}_0$, but the converse inclusion is evident, and henceforth we have $\mathcal{D}_0 = \mathcal{D}_1$. Thus, $\mathcal{D}_0$ is $(C, \alpha, \beta, \cdots)$-stable. If $(\alpha, \beta) = (\cup c, \cap c)$, then $\mathcal{D}_0$ is $(C, \cup c)$-stable and so $\mathcal{D}_0 = \sigma(\mathcal{G})$. A similar argument shows that if $(\alpha, \beta) = (\cup f, \cap f)$, then $\mathcal{D}_0 = \alpha(\mathcal{G})$.

(1.6.3): By (1) and (2) we have that $\mathcal{D}_0$ is $(C, \cap f, \alpha, \beta, \cdots)$-stable, and since

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

we see that $\mathcal{D}_0$ is $(C, \cap f, \cup f, \alpha, \beta, \cdots)$. In particular, $\mathcal{D}_0$ is an algebra containing $\mathcal{G}$, and so $\alpha(\mathcal{G}) \subseteq \mathcal{D}_0$.

If $(\alpha, \beta) = (\uparrow, \downarrow)$, then $\mathcal{D}_0$ is $(C, \cup f, \uparrow)$-stable. Let $\{A_n\} \subseteq \mathcal{D}_0$, then $A^n := A_1 \cup \cdots \cup A_n \in \mathcal{D}_0$ by $(\cup f)$-stability of $\mathcal{D}_0$ and $A^n \uparrow \cup_1^{\infty} A_n$. Thus, $\mathcal{D}_0$ is $(C, \cup c)$-stable and so $\mathcal{D}_0$ is a σ-algebra containing $\mathcal{G}$. This means that $\sigma(\mathcal{G}) \subseteq \mathcal{D}_0$, but $\sigma(\mathcal{G})$ is a $(\uparrow, \downarrow)$-stable paving containing $\mathcal{G}$; so by minimality of $\mathcal{D}_0$, we have that $\mathcal{D}_0 = \sigma(\mathcal{G})$. □

1.7 The Uniqueness Theorem I

Let μ_1 and μ_2 be two measures on the σ-algebra $\mathcal{F}$, let $\mathcal{G} \subseteq \mathcal{F}$ be a $(\cap f)$-stable paving on Ω , and let $F_1, F_2, \ldots \in \sigma(\mathcal{G})$ be given sets such that

(1.7.1) $$\mu_1(G \cap F_n) = \mu_2(G \cap F_n) \qquad \forall G \in \mathcal{G}\ \forall n \geq 1$$

(1.7.2) $$\Omega = \bigcup_{n=1}^{\infty} F_n \quad \textit{and} \quad \mu_1(F_k) = \mu_2(F_k) < \infty \qquad \forall k \geq 1$$

Then we have that $\mu_1(H) = \mu_2(H)$ *for all* $H \in \sigma(\mathcal{G})$.

Proof: Let $\mathcal{G}_0 = \mathcal{G} \cup \{\Omega\}$; then it is easily seen that $\mathcal{G}_0$ is $(\cap f)$-stable. If we set

$$\mathcal{D} = \{ A \in \mathcal{F} \mid \mu_1(A \cap F_n) = \mu_2(A \cap F_n) \quad \forall n \geq 1 \}$$

then by (1.1.5), (1.3.3), (1.7.1) and (1.7.2) we have that $\mathcal{D}$ is a $(\sum c, -)$-stable paving containing $\mathcal{G}_0$. Hence, if $\mathcal{D}_0$ is the smallest $(\sum c, -)$-stable paving containing $\mathcal{G}_0$, then by Sierpińsky's stability lemma we have that $\mathcal{D}_0$ is $(\cap f, \sum c, -)$-stable. By (1.2.4) and (1.2.6) we see that $\mathcal{D}_0$ is $(\cup c, \setminus)$-stable; moreover, since $\Omega \in \mathcal{G}_0 \subseteq \mathcal{D}_0$, then by (1.2.7) we conclude that $\mathcal{D}_0$ is a σ-algebra containing $\mathcal{G}$. Thus, $\sigma(\mathcal{G}) \subseteq \mathcal{D}$ and so

$$\mu_1(F_n \cap H) = \mu_2(F_n \cap H) \qquad \forall n \geq 1 \; \forall H \in \sigma(\mathcal{G})$$

Let $H \in \sigma(\mathcal{G})$ be given and let $\{D_n\}$ be the disjungation of $\{F_n\}$, that is, $D_1 = F_1$ and $D_n = F_n \setminus F^{n-1}$ for $n \geq 2$, where $F^n = F_1 \cup \cdots \cup F_n$. Recall that $F_1, F_2, \ldots \in \sigma(\mathcal{G})$; thus, we have that $H \setminus F^{n-1} \in \sigma(\mathcal{G})$, and since $H \cap D_n = F_n \cap (H \setminus F^{n-1})$, we conclude that $\mu_1(H \cap D_n) = \mu_2(H \cap D_n)$. Moreover, since $D_1, D_2, \ldots$ are disjoint with union Ω , then summing over $n \in \mathbf{N}$ we get

$$\begin{aligned} \mu_1(H) &= \mu_1\left(\bigcup_{n=1}^{\infty} (H \cap D_n) \right) = \sum_{n=1}^{\infty} \mu_1(H \cap D_n) \\ &= \sum_{n=1}^{\infty} \mu_2(H \cap D_n) = \mu_2(H) \end{aligned}$$

Thus, the theorem is proved. □

1.8 The Borel σ-Algebra

I shall now describe an important type of σ-algebra. Let (Ω, d) be a pseudometric space. Then we let $\mathcal{F}$ denote the paving of all closed subsets of Ω , and we let $\mathcal{G}$ denote the paving of all open subsets of Ω . Then $\mathcal{F}$ is $(\cap c, \cup f)$-stable and $\mathcal{G}$ is $(\cup c, \cap f)$-stable, and

$$F \in \mathcal{F} \iff F^c \in \mathcal{G} \qquad \text{and} \qquad G \in \mathcal{G} \iff G^c \in \mathcal{F}$$

Hence, we have that $\sigma(\mathcal{F}) = \sigma(\mathcal{G})$. This σ-algebra plays a cardinal role in measure theory and it is called *the Borel σ-algebra*, denoted $\mathcal{B}(\Omega)$; that is,

$$\mathcal{B}(\Omega) := \sigma(\mathcal{G}) = \sigma(\mathcal{F})$$

Thus, $\mathcal{B}(\Omega)$ is the smallest σ-algebra on Ω containing all open sets (or all closed sets). A set $B \in \mathcal{B}(\Omega)$ is called a *Borel set* and is said to be *Borel measurable.* A content or measure on $\mathcal{B}(\Omega)$ is called a *Borel content* respectively a *Borel measure* on Ω . The Borel σ-algebras on $\mathbf{R}$ or $\mathbf{R}^m$ (with their usual euclidean metrics, of course) plays a particularly important role.

1.9 Generation of the Borel σ-Algebra

Let (Ω, d) be a separable metric space and let $\mathcal{H} \subseteq \mathcal{B}(\Omega)$ be a paving on Ω. Suppose that there exists positive numbers $r_1 \geq r_2 \geq \cdots$ such that $r_n \to 0$ and

(1.9.1) $$\forall \omega \in \Omega \, \forall n \geq 1 \, \exists H \in \sigma(\mathcal{H}) \quad : \quad b(\omega, r_{n+1}) \subseteq H \subseteq b(\omega, r_n)$$

Then we have that $\sigma(\mathcal{H}) = \mathcal{B}(\Omega)$. [Recall that $b(\omega, r)$ is the open ball with center at ω and radius r].

Proof: Since $\mathcal{H} \subseteq \mathcal{B}(\Omega)$ and $\mathcal{B}(\Omega)$ is a σ-algebra, we have that $\mathcal{B}(\Omega) \supseteq \sigma(\mathcal{H})$ because $\sigma(\mathcal{H})$ is the smallest σ-algebra containing $\mathcal{H}$. So let G be an open set, I shall then show that $G \in \sigma(\mathcal{H})$. Let D be a countable dense subset of Ω , then $G \cap D$ is countable, so let us enumerate $G \cap D$: $G \cap D = \{\omega_1, \omega_2, \ldots, \omega_j, \cdots\}$. Since G is open and $\omega_j \in G$, we have that $b(\omega_j, r) \subseteq G$ for some $r > 0$. Since $r_n \to 0$, we see that $b(\omega_j, r_n) \subseteq G$ for some (sufficiently large) n . So we may define

$$\sigma(j) = \min\{n \in \mathbf{N} \mid b(\omega_j, r_n) \subseteq G\}$$

By (1.9.1) we can find $H_1, H_2, \ldots \in \sigma(\mathcal{H})$ such that

(i) $$b\big(\omega_j, r_{\sigma(j)+1}\big) \subseteq H_j \subseteq b\big(\omega_j, r_{\sigma(j)}\big) \qquad \forall j \geq 1$$

I now claim that we have

(ii) $$G = \bigcup_{j=1}^{\infty} H_j$$

So let $\omega \in G$. Since G is open, there exists $r > 0$ such that $b(\omega, r) \subseteq G$. Since $r_n \to 0$, there exists $p \geq 1$ such that $r_p + r_{p+1} \leq r$; and since D is dense in Ω , there exists $\omega^* \in D$ such that $d(\omega^*, \omega) < r_{p+1}$. Since $r_{p+1} \leq r$, we have $d(\omega^*, \omega) < r$, and so $\omega^* \in D \cap G$. This means that ω^* has some number j in our enumeration of $D \cap G$, that is, $\omega^* = \omega_j$. Let $\omega^{**} \in b(\omega_j, r_p)$ then by the triangle inequality we have

$$d(\omega, \omega^{**}) \leq d(\omega, \omega_j) + d(\omega_j, \omega^{**}) < r_{p+1} + r_p \leq r$$

Hence, we find that $b(\omega_j, r_p) \subseteq b(\omega, r) \subseteq G$, and so $\sigma(j) \leq p$ and $d(\omega, \omega_j) < r_{p+1} \leq r_{\sigma(j)+1}$ because the r_n 's are decreasing. Thus, by (i) we get

$$\omega \in b(\omega_j, r_{\sigma(j)+1}) \subseteq H_j \subseteq \bigcup_{i=1}^{\infty} H_i \qquad \forall \omega \in G$$

and so we have $G \subseteq \cup_{i=1}^{\infty} H_i$. Moreover, by (i) and the definition of $\sigma(i)$, we have

$$\bigcup_{i=1}^{\infty} H_i \subseteq \bigcup_{i=1}^{\infty} b(\omega_i, r_{\sigma(i)}) \subseteq G$$

thus, (ii) is proved.

Recall that the H_j 's were chosen such that $H_j \in \sigma(\mathcal{H})$, and so by (ii), we see that $\sigma(\mathcal{H})$ contains any open subset G of Ω . Since $\mathcal{B}(\Omega)$ is the smallest σ-algebra containing all open sets, we conclude that $\mathcal{B}(\Omega) \subseteq \sigma(\mathcal{H})$. Having already established the converse inclusion, we have proved that $\mathcal{B}(\Omega) = \sigma(\mathcal{H})$. □

1.10 The Borel σ-Algebra on $\mathbf{R}$ and $\mathbf{R}^k$

Consider the real line $\mathbf{R}$ with its usual metric, and set

$$\begin{aligned}
\mathcal{F}_1 &= \{\,]-\infty, a] \mid a \in \mathbf{R}\} &,\quad \mathcal{F}_2 &= \{\,]-\infty, a[\mid a \in \mathbf{R}\} \\
\mathcal{F}_3 &= \{\, [a, \infty[\mid a \in \mathbf{R}\} &,\quad \mathcal{F}_4 &= \{\,]a, \infty[\mid a \in \mathbf{R}\} \\
\mathcal{F}_5 &= \{\, [a, b] \mid a, b \in \mathbf{R}\} &,\quad \mathcal{F}_6 &= \{\,]a, b[\mid a, b \in \mathbf{R}\} \\
\mathcal{F}_7 &= \{\,]a, b] \mid a, b \in \mathbf{R}\} &,\quad \mathcal{F}_8 &= \{\, [a, b[\mid a, b \in \mathbf{R}\}
\end{aligned}$$

Since the set of rationals is countable and dense in $\mathbf{R}$, we have that $\mathbf{R}$ is separable, and if $r_n = 2^{-n}$, then it can be easily checked that each of the eight pavings $\mathcal{F}_j$, $j = 1, \cdots, 8$, satisfies (1.9.1). Hence,

(1.10.1) $\qquad \mathcal{B}(\mathbf{R}) = \sigma(\mathcal{F}_1) = \sigma(\mathcal{F}_2) = \cdots = \sigma(\mathcal{F}_8)$

Let $k \geq 1$ and consider the k-dimensional euclidean space $\mathbf{R}^k$ (with its usual euclidean metric of course). If $a = (a_1, \cdots, a_k)$ and $b = (b_1, \cdots, b_k)$ are vectors in $\mathbf{R}^k$, we define

$$]a, b] = \{ (x_1, \ldots, x_n) \mid a_j < x_j \leq b_j \quad \forall j = 1, \ldots, n \}$$

$$]-\infty, b] = \{ (x_1, \ldots, x_n) \mid x_j \leq b_j \quad \forall j = 1, \ldots, n \}$$

$$\vdots$$

etc.

Thus, we may define pavings $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_8$ on $\mathbf{R}^k$ exactly as above, and exactly as above, we find

$$\mathcal{B}(\mathbf{R}^k) = \sigma(\mathcal{F}_1) = \sigma(\mathcal{F}_2) = \cdots = \sigma(\mathcal{F}_8) \tag{1.10.2}$$

1.11 Finite σ-Algebras

It is time to leave the abstractions and turn to some simple examples on σ-algebras. Any σ-algebra on Ω contains $\emptyset$ and Ω by (1.2.10), and since $\mathcal{F} = \{\emptyset, \Omega\}$ is indeed a σ-algebra, this is the smallest σ-algebra on Ω. Let $A \subseteq \Omega$, then the smallest σ-algebra containing A, has four elements and is given by

$$\sigma\{A\} = \{\emptyset, A, A^c, \Omega\} \tag{1.11.1}$$

Let A and B be subsets of Ω, then by a straightforward (but lengthy) argument one finds that the smallest σ-algebra containing A and B has 16 elements and is given by

$$\begin{aligned} \sigma\{A, B\} = \{ & \emptyset, A \cap B, A \setminus B, B \setminus A, A^c \cap B^c, A, B, A^c, B^c, \\ & A \Delta B, (A \Delta B)^c, A \cup B, A \cup B^c, A^c \cup B, A^c \cup B^c, \Omega \} \end{aligned} \tag{1.11.2}$$

The smallest σ-algebra containing three sets A, B and C has 256 elements, so I shall abstain from writing the list here.

Consider n subsets, $A_1, \cdots, A_n$ of Ω, and set $\gamma = \{1, \cdots, n\}$ and $\Gamma = 2^\gamma$. Let

$$N(\omega) = \{ j \in \gamma \mid \omega \in A_j \} \qquad \forall \omega \in \Omega$$

Then N is a function from Ω into Γ, and if we set

$$A_\delta = \{\omega \mid N(\omega) = \delta\} = \bigcap_{j\in\delta} A_j \cap \bigcap_{j\notin\delta} A_j^c \quad \forall \delta \subseteq \gamma$$

$$A_\Delta = \{\omega \mid N(\omega) \in \Delta\} = \bigcup_{\delta\in\Delta} A_\delta \quad \forall \Delta \subseteq \Gamma$$

then $A_\Delta \in \sigma\{A_1, \cdots, A_n\}$. Moreover, because

$$(A_\Delta)^c = A_{\Delta^c} \quad , \quad \bigcup_{j=1}^{\infty} A_{\Delta_j} = A_\Delta \quad \text{where} \quad \Delta = \bigcup_{j=1}^{\infty} \Delta_j$$

we see that

(1.11.3) $\quad \sigma\{A_1, \cdots, A_n\} = \{A_\Delta \mid \Delta \subseteq \Gamma\}$

Note that γ has n elements, so $\Gamma = 2^\gamma$ has 2^n elements, and 2^Γ has 2^{2^n} elements. Hence in general the σ-algebra generated by n sets has 2^{2^n} elements - a number which increases very rapidly with n. For instance, a σ-algebra generated by 10 sets has about $1.8 \cdot 10^{308}$ elements. The moral of this is that σ-algebras are complicated objects.

1.12 The σ-Algebra Generated by Singletons

Let Ω be a set. Then a *singleton* in Ω is a subset of Ω with exactly one element; that is, a singleton is a set of the form $A = \{\omega_0\}$ for some $\omega_0 \in \Omega$. Let $\mathcal{S}$ be the paving consisting of all singletons: $\mathcal{S} := \{\{\omega\} \mid \omega \in \Omega\}$. Then it is easily checked that we have

(1.12.1) $\quad \sigma(\mathcal{S}) = \{A \subseteq \Omega \mid \text{either } A \text{ or } A^c \text{ is countable}\}$

and if we define

(1.12.2) $\quad P(A) = \begin{cases} 0 & \text{if } A \text{ is countable} \\ 1 & \text{if } A \text{ is uncountable} \end{cases}$

(1.12.3) $\quad \delta(\omega_1, \omega_2) = \begin{cases} 0 & \text{if } \omega_1 = \omega_2 \\ 1 & \text{if } \omega_1 \neq \omega_2 \end{cases}$

then P is a measure on $\sigma(\mathcal{S})$, and δ is a metric on Ω called *the discrete metric* or *the Kronecker metric*. Moreover, $b(\omega, r) = \{\omega\}$ if $0 \leq r \leq 1$, and $b(\omega, r) = \Omega$ if $r > 1$. Hence $\sigma(\mathcal{S})$ equals the smallest σ-algebra on Ω containing all open δ-balls. With this metric, any subset of Ω becomes both open and closed, so $\mathcal{B}(\Omega) = 2^{\Omega}$. If Ω is uncountable then there exist a subset A of Ω such that neither A nor A^c is countable. Hence, $\mathcal{B}(\Omega) \neq \sigma(\mathcal{S})$ but $\mathcal{S}$ clearly satisfies (1.9.1) with $r_n = 2^{-n}$. Thus, we see that the separability condition in (1.9) cannot be omitted in general.

1.13 The Counting Measure

The power set 2^{Ω} is of course the largest possible σ-algebra on Ω . We define *the counting measure*, denoted $\#$, and *the finite-co-finite content*, denoted Δ , as follows:

$$\#(A) = \begin{cases} \text{the number of elements of } A & \text{if } A \text{ is finite} \\ \infty & \text{if } A \text{ is infinite} \end{cases} \tag{1.13.1}$$

$$\Delta(A) = \begin{cases} 0 & \text{if } A \text{ is finite} \\ \infty & \text{if } A \text{ is infinite} \end{cases} \tag{1.13.2}$$

Then $\#$ is a measure on 2^{Ω} and Δ is a content on 2^{Ω} , but if Ω is infinite, then Δ is **not** a measure on 2^{Ω} .

1.14 Measures on $(\mathbf{N}, 2^{\mathbf{N}})$

Let $\{p_n\}$ be a sequence of non-negative numbers on $\mathbf{N}$ and define

$$\mu(A) = \sum_{n \in A} p_n \qquad \forall A \subseteq \mathbf{N} \tag{1.14.1}$$

Since the sum of non-negative numbers does not depend on the order in which they are summed, we have that μ is a measure on $2^{\mathbf{N}}$ such that

$$\mu(\{n\}) = p_n \qquad \text{and} \qquad \mu(\mathbf{N}) = \sum_{n=1}^{\infty} p_n \tag{1.14.2}$$

Conversely, if μ is a measure on $2^{\mathbf{N}}$ and $p_n = \mu(\{n\})$, then μ is given by (1.14.1). To see this, let $A \subseteq \mathbf{N}$, then $A = \cup_{n \in A} \{n\}$ and the union is disjoint and countable. Hence, by (1.1.5) we find

$$\mu(A) = \sum_{n \in A} \mu(\{n\}) = \sum_{n \in A} p_n$$

and thus, μ is given by (1.14.1).

The content Δ defined in (1.13.2) (with $\Omega = \mathbf{N}$) is a content on $2^{\mathbf{N}}$ that is not a measure. It is not possible to ***construct*** a ***finite*** content on $2^{\mathbf{N}}$ which is not a measure. However, such contents do exist! But their existence is a consequence of the axiom of choice, and no explicit construction can be given.

1.15 The Dirac Measure

Let Ω be a set, and let $c \in \Omega$ be a given point in Ω . We may then define a measure on 2^{Ω} , called *the Dirac measure* at c , as follows

$$\delta_c(A) = \begin{cases} 1 & \text{if } c \in A \\ 0 & \text{if } c \notin A \end{cases} \tag{1.15.1}$$

for all $A \subseteq \Omega$. Note that δ_c assumes only the values 0 and 1 . A measure which only assumes the values 0 and 1 is called a $0-1$ *measure*. It is natural to ask if every $0-1$ measure on 2^{Ω} is a Dirac measure? The answer is confusing: It is sometimes true and it is sometimes false depending on the set theory in which we are working.

1.16 Construction of Algebras

Let $\mathcal{F}$ be a paving on Ω ; then the definition of the algebra $\alpha(\mathcal{F})$ generated by $\mathcal{F}$ is descriptive (the smallest algebra containing $\mathcal{F}$) but not constructive. However, it is possible to give several constructions leading to $\alpha(\mathcal{F})$. Let the pavings $\mathcal{F}_1, \mathcal{F}_2, \cdots$ be defined by induction in the following way:

$$\mathcal{F}_1 = \mathcal{F} \cup \{\emptyset, \Omega\} \quad , \quad \mathcal{F}_{n+1} = \{A \cup B^c \mid A \in \mathcal{F}_n, B \in \mathcal{F}_n\}$$

Then I claim that

$$\alpha(\mathcal{F}) = \bigcup_{n=1}^{\infty} \mathcal{F}_n \tag{1.16.1}$$

The idea is that first we add $\emptyset$ and Ω which we know belong to $\alpha(\mathcal{F})$; then knowing that $A \cup B^c \in \alpha(\mathcal{F})$ for all $A, B \in \alpha(\mathcal{F})$, we perform this operation on all sets in $\mathcal{F}_1$ and so on. As the final result we obtain the whole algebra $\alpha(\mathcal{F})$. I shall leave the proof of (1.16.1) as an exercise to the reader [see Exc.1.11].

The construction (1.16.1) involves a countable number of steps. There is another construction which only involves two steps. Let $\mathcal{F}_0 = \mathcal{F} \cup \{\emptyset, \Omega\} \cup \{F^c \mid F \in \mathcal{F}\}$. Then I claim that

$$(1.16.2) \qquad A \in \alpha(\mathcal{F}) \iff \exists F_{ij} \in \mathcal{F}_0 \quad \text{so that} \quad A = \bigcup_{i=1}^{n} \bigcap_{j=1}^{m} F_{ij}$$

Again the idea is to perform some of the operations, which we know leave $\alpha(\mathcal{F})$ stable, on all sets in $\mathcal{F}$ in the hope that in that way we obtain all of $\alpha(\mathcal{F})$. I shall leave the proof as an exercise to the reader [see Exc.1.12].

1.17 The Borel σ-Algebra Revisited

Consider the Borel σ-algebra $\mathcal{B}(\mathbf{R})$ on the real line. Then all intervals, all open sets and all closed sets belong to $\mathcal{B}(\mathbf{R})$, and so do all sets formed by applying the usual set-theoretic operations on intervals, open sets or closed sets at most a countable number of times. It is hard to imagine a subset of $\mathbf{R}$ that cannot be formed in this way, and one may ask whether all subsets of $\mathbf{R}$ are Borel sets? The answer is no! The axiom of choice implies the existence of a non-Borel subset of $\mathbf{R}$. However, this is a non-constructive method and the point is that virtually all subsets of $\mathbf{R}$ that you ever could think about are Borel sets.

It is possible to construct non-Borel sets without appealing to the axiom of choice, but the construction is weird, and the simplest example I know goes as follows: Let $\mathbf{N}^{\mathbf{N}}$ denote the set of all sequences $\sigma = \{\sigma(1), \sigma(2), \cdots\}$ of positive integers, and let $F(\sigma)$ denote the *continued fraction*

$$F(\sigma) = \sigma(1) + \cfrac{1}{\sigma(2) + \cfrac{1}{\sigma(3) + \cfrac{1}{\sigma(4) + \cfrac{1}{\sigma(5) + \cfrac{1}{\sigma(6) + \ddots}}}}}$$

where this is to be understood as follows: Stop the process after k-times, compute the fraction, and let $k \to \infty$. It can be shown that the kth fraction converges no matter which $\sigma \in \mathbf{N}^{\mathbf{N}}$ we take, and that F is injective and maps $\mathbf{N}^{\mathbf{N}}$ onto the set of all irrational numbers > 1 . Let

$$D = \left\{ \sigma \in \mathbf{N}^{\mathbf{N}} \;\middle|\; \begin{array}{l} \exists n_1 < n_2 < \cdots \quad \exists k_1, k_2, \ldots \in \mathbf{N} \\ \text{so that } \sigma(n_j) = k_j \cdot \sigma(n_{j+1}) \quad \forall j \geq 1 \end{array} \right\}$$

It can be shown (but this is very difficult) that the set $F(D) = \{F(\sigma) \mid \sigma \in D\}$ is a non-Borel set.

In a general metric space, the situation is much the same: It takes a weird fantasy to imagine a subset which is not Borel. The rule-of-thumb is that any subset of a metric space which is given by a concrete construction, is a Borel set. However, any rule-of-thumb has exceptions, so let me give the exception of this one:

Let $C[0,1]$ be the set of all continuous functions from the unit interval $[0,1]$ into the real line $\mathbf{R}$ with its sup-norm metric:

$$||f - g||_\infty = \sup_{0 \le t \le 1} |f(t) - g(t)|$$

Then $C[0,1]$ is complete (a uniform limit of continuous functions is continuous) and separable (the set of polynomials with rational coefficients is dense and countable [see Exc.4.29]). Let

$$C^{(1)}[0,1] = \{ f \in C[0,1] \mid f \text{ is differentiable}\}$$

Then $C^{(1)}[0,1] \subseteq C[0,1]$, and it can be shown (but this is extremely difficult) that $C^{(1)}[0,1]$ is **not** a Borel set in $C[0,1]$.

1.18 Construction of σ-Algebras

In (1.16) we saw how to construct algebras by successive use of some of the set-theoretic operations. It is natural to ask if something similar is possible for σ-algebras. Let $\mathcal{F}$ be a paving on Ω. Since $\emptyset$ and Ω belong to $\sigma(\mathcal{F})$, and $\sigma(\mathcal{F})$ is the smallest $(C, \cup c)$-stable paving containing $\mathcal{F}$, it is natural to first add $\emptyset$ and Ω to $\mathcal{F}$, then add all countable unions, then add all complements and so on:

$$\mathcal{F}_0 = \mathcal{F} \cup \{\emptyset, \Omega\} \quad , \quad \mathcal{F}_1 = \mathcal{F}_0 \cup \left\{ \bigcup_{n=1}^{\infty} F_n \;\middle|\; F_1, F_2, \ldots \in \mathcal{F}_0 \right\}$$

$$\mathcal{F}_2 = \mathcal{F}_1 \cup \{ F^c \mid F \in \mathcal{F}_1\} \quad , \quad \mathcal{F}_3 = \mathcal{F}_2 \cup \left\{ \bigcup_{n=1}^{\infty} F_n \;\middle|\; F_1, F_2, \ldots \in \mathcal{F}_2 \right\}$$

$$\mathcal{F}_4 = \mathcal{F}_3 \cup \{ F^c \mid F \in \mathcal{F}_3\} \quad , \quad \mathcal{F}_5 = \mathcal{F}_4 \cup \left\{ \bigcup_{n=1}^{\infty} F_n \;\middle|\; F_1, F_2, \ldots \in \mathcal{F}_4 \right\}$$

$$\vdots$$

$$\mathcal{F}_\infty = \bigcup_{n=0}^{\infty} \mathcal{F}_n$$

It would be natural to expect that we have reached the σ-algebra $\sigma(\mathcal{F})$, but this **not** so in general! For instance, if $\mathcal{F}$ is the set of all intervals on $\mathbf{R}$, then $\mathcal{F}_\infty \neq \sigma(\mathcal{F})$, so we could try to repeat the process with $\mathcal{F}$ replaced by $\mathcal{F}_\infty$; but even if we do this, an infinite $\times$ infinite $\times$ infinite $\times \cdots$ number of times we will **not** reach $\sigma(\mathcal{F})$. However, if you repeat this process a superinfinite number of times (more precisely $\aleph_1$-times), then it can be shown that we reach the σ-algebra $\sigma(\mathcal{F})$. But it takes more than a countable number of successive operations to construct the σ-algebra.

1.19 Null Sets and Almost Sure Properties

Let $\mathcal{F}$ be a σ-algebra on Ω and let μ be a measure on $\mathcal{F}$. Then $N \subseteq \Omega$ is called a *μ-null set*, if there exists $N_0 \in \mathcal{F}$ such that $N \subseteq N_0$ and $\mu(N_0) = 0$. We let $\mathcal{N}(\mu)$ denote the paving of all μ-null sets:

(1.19.1) $\quad \mathcal{N}(\mu) = \{ N \subseteq \Omega \mid N \text{ is a } \mu\text{-null set} \}$

By (1.4.1) and definition of μ-null sets we have

(1.19.2) $\quad$ Any subset of a μ-null set is a μ-null set

(1.19.3) $\quad \mathcal{N}(\mu)$ is $(\cup c)$-stable and $\emptyset \in \mathcal{N}(\mu)$

If $p(\omega)$ is a property or statement depending on $\omega \in \Omega$, we say that $p(\omega)$ holds *μ-almost surely* (abbreviated *μ-a.s.*) if there exists a μ-null set N such that $p(\omega)$ is true for all $\omega \in \Omega \setminus N$. If $F \subseteq \Omega$ we say that $p(\omega)$ holds *μ-a.s. on* F , or that $p(\omega)$ holds for *μ-almost all* $\omega \in F$ (abbreviated *μ-a.a.* $\omega \in F$) if there exists a μ-null set N such that $p(\omega)$ is true for all $\omega \in F \setminus N$. For instance, if $A \subseteq \Omega$, $B \subseteq \Omega$ and $f, g, f_1, f_2, \cdots$ are functions from Ω into $\overline{\mathbf{R}}$, then we write

$$A \subseteq B \quad \mu\text{-a.s.} \iff A \setminus B \text{ is a } \mu\text{-null set}$$

$$A = B \quad \mu\text{-a.s.} \iff A \Delta B = (A \setminus B) \cup (B \setminus A) \text{ is a } \mu\text{-null set}$$

$$f \leq g \quad \mu\text{-a.s.} \iff \{ \omega \mid f(\omega) > g(\omega) \} \text{ is a } \mu\text{-null set}$$

$$f = g \quad \mu\text{-a.s.} \iff \{ \omega \mid f(\omega) \neq g(\omega) \} \text{ is a } \mu\text{-null set}$$

$$f_n \to f \quad \mu\text{-a.s.} \iff \exists N \in \mathcal{N}(\mu) \ : \ f(\omega) = \lim_{n\to\infty} f_n(\omega) \quad \forall \omega \in \Omega \setminus N$$

$$f_n \uparrow f \quad \mu\text{-a.s.} \iff f_n \to f \quad \mu\text{-a.s. and } f_n \leq f_{n+1} \quad \mu\text{-a.s.} \quad \forall n \geq 1$$

etc.

Null sets play an important role in measure theory for the strange reason that they can be ignored for most purposes. For instance, if $F, G, F_1, F_2, \cdots$ are sets in $\mathcal{F}$, then we have

(1.19.4) $\quad F \subseteq G \quad \mu\text{-a.s.} \quad \Rightarrow \quad \mu(F) \leq \mu(G)$

(1.19.5) $\quad F = G \quad \mu\text{-a.s.} \quad \Rightarrow \quad \mu(F) = \mu(G)$

(1.19.6) $\quad F_n \cap F_k = \emptyset \quad \mu\text{-a.s.} \quad \forall n \neq k \quad \Rightarrow \quad \mu\Big(\bigcup_{n=1}^{\infty} F_n\Big) = \sum_{n=1}^{\infty} \mu(F_n)$

I shall leave the proof of these formulas to the reader, they follow straightforwardly from the formulas for contents and measures found in (1.3) and (1.4).

1.20 Operations on Measures

Let $\mathcal{F}$ be a σ-algebra on Ω, let $A \in \mathcal{F}$, let $\mu, \nu, \mu_1, \mu_2 \ldots$ be measures on $\mathcal{F}$, and let $a, a_1, a_2, \ldots$ be non-negative numbers. Then we have

(1.20.1) $\quad \kappa(F) = \mu(F) + \nu(F)$ *for* $F \in \mathcal{F}$ *is a measure on* $(\Omega, \mathcal{F})$

(1.20.2) $\quad \lambda(F) = a\,\mu(F)$ *for* $F \in \mathcal{F}$ *is a measure on* $(\Omega, \mathcal{F})$

(1.20.3) $\quad \mu_A(F) = \mu(A \cap F)$ *for* $F \in \mathcal{F}$ *is a measure on* $(\Omega, \mathcal{F})$

(1.20.4) $\quad \xi(F) = \sum_{n=1}^{\infty} a_n\,\mu_n(F)$ *for* $F \in \mathcal{F}$ *is a measure on* $(\Omega, \mathcal{F})$

Proof: All the propositions follow easily from the definition of a measure. So let me only show (1.20.4). Let $F_1, F_2, \ldots$ be disjoint sets in $\mathcal{F}$. Then we find

$$\xi\Big(\bigcup_{j=1}^{\infty} F_j\Big) = \sum_{n=1}^{\infty} a_n\,\mu_n\Big(\bigcup_{j=1}^{\infty} F_j\Big) = \sum_{n=1}^{\infty}\sum_{j=1}^{\infty} a_n\,\mu_n(F_j)$$

$$= \sum_{j=1}^{\infty}\sum_{n=1}^{\infty} a_n\,\mu_n(F_j) = \sum_{j=1}^{\infty} \xi(F_j)$$

where the interchange of the summation is justified by positivity of the terms. □

1.21 Set Functions

Apart from the more or less trivial examples given in (1.13), (1.14) and (1.15) I have not yet given a nontrivial example of a measure. For probability theory it

is of vital interest to have lots of nontrivial measures; the problem, however, is that this is not an easy task. The road to a nontrivial measure is long and arduous, but it is a road we have to take. In (1.3) and (1.4) we found a series of properties of contents and measures. The idea in the construction is to isolate the properties, analyze them, and then finally, put the pieces together to a content or measure.

Let $\mathcal{F}$ be a paving on the set Ω such that $\emptyset \in \mathcal{F}$. Recall that a function $\mu : \mathcal{F} \to [0, \infty]$ is called a *set function* if $\mu(\emptyset) = 0$. If μ is a set function on $\mathcal{F}$, then we say that μ is

- *finite* on $\mathcal{F}$ if $\mu(F) < \infty \quad \forall F \in \mathcal{F}$
- *increasing* on $\mathcal{F}$ if (1.3.2) holds whenever $F_0, F_1 \in \mathcal{F}$
- *superadditive* on $\mathcal{F}$ if (1.3.7) holds whenever $F, F_1, \ldots F_n \in \mathcal{F}$
- *subadditive* on $\mathcal{F}$ if (1.3.6) holds whenever $F, F_1, \ldots F_n \in \mathcal{F}$
- *σ-subadditive* on $\mathcal{F}$ if (1.4.1) holds whenever $F, F_1, F_2 \ldots \in \mathcal{F}$
- *additive* on $\mathcal{F}$ if (1.3.7) and (1.3.6) hold whenever $F, F_1, \ldots F_n \in \mathcal{F}$
- *σ-additive* on $\mathcal{F}$ if (1.3.7) and (1.4.1) hold whenever $F, F_1, F_2 \ldots \in \mathcal{F}$

Hence, if $\mathcal{F}$ is an algebra, then μ is a content on $\mathcal{F}$ if and only if μ is additive on $\mathcal{F}$. Similarly, if $\mathcal{F}$ is an σ-algebra, then μ is a measure on $\mathcal{F}$ if and only if μ is σ-additive on $\mathcal{F}$. Letting $n \to \infty$ in (1.3.7), we see that we have

(1.21.1) If μ is superadditive on $\mathcal{F}$, then (1.3.8) holds

Hence, there is no need for defining "σ-superadditivity." Since $\mu(\emptyset) = 0$ for any set function μ we have

(1.21.2) Any superadditive set function is increasing

(1.21.3) Any subadditive set function is increasing

(1.21.4) Any σ-subadditive set function is subadditive

Let $\mathcal{F}$ and $\mathcal{G}$ be two pavings on Ω, and let $\mu : \mathcal{F} \to \overline{\mathbf{R}}_+$ and $\nu : \mathcal{G} \to \overline{\mathbf{R}}_+$ be two set functions. Then we say μ is an *extension* of ν or that μ *extends* ν if $\mathcal{F} \supseteq \mathcal{G}$ and $\mu(G) = \nu(G)$ for all $G \in \mathcal{G}$. The point of Sections (1.22)–(1.33) is to give various solutions to the following extension problem: Let μ be a concretely given set function on $\mathcal{F}$; when does μ extend to a measure or a content?

1.22 The μ-Measurable Sets

Let Ω be a set and let μ be a set function: $2^{\Omega} \to \overline{\mathbf{R}}_+$, that is, μ is defined on all subsets of Ω . A set $M \subseteq \Omega$ is called *μ-measurable* if M splits all subsets of Ω in the following sense:

$$\mu(A) = \mu(A \cap M) + \mu(A \setminus M) \qquad \forall A \subseteq \Omega \tag{1.22.1}$$

The paving of all μ-measurable sets is denoted $\mathcal{M}(\mu)$, that is,

$$\mathcal{M}(\mu) = \{M \subseteq \Omega \mid \mu(A) = \mu(A \cap M) + \mu(A \setminus M) \quad \forall A \subseteq \Omega\} \tag{1.22.2}$$

Our next theorem shows that this paving has surprisingly good properties and it plays the main role in our construction of measures and contents.

Let $\mathcal{F}$ be a paving on Ω such that $\emptyset \in \mathcal{F}$, and let $\mu : \mathcal{F} \to \overline{\mathbf{R}}_+$ be a given non-negative function such that $\mu(\emptyset) = 0$. Then we define *the inner μ-measure* as follows:

$$\mu_*(A) = \sup\left\{ \sum_{j=1}^{n} \mu(F_j) \;\middle|\; F_1, \ldots, F_n \text{ disjoint } \in \mathcal{F} \,,\; \bigcup_{j=1}^{n} F_j \subseteq A \right\} \tag{1.22.3}$$

for all $A \subseteq \Omega$. Exactly as in (1.21.1), we find

$$\mu_*(A) = \sup\left\{ \sum_{j=1}^{\infty} \mu(F_j) \;\middle|\; F_1, F_2 \ldots \text{ disjoint } \in \mathcal{F} \,,\; \bigcup_{j=1}^{\infty} F_j \subseteq A \right\} \tag{1.22.4}$$

for all $A \subseteq \Omega$. Note that μ_* is constructed by approximation by packings from inside. There is a similar construction by approximation by coverings from outside, called *the outer μ-measure*, which is defined by

$$\mu^*(A) = \inf\left\{ \sum_{j=1}^{\infty} \mu(F_j) \;\middle|\; F_1, F_2, \ldots \in \mathcal{F} \,,\; A \subseteq \bigcup_{j=1}^{\infty} F_j \right\} \tag{1.22.5}$$

for all $A \subseteq \Omega$. Let $\mathcal{G}$ be a paving on Ω , then we say that $\mathcal{G}$ *approximates* μ on $\mathcal{F}$ if we have

$$\mu(F) \leq \sup\{\mu_*(G) \mid G \in \mathcal{G}, G \subseteq F\} \qquad \forall F \in \mathcal{F} \tag{1.22.6}$$

1.23 Carathéodory's Theorem

Let $\mu : 2^{\Omega} \to \overline{\mathbf{R}}_+$ be a set function. Then $\mathcal{M}(\mu)$ is an algebra, μ is a content on $\mathcal{M}(\mu)$, and if $M_1, M_2, \cdots$ is a sequence of disjoint μ-measurable sets, then we have

$$(1.23.1) \qquad \mu\Big(A \cap \bigcup_{j=1}^{n} M_j\Big) = \sum_{j=1}^{n} \mu(A \cap M_j)$$

$$(1.23.2) \qquad \mu\Big(A \cap \bigcup_{j=1}^{\infty} M_j\Big) = \sum_{j=1}^{\infty} \mu(A \cap M_j) + \lim_{n\to\infty} \mu\Big(A \cap \bigcup_{j=n}^{\infty} M_j\Big)$$

for all $A \subseteq \Omega$. *Moreover, if* μ *is* σ*-subadditive, then* $\mathcal{M}(\mu)$ *is a* σ*-algebra and* μ *is a measure on* $\mathcal{M}(\mu)$.

Remark: The idea in the construction of measures and contents is as follows: First we start with a "concretely" defined set function μ on a "small" paving $\mathcal{F}$. Then we use μ to define a set function on all subsets of Ω , for instance, μ_* or μ^* defined in (1.22.3) and (1.22.5). Finally, we apply Carathéodory's theorem in order to obtain a measure or a content.

Proof: Since $A \cap M^c = A \setminus M$ and $A \setminus M^c = A \cap M$, we see that $\mathcal{M}(\mu)$ is (C)-stable. Moreover, since $\mu(\emptyset) = 0$, we have that $\emptyset \in \mathcal{M}(\mu)$. Let $M_1, M_2 \in \mathcal{M}(\mu)$ and let $M = M_1 \cup M_2$. Let $A \subseteq \Omega$ be given and set $B = A \cap M$, then we have

$$\begin{aligned} B \cap M_1 &= A \cap M_1 \quad , \quad B \setminus M_1 = (A \setminus M_1) \cap M_2 \\ A \cap M &= B \qquad\qquad , \quad A \setminus M = (A \setminus M_1) \setminus M_2 \end{aligned}$$

Hence, by μ-measurability of M_1 and M_2 we find

$$\begin{aligned} \mu(A) &= \mu(A \cap M_1) + \mu(A \setminus M_1) \\ &= \mu(A \cap M_1) + \mu((A \setminus M_1) \cap M_2) + \mu((A \setminus M_1) \setminus M_2) \\ &= \mu(B \cap M_1) + \mu(B \setminus M_1) + \mu(A \setminus M) \\ &= \mu(A \cap M) + \mu(A \setminus M) \end{aligned}$$

and so $M = M_1 \cup M_2 \in \mathcal{M}(\mu)$. Hence, $\mathcal{M}(\mu)$ is $(\cup f, C)$-stable and, thus, an algebra. Suppose that $M_1, M_2, \cdots$ are disjoint μ-measurable sets and set

$$M^n = \bigcup_{j=1}^{n} M_j \quad , \quad R_n = \bigcup_{j=n}^{\infty} M_j$$

Since $M^n \cap M_n = M_n$ and $M^n \setminus M_n = M^{n-1}$, we have

$$\mu(A \cap M^n) = \mu(A \cap M_n) + \mu(A \cap M^{n-1})$$

for all $A \subseteq \Omega$. Hence, (1.23.1) follows easily by induction in n , and setting $A = \Omega$, we see that μ is a content on the algebra $\mathcal{M}(\mu)$. Moreover, since $R_1 \cap M^{n-1} = M^{n-1}$ and $R_1 \setminus M^{n-1} = R_n$, we have

$$\mu(A \cap R_1) = \mu(A \cap M^{n-1}) + \mu(A \cap R_n) = \sum_{j=1}^{n-1} \mu(A \cap M_j) + \mu(A \cap R_n)$$

Note that $R_{n+1} \subseteq R_n$ and $R_n \setminus R_{n+1} = M_n \in \mathcal{M}(\mu)$. Hence, by Exc.1.67 we have that $\{\mu(A \cap R_n)\}$ is an decreasing sequence of non-negative numbers. Thus, (1.23.2) follows by letting $n \to \infty$.

Suppose that μ is σ-subadditive on 2^Ω , and let $\{M_n\}$ be a disjoint sequence in $\mathcal{M}(\mu)$. Set $M = \cup_1^\infty M_n$, and let $A \subseteq \Omega$; then we have

$$(*) \qquad \mu(A \cap M) = \sum_{j=1}^{\infty} \mu(A \cap M_j)$$

because $\geq$ follows from (1.23.2) and $\leq$ follows from σ-subadditivity. Since $M^n = M_1 \cup \cdots \cup M_n \in \mathcal{M}(\mu)$, then by (1.23.1) we have

$$\mu(A) = \mu(A \cap M^n) + \mu(A \setminus M^n) \geq \sum_{j=1}^{n} \mu(A \cap M_j) + \mu(A \setminus M)$$

because $A \setminus M \subseteq A \setminus M^n$ and μ is increasing. Letting $n \to \infty$, we find by (*) and subadditivity of μ that

$$\mu(A) \geq \mu(A \cap M) + \mu(A \setminus M) \geq \mu(A)$$

Hence, $M \in \mathcal{M}(\mu)$, and so $\mathcal{M}(\mu)$ is $(\sum c)$-stable. But then $\mathcal{M}(\mu)$ is a σ-algebra by (1.2.6), and setting $A = \Omega$ in (*) we see that μ is a measure on $\mathcal{M}(\mu)$. □

1.24 Additivity of Inner and Outer Measures

Let $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ be a pavings on the set Ω such that $\emptyset \in \mathcal{F}$, and let μ be a set function on $\mathcal{F}$. Then we have the following:

(1.24.1) μ_* *is superadditive on* 2^Ω *and* $\mu_*(F) \geq \mu(F) \quad \forall F \in \mathcal{F}$

(1.24.2) μ *is superadditive on* $\mathcal{F}$ $\Leftrightarrow$ μ_* *is an extension of* μ

(1.24.3) μ^* *is σ-subadditive on* 2^Ω *and* $\mu^*(F) \leq \mu(F) \quad \forall F \in \mathcal{F}$

(1.24.4) *μ is σ-subadditive on $\mathcal{F}$ $\Leftrightarrow$ μ^* is an extension of μ*

(1.24.5) *If μ is superadditive and $\mathcal{F}$ is $(\sum f)$-stable, then*

$$\mu_*(A) = \sup\{\mu(F) \mid F \in \mathcal{F}\ ,\ F \subseteq A\} \qquad \forall A \subseteq \Omega$$

(1.24.6) *If μ is σ-subadditive and $\mathcal{F}$ is $(\cup c)$-stable, then*

$$\mu^*(A) = \inf\{\mu(F) \mid F \in \mathcal{F}\ ,\ F \supseteq A\} \qquad \forall A \subseteq \Omega$$

(1.24.7) *If $\mathcal{G}$ approximates μ on $\mathcal{F}$ and $G_1 \cup \cdots G_n \in \mathcal{H}$ for all disjoint sets $G_1, \ldots, G_n \in \mathcal{G}$, then $\mathcal{H}$ approximates μ_* on 2^Ω*

Proof: (1.24.1): Clearly, μ_* is increasing and $\mu_*(F) \geq \mu(F)$ for all $F \in \mathcal{F}$. Since $\mu(\emptyset) = 0$, we see that $\mu_*(\emptyset) = 0$. Let $A \supseteq A_1 \cup \cdots \cup A_k$ where $A_1, \ldots, A_k$ are disjoint subsets of Ω. I shall then show that

$$\mu_*(A) \geq \sum_{i=1}^{k} \mu_*(A_i)$$

If $\mu_*(A) = \infty$ this is trivially satisfied. So suppose that $\mu_*(A) < \infty$ and let $\varepsilon > 0$ be given. Then $\mu_*(A_i) \leq \mu_*(A) < \infty$, and so there exist disjoint sets $\{F_{i1}, \ldots, F_{i,n(i)}\}$ in $\mathcal{F}$ such that

$$\bigcup_{j=1}^{n(i)} F_{ij} \subseteq A_i \quad \text{and} \quad \mu_*(A_i) \leq \varepsilon 2^{-i} + \sum_{j=1}^{n(i)} \mu(F_{ij})$$

for all $i = 1, \ldots, k$. Since $A_1, \ldots, A_k$ are disjoint we see that the sets $\{F_{ij} \mid 1 \leq i \leq k,\ 1 \leq j \leq n(i)\}$ are disjoint sets belonging to $\mathcal{F}$ such that

$$\bigcup_{i=1}^{k} \bigcup_{j=1}^{n(i)} F_{ij} \subseteq \bigcup_{i=1}^{k} A_i \subseteq A$$

Hence, we have

$$\sum_{i=1}^{k} \mu_*(A_i) \leq \sum_{i=1}^{k} \varepsilon 2^{-i} + \sum_{i=1}^{k} \sum_{j=1}^{n(i)} \mu(F_{ij}) \leq \varepsilon + \mu_*(A)$$

Letting $\varepsilon \to 0$, we see that μ_* is superadditive.

(1.24.2): Suppose that μ is superadditive and let $F \in \mathcal{F}$. If $F_1, \ldots, F_n$ are disjoint sets in $\mathcal{F}$ such that $F \supseteq F_1 \cup \cdots \cup F_n$, then by superadditivity of μ we have $\mu(F) \geq \mu(F_1) + \cdots + \mu(F_n)$. Hence $\mu(F) \geq \mu_*(F)$, but the

converse inequality holds by (1.22.3), and so μ_* is an extension of μ . The converse implication follows from (1.24.1).

(1.24.3) and (1.24.4) are proved in exactly the same manner.

(1.24.5): Suppose that μ is superadditive and that $\mathcal{F}$ is $(\sum f)$-stable. Let $A \subseteq \Omega$ and let $F_1, \ldots, F_n$ be disjoint sets in $\mathcal{F}$ such that $F_0 = \cup_1^n F_j \subseteq A$. Then $F_0 \in \mathcal{F}$ and

$$\sum_{j=1}^{n} \mu(F_j) \leq \mu(F_0) \leq \sup\{\mu(F) \mid F \in \mathcal{F}, F \subseteq A\}$$

Hence, we have

$$\mu_*(A) \leq \sup\{\mu(F) \mid F \in \mathcal{F}, F \subseteq A\}$$

and the converse inequality is evident.

(1.24.6) is proved similarly.,

(1.24.7): Suppose that $\mathcal{G}$ approximates μ on $\mathcal{F}$, and let $A \subseteq \Omega$ be given. If $a < \mu_*(A)$ then by (1.22.3) there exist disjoint set $F_1, \ldots, F_n \in \mathcal{F}$ such that $a < \mu(F_1) + \cdots + \mu(F_n)$ and $F_1 \cup \cdots \cup F_n \subseteq A$. Hence, by (1.22.6) we can find $G_1, \ldots, G_n \in \mathcal{G}$ such that $G_j \subseteq F_j$ and $a < \mu_*(G_1) + \cdots + \mu_*(G_n)$. Then $G_1, \ldots, G_n$ are disjoint. Hence, if we set $H_0 = G_1 \cup \cdots \cup G_n$ and $c = \sup\{\mu_*(H) \mid H \in \mathcal{H}\ ,\ H \subseteq A\}$, then $H_0 \in \mathcal{H}$ by assumption, $H \subseteq A$ and by superadditivity of μ_* , we get

$$a < \mu_*(G_1) + \cdots + \mu_*(G_n) \leq \mu_*(H) \leq c$$

Letting $a \uparrow \mu_*(A)$ we see that $\mu_*(A) \leq c$. Thus, (1.24.7) is proved. □

1.25 Tightness

Let $\mu : \mathcal{F} \to \overline{\mathbf{R}}_+$ be a non-negative set function. Then the inner and outer μ-measures, μ_* and μ^* , are everywhere defined set functions. Hence, by Carathéodory's theorem we know that μ_* is a content on the algebra $\mathcal{M}(\mu_*)$, and μ^* is a measure on the σ-algebra $\mathcal{M}(\mu^*)$ because μ^* is σ-subadditive on 2^Ω by (1.24.3). To solve the extension problem stated in (1.21) we would like to decide when $\mathcal{F} \subseteq \mathcal{M}(\mu_*)$ and μ_* is an extension of μ , or when $\mathcal{F} \subseteq \mathcal{M}(\mu^*)$ and μ^* is an extension of μ . We define ourselves out of these problems: Let $\mathcal{F}$ be a paving on Ω , and let $\mu : \mathcal{F} \to \overline{\mathbf{R}}_+$ be a non-negative set function. Then we say that μ is *inner tight* if

(1.25.1) $\quad \mu(F_1) = \mu_*(F_1 \cap F_2) + \mu_*(F_1 \setminus F_2) \qquad \forall F_1, F_2 \in \mathcal{F}$

Similarly, we say that μ is *outer tight* if

(1.25.2) $\mu(F_1) = \mu^*(F_1 \cap F_2) + \mu^*(F_1 \setminus F_2) \qquad \forall F_1, F_2 \in \mathcal{F}$

Then we have the following criteria for inner and outer tightness.

1.26 Criteria for Inner and Outer Tightness

Let $\mathcal{F}$ be a paving on the set Ω and let μ be a set function on $\mathcal{F}$. Then we have

(1.26.1) μ *is inner tight* $\iff$ μ_* *is an extension of* μ *and* $\mathcal{F} \subseteq \mathcal{M}(\mu_*)$

(1.26.2) μ *is outer tight* $\iff$ μ^* *is an extension of* μ *and* $\mathcal{F} \subseteq \mathcal{M}(\mu^*)$

If μ is a superadditive on $\mathcal{F}$ and $\mathcal{F}$ is a $(\cap f)$-stable paving such that for every $F', F \in \mathcal{F}$ with $F' \supseteq F$, there exist disjoint sets $F_1, F_2 \cdots \in \mathcal{F}$, satisfying

(1.26.3) $F' \setminus F \supseteq \bigcup_{j=1}^{\infty} F_j$ *and* $\mu(F') \leq \mu(F) + \sum_{j=1}^{\infty} \mu(F_j)$

then μ is inner tight. In particular, every content μ on an algebra $\mathcal{F}$ is inner tight.

Similarly, if μ is σ-subadditive on the $(\cap f)$-stable paving $\mathcal{F}$ such that for every $F', F \in \mathcal{F}$, with $F' \supseteq F$, there exist sets $F_1, F_2, \cdots \in \mathcal{F}$ satisfying

(1.26.4) $F' \setminus F \subseteq \bigcup_{j=1}^{\infty} F_j$ *and* $\mu(F') \geq \mu(F) + \sum_{j=1}^{\infty} \mu(F_j)$

then μ is outer tight. In particular, every σ-subadditive content μ on an algebra is outer tight.

Proof: (1.26.1) "$\Rightarrow$" : Suppose that μ is inner tight. Setting $F_2 = \emptyset$ in (1.25.1), we see that μ_* is an extension of μ. Let $F \in \mathcal{F}$ and let $A \subseteq \Omega$ be given. Suppose that $F_1, \ldots, F_n$ are disjoint sets in $\mathcal{F}$ such that $F_1 \cup \cdots \cup F_n \subseteq A$; then by (1.25.1) and (1.24.1) we have

$$\begin{aligned}\sum_{j=1}^{n} \mu(F_j) &= \sum_{j=1}^{n} \mu_*(F_j \cap F) + \sum_{j=1}^{n} \mu_*(F_j \setminus F) \\ &\leq \mu_*(A \cap F) + \mu_*(A \setminus F) \leq \mu_*(A)\end{aligned}$$

Hence, taking supremum over all $F_1, \ldots, F_n$, we get $\mu_*(A) = \mu_*(A \cap F) + \mu_*(A \setminus F)$ for all $A \subseteq \Omega$ and all $F \in \mathcal{F}$ and, thus, $\mathcal{F} \subseteq \mathcal{M}(\mu_*)$.

(1.26.1) "$\Leftarrow$" : Suppose that $\mathcal{F} \subseteq \mathcal{M}(\mu_*)$ and that μ_* is an extension of μ . Then $\mathcal{M}(\mu_*)$ is an algebra and μ_* is a content on $\mathcal{M}(\mu_*)$ by Carathéodory's theorem (1.23). Hence, by (1.3.1) we have

$$\mu(F_1) = \mu_*(F_1) = \mu_*(F_1 \cap F_2) + \mu_*(F_1 \setminus F_2) \qquad \forall F_1, F_2 \in \mathcal{F}$$

and so μ is inner tight.

(1.26.2) is proved similarly.

Suppose that μ is superadditive and satisfies (1.26.3). Let $F_1, F_2 \in \mathcal{F}$ be given and set $F' = F_1$ and $F = F_1 \cap F_2$. Then $F' \supseteq F$ and $F', F \in \mathcal{F}$ by $(\cap f)$-stability of $\mathcal{F}$. So let $F_1', F_2', \cdots \in \mathcal{F}$ be disjoint sets satisfying (1.26.3), then by (1.21.1), (1.22.4) (1.24.1) and (1.24.2) we have

$$\begin{aligned}\mu(F_1) = \mu(F') &\leq \mu(F) + \sum_{j=1}^{\infty} \mu(F_j') \leq \mu(F_1 \cap F_2) + \mu_*(F_1 \setminus F_2) \\ &= \mu_*(F_1 \cap F_2) + \mu_*(F_1 \setminus F_2) \leq \mu_*(F_1) = \mu(F_1)\end{aligned}$$

Hence, μ is inner tight. The last part of the proposition is proved similarly., □

1.27 The Extension Theorem

Let $\mathcal{F}$ be a paving on Ω and let $\mu : \mathcal{F} \to \overline{\mathbf{R}}_+$ be a non-negative set function. Then we have the following:

(1): If μ is outer tight, then there exists a measure μ_0 on $\sigma(\mathcal{F})$ such that μ_0 is an extension of μ .

(2): If μ is inner tight, then there exists a content μ_1 on $\alpha(\mathcal{F})$ such that μ_1 is an extension of μ .

(3): If μ is inner tight and μ_* is σ-subadditive on $\alpha(\mathcal{F})$, then there exists a measure μ_2 on $\sigma(\mathcal{F})$ such that μ_2 is an extension of μ .

Remarks: (1): In Sections (1.32), (1.33) and (1.50) we shall see how this theorem allows us to construct all the measures we want. Note that if $\mathcal{F}$ is $(\cap f)$-stable and if there exists $F_n \in \mathcal{F}$ such that $\mu(F_n) < \infty$ for all $n = 1, 2, \cdots$ and $\Omega = \cup_1^\infty F_n$, then by the uniqueness theorem (1.7), there exists at most one measure on $\sigma(\mathcal{F})$, which extends μ .

(2): By the proof below we have that the measure μ_0 in (1) is the restriction of the outer μ-measure μ^* to the σ-algebra $\sigma(\mathcal{F})$. The content μ_1 in (2) is the restriction of the inner μ-measure μ_* to the algebra $\alpha(\mathcal{F})$. The measure μ_2 in (3) is a bit more complicated: if ν is the restriction of the inner measure μ_* to the algebra $\alpha(\mathcal{F})$, then μ is the restriction of the outer ν -measure ν^* to the σ-algebra $\sigma(\mathcal{F})$.

Proof: (1): By (1.24.3) and (1.23) we have that $\mathcal{M}(\mu^*)$ is a σ-algebra and that μ^* is a measure on $\mathcal{M}(\mu^*)$. By (1.26.2) we have $\mathcal{F} \subseteq \mathcal{M}(\mu^*)$ and μ^* is an extension of μ. Hence, $\sigma(\mathcal{F}) \subseteq \mathcal{M}(\mu^*)$ and if μ_0 is the restriction of μ^* to $\sigma(\mathcal{F})$, then μ_0 is a measure on $\sigma(\mathcal{F})$ which extends μ.

(2): Exactly as above we have that $\alpha(\mathcal{F}) \subseteq \mathcal{M}(\mu_*)$, and if μ_1 is the restriction of μ_* to $\alpha(\mathcal{F})$, then μ_1 is a content, which extends μ.

(3): Let μ_1 be the restriction of μ_* to $\alpha(\mathcal{F})$. Then from the argument above, we have that μ_1 is a content which extends μ. By assumption we have that μ_1 is σ-subadditive on $\alpha(\mathcal{F})$, and so by (1.26), we have that μ_1 is outer tight. Hence by (1) there exists a measure μ_2 on $\sigma(\mathcal{F})$, which extends μ_1 and thus, μ. □

1.28 Completion of Measures

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and consider the outer μ-measure μ^* which, by (1.24.6), is given by

(1.28.1) $$\mu^*(A) = \inf\{\mu(F) \mid F \in \mathcal{F}\ ,\ F \supseteq A\} \qquad \forall A \supseteq \Omega$$

By (1.24.3) and (1.24.4) we know that μ^* is σ-subadditive on 2^Ω and that μ^* is an extension of μ. Moreover, by (1.26) we have μ is outer tight on $\mathcal{F}$, and so $\mathcal{F} \subseteq \mathcal{M}(\mu^*)$. Hence, by (1.23) we see that $\mathcal{M}(\mu^*)$ is a σ-algebra and μ^* is σ-additive on $\mathcal{M}(\mu^*)$. Let us define

(1.28.2) $$\mathcal{M}(\mu) := \mathcal{M}(\mu^*) \quad \text{and} \quad \bar{\mu} := \text{the restriction of } \mu^* \text{ to } \mathcal{M}(\mu)$$

The σ-algebra $\mathcal{M}(\mu)$ is called *the completion* of $\mathcal{F}$ under μ and $\bar{\mu}$ is called *the completion* of μ. By the argument above we have

(1.28.3) $\mathcal{M}(\mu)$ is a σ-algebra on Ω containing $\mathcal{F}$, and $\bar{\mu}$ is a measure on $(\Omega, \mathcal{M}(\mu))$, which extends μ

A set $A \in \mathcal{M}(\mu)$ is said to be *μ-measurable*, and if $\mathcal{F} = \mathcal{M}(\mu)$ we say that μ is *complete*.

Let $A \subseteq \Omega$; then we say that A is *Lebesgue μ-measurable* if there exists $F_0, F_1 \in \mathcal{F}$ such that $F_0 \subseteq A \subseteq F_1$ and $\mu(F_1 \setminus F_0) = 0$. The paving of all Lebesgue μ-measurable sets is denoted $\mathcal{L}(\mu)$. We say that the measure μ is *σ-finite* if there exists a sequence $\{F_n\}$ in $\mathcal{F}$ such that $\mu(F_n) < \infty$ and $\Omega = \cup_1^\infty F_n$. It is a straightforward exercise [see Exc.1.39] to show the following:

(1.28.4) $\mathcal{L}(\mu)$ is a σ-algebra with $\mathcal{F} \subseteq \mathcal{L}(\mu) \subseteq \mathcal{M}(\mu)$

(1.28.5) $\mu_*(A) = \mu^*(A) < \infty \quad \Rightarrow \quad A \in \mathcal{L}(\mu)$

(1.28.6) $\mu_*(A) = \mu^*(A) \qquad \forall A \in \mathcal{L}(\mu)$

(1.28.7) If μ is σ-finite, then $\mathcal{L}(\mu) = \mathcal{M}(\mu)$

If μ is a measure, then in general the paving of all μ-measurable sets is much larger than the original σ-algebra $\mathcal{F}$. Thus, the extension theorem allows us to extend measures to a (usually) much larger σ-algebra. If μ is just a set function on some paving $\mathcal{F}$, then in order to apply Theorem 3.7 we need "easy to check" criteria for inner and outer tightness; such criteria are provided by (1.26). Moreover, we need a good criterion for σ-subadditivity of μ_* on $\alpha(\mathcal{F})$, and we shall do this in terms of approximating compact pavings. A paving $\mathcal{K}$ is said to be *compact* if for every sequence $\{K_n\}$ in $\mathcal{K}$ with $\cap_1^\infty K_n = \emptyset$ there exists an integer m such that $K_1 \cap \cdots \cap K_m = \emptyset$; that is, $\mathcal{K}$ is compact if and only if

(1.28.8) $K_n \in \mathcal{K} \ \ \forall n$ and $\bigcap_{n=1}^{\infty} K_n = \emptyset \quad \Rightarrow \quad \exists m \ : \ \bigcap_{j=1}^{m} K_j = \emptyset$

Suppose that Ω is a metric space and that $\mathcal{K}$ is a paving on Ω such that every $K \in \mathcal{K}$ is compact. If $K_n \in \mathcal{K}$ and $\cap_1^\infty K_n = \emptyset$, then K_n^c is open and $K_1 \subseteq \Omega = \cup_1^\infty K_n^c$. Hence, by compactness of K_1 there is an integer m such that

$$K_1 \subseteq \bigcup_{j=1}^{m} K_j^c = \left(\bigcap_{j=1}^{m} K_j \right)^c$$

Hence, $K_1 \cap \cdots \cap K_m = \emptyset$ because $K_1 \cap \cdots \cap K_m \subseteq K_1$. Thus, we have proved

(1.28.9) If Ω is a metric space and $\mathcal{K}$ is a paving consisting of compact subsets of Ω, then $\mathcal{K}$ is a compact paving

Note the paving of all compact subsets of a metric space is $(\cup f, \cap c)$-stable; this property carries over to arbitrary compact pavings, in the following sense:

1.29 Stability of Compact Pavings

If $\mathcal{K}$ is compact paving on the set Ω and $\mathcal{C}$ is the smallest $(\cap c, \cup f)$-stable paving containing $\mathcal{K}$, then $\mathcal{C}$ is a compact paving.

Proof: Let $\mathcal{C}_0$ be the paving of all finite unions of sets from $\mathcal{K}$; then I claim that $\mathcal{C}_0$ is compact. So suppose for a moment that $\mathcal{C}_0$ is **not** compact. Then there exists $\{C_n\} \in \mathcal{C}_0$ such that $\cap_1^\infty C_n = \emptyset$ and $C^n \neq \emptyset$ for all n , where $C^n = C_1 \cap \cdots \cap C_n$. By the definition of $\mathcal{C}_0$ there exist sets $K_1, \cdots, K_q$ in $\mathcal{K}$ such that $C_1 = K_1 \cup \cdots \cup K_q$. Let

$$N_j = \{n \in \mathbf{N} \mid K_j \cap C^n \neq \emptyset\}$$

Since

$$C^n = C_1 \cap C^n = \bigcup_{j=1}^{q} K_j \cap C^n \neq \emptyset \qquad \forall n \in \mathbf{N}$$

we have $\mathbf{N} = N_1 \cup \cdots \cup N_q$, and so at least one of the sets $N_1, \ldots, N_q$ is infinite, and by renumerating the K_j 's if necessary we may assume that N_1 is infinite. Observe that if $n \in N_1$, then $1, \ldots, n \in N_1$ because $C^1 \supseteq C^2 \supseteq \cdots \supseteq C^n$. Thus, we have $K_1 \in \mathcal{K}$, $K_1 \subseteq C_1$ and $K_1 \cap C^n \neq \emptyset$ for all $n \geq 1$. Repeating the same argument there exist a set $K_2 \in \mathcal{K}$ satisfying $K_2 \subseteq C_2$ and $K_1 \cap K_2 \cap C^n \neq \emptyset$ for all $n \geq 1$, etc. Hence, in this way we may construct a sequence $\{K_n\}$ in $\mathcal{K}$ such that

$$K_n \subseteq C_n \quad \text{and} \quad K_1 \cap \cdots \cap K_p \cap C^n \neq \emptyset \qquad \forall n, p \geq 1$$

But then $\cap_1^\infty K_n = \emptyset$ and $K_1 \cap \cdots \cap K_p \neq \emptyset$ for all $p \in \mathbf{N}$, which contradicts the compactness of $\mathcal{K}$. Hence, our assumption is false and so $\mathcal{C}_0$ must be compact. By definition of compactness we see that the paving $\mathcal{C}_1$ of all countable intersections of sets from $\mathcal{C}_0$ is compact. Moreover, $\mathcal{C}_1$ is, evidently, a $(\cap c)$-stable paving containing $\mathcal{K}$. Let $L, M \in \mathcal{C}_1$ be given; then there exists $L_i, M_j \in \mathcal{C}_0$ such that $L = \cap_1^\infty L_i$ and $M = \cap_1^\infty M_j$. Hence we have

$$L \cup M = \left(\bigcap_{i=1}^{\infty} L_i\right) \cup \left(\bigcap_{j=1}^{\infty} M_j\right) = \bigcap_{i=1}^{\infty} \bigcap_{j=1}^{\infty} (L_i \cup M_j)$$

Since $L_i \cup M_j \in \mathcal{C}_0$ we see that $L \cup M \in \mathcal{C}_1$. Thus, $\mathcal{C}_1$ is a compact $(\cap c, \cup f)$-stable paving containing $\mathcal{K}$, and so we have proved the proposition. □

1.30 The Compactness Lemma

Let $\mathcal{F}$ be a paving and let $\mu : \mathcal{F} \to \overline{\mathbf{R}}_+$ be an inner tight set function. If there exists a compact paving $\mathcal{K}$ which approximates μ on $\mathcal{F}$, then we have

(1.30.1) $\lim_{n\to\infty} \mu_*(A_n) = 0$ *if* $A_n \downarrow \emptyset$ *and* $\mu_*(A_1) < \infty$

(1.30.2) μ_* *is σ-subadditive on* $\mathcal{M}(\mu_*)$

Proof: (1.30.1): Let $A_n \downarrow \emptyset$ such that $\mu_*(A_1) < \infty$, and let $\varepsilon > 0$ be given. Let $\mathcal{C}$ be defined as in (1.29), then $\mathcal{C}$ is a compact $(\cap c, \cup f)$-stable paving. By (1.24.7) there exist $C_n \in \mathcal{C}$ such that $C_n \subseteq A_n$ and $\mu_*(A_n) \leq \varepsilon 2^{-n-1} + \mu_*(C_n)$. Since $\mathcal{F}$ evidently approximates μ on $\mathcal{F}$ and $\mathcal{F} \subseteq \mathcal{M}(\mu_*)$, then by (1.23) and (1.24.7) there exists $M_n \in \mathcal{M}(\mu_*)$ such that $M_n \subseteq C_n \subseteq A_n$ and $\mu_*(C_n) \leq \varepsilon 2^{-n-1} + \mu_*(M_n)$. Moreover, since $M_n \subseteq A_n$, we find

$$\mu_*(A_n \setminus M_n) + \mu_*(M_n) = \mu_*(A_n) \leq \varepsilon 2^{-n-1} + \mu_*(C_n) \leq \varepsilon 2^{-n} + \mu_*(M_n)$$

and so we see that $\mu_*(A_n \setminus M_n) \leq \varepsilon 2^{-n}$ for all $n \geq 1$. Since $C_n \subseteq A_n$ we have that $\cap_1^\infty C_n = \emptyset$; so by (1.29) there exists an integer $p \geq 1$ with $C_1 \cap \cdots \cap C_p = \emptyset$. Collecting all these facts, we get

$$\begin{aligned}
\mu_*(A_n) &\leq \varepsilon\, 2^{-n-1} + \mu_*(C_n) \leq \varepsilon 2^{-n} + \mu_*(M_n) \\
&= \varepsilon 2^{-n} + \mu_*\left(M_n \cap \bigcap_{j=1}^{p} M_j\right) + \mu_*\left(M_n \setminus \bigcap_{j=1}^{p} M_j\right) \\
&\leq \varepsilon 2^{-n} + \sum_{j=1}^{p} \mu_*(M_n \setminus M_j)
\end{aligned}$$

because $M_1 \cap \cdots \cap M_p \subseteq C_1 \cap \cdots \cap C_p = \emptyset$ and μ_* is a content on $\mathcal{M}(\mu_*)$. Suppose that $n > p \geq j$; then $M_n \subseteq A_n \subseteq A_j$, and we find

$$\mu_*(A_n) \leq \varepsilon 2^{-n} + \sum_{j=1}^{p} \mu_*(A_j \setminus M_j) \leq \sum_{j=1}^{\infty} \varepsilon 2^{-j} = \varepsilon$$

for $n > p$. Hence, $\mu_*(A_n) \to 0$ and so (1.30.1) holds.

(1.30.2): Let $M, M_1, M_2, \ldots \in \mathcal{M}(\mu_*)$ such that $M \subseteq \cup_1^\infty M_n$. I shall then show

$$(*) \qquad \mu_*(M) \leq \sum_{n=1}^{\infty} \mu_*(M_n)$$

Let a denote the infinite sum on the right-hand side of (*), and suppose that (*) fails, that is, that $\mu_*(M) > a$. Let us disjungate $\{M_n\}$ in the usual way [see (1.2.11–(1.2.13)]: $B_1 = M_1$ and $B_n = M_n \setminus (M_1 \cup \cdots \cup M_{n-1})$ for $n \geq 2$. Then $B_1, B_2, \cdots$ is a disjoint sequence in $\mathcal{M}(\mu_*)$ and

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} M_n \supseteq M$$

Hence if $C \subseteq M$ such that $\mu_*(C) > a$, then setting $R_n = \cup_{j=n}^{\infty} B_j$ and using (1.23.2), we get

$$\begin{aligned} a < \mu_*(C) &= \sum_{n=1}^{\infty} \mu_*(C \cap B_n) + \lim_{n\to\infty} \mu_*(C \cap R_n) \\ &\le a + \lim_{n\to\infty} \mu_*(C \cap R_n) \end{aligned}$$

because $C \cap B_n \subseteq M_n$. Since $B_1, B_2, \ldots$ are disjoint, we have that $R_n \downarrow \emptyset$. Hence, by (1.30.1) we see that $\mu_*(C \cap R_n) = \infty$ for all $C \subseteq M$ with $\mu_*(C) > a$. Since $\mu_*(M) > a$, then by (1.24.7) there exists $C_0 \in \mathcal{C}$ such that $C_0 \subseteq M$ and $\mu_*(C_0) > a$. Then $\mu_*(C_0 \cap R_1) = \infty > a$, and so there exists $C_1 \in \mathcal{C}$ such that $C_1 \subseteq C_0 \cap R_1$ and $\mu_*(C_1) > a$. Continuing like this we can find $C_n \in \mathcal{C}$ such that

$$C_{n+1} \subseteq C_n \cap R_{n+1} \quad \text{and} \quad \mu_*(C_n) > a \qquad \forall n \ge 0$$

Hence, $C_n \downarrow \emptyset$ because $R_n \downarrow \emptyset$. Thus, by (1.29) there exists $m \in \mathbf{N}$ such that $C_m = C_0 \cap \cdots \cap C_m = \emptyset$, but this is impossible since $\mu_*(\emptyset) = 0$. Hence, (*) must hold and (1.30.2) is proved. □

1.31 Inner Tight Set Functions on Compact Pavings

Let $\mathcal{K}$ be a $(\cup f, \cap c)$-stable compact paving on Ω and let $\mu : \mathcal{K} \to \mathbf{R}_+$ be a finite, inner tight set function on $\mathcal{K}$. Then μ_* is an extension of μ and $\mathcal{K} \subseteq \mathcal{M}(\mu_*)$. Moreover, we have the following:

(1.31.1) $\mu_*(A) = \sup\{\mu(K) \mid K \in \mathcal{K}\ ,\ K \subseteq A\} \qquad \forall A \subseteq \Omega$

(1.31.2) $\lim_{n\to\infty} \mu_*(A_n) = \mu_*(A)$ *if* $A_n \downarrow A$ *and* $\mu_*(A_1) < \infty$

(1.31.3) $\mathcal{M}(\mu_*)$ *is a σ-algebra on Ω and μ_* is a measure on* $(\Omega, \mathcal{M}(\mu_*))$

(1.31.4) $M \in \mathcal{M}(\mu_*) \iff M \cap K \in \mathcal{M}(\mu_*) \quad \forall K \in \mathcal{K}$

Proof: (1.31.1): By (1.26.1) we have that μ_* is an extension of μ and $\mathcal{K} \subseteq \mathcal{M}(\mu_*)$. So by (1.24.2) we have that μ is superadditive and (1.31.1) follows from (1.24.5).

(1.31.2): Let $A_n \downarrow A$ and $\mu_*(A_1) < \infty$, and let $\varepsilon > 0$ be given. Then by (1.31.1) there exists $K_1 \in \mathcal{K}$ with $K_1 \subseteq A_1$ and $\mu_*(A_1) \le \frac{1}{2}\varepsilon + \mu(K_1)$. Similarly, we can find $K_2 \in \mathcal{K}$ with $K_2 \subseteq A_2 \cap K_1$ and $\mu_*(A_2 \cap K_1) \le$

$\frac{1}{4}\varepsilon + \mu(K_2)$. Continuing like this we can find a decreasing sequence $\{K_n\} \subseteq \mathcal{K}$ satisfying

$$K_1 \subseteq A_1 \quad , \quad \mu_*(A_1) \leq \tfrac{\varepsilon}{2} + \mu(K_1)$$

$$K_n \subseteq A_n \cap K_{n-1} \quad , \quad \mu_*(A_n \cap K_{n-1}) \leq 2^{-n}\varepsilon + \mu(K_n) \qquad \forall n \geq 2$$

Since $K_j \in \mathcal{K} \subseteq \mathcal{M}(\mu_*)$, $A_{j+1} \subseteq A_j$ and $K_j \subseteq A_j$ we find:

$$\begin{aligned}
\mu_*(A_{j+1}) + \mu(K_j) &= \mu_*(A_{j+1} \cap K_j) + \mu_*(A_{j+1} \setminus K_j) + \mu(K_j) \\
&\leq 2^{-j-1}\varepsilon + \mu(K_{j+1}) + \mu_*(A_j \setminus K_j) + \mu(K_j) \\
&\leq 2^{-j-1}\varepsilon + \mu(K_{j+1}) + \mu_*(A_j \setminus K_j) + \mu_*(A_j \cap K_j) \\
&\leq 2^{-j-1}\varepsilon + \mu(K_{j+1}) + \mu_*(A_j)
\end{aligned}$$

Set $K = \cap_1^\infty K_n$, then $K \subseteq A$ and $K \in \mathcal{K} \subseteq \mathcal{M}(\mu_*)$ by $(\cap c)$-stability of $\mathcal{K}$. Thus, we get

$$\begin{aligned}
\mu_*(A_n) &= \mu_*(A_1) + \sum_{j=1}^{n-1} \big(\mu_*(A_{j+1}) - \mu_*(A_j)\big) \\
&\leq \tfrac{\varepsilon}{2} + \mu(K_1) + \sum_{j=1}^{n-1} \big(2^{-j-1}\varepsilon + \mu_*(K_{j+1}) - \mu_*(K_j)\big) \\
&\leq \varepsilon + \mu(K_n) = \varepsilon + \mu_*(K) + \mu_*(K_n \setminus K) \\
&\leq \varepsilon + \mu_*(A) + \mu_*(K_n \setminus K)
\end{aligned}$$

and since $A \subseteq A_n$ and $K_n \setminus K \downarrow \emptyset$, then by (1.30.1) we find

$$\mu_*(A) \leq \lim_{n\to\infty} \mu_*(A_n) \leq \lim_{n\to\infty} \big\{\varepsilon + \mu_*(A) + \mu_*(K_n \setminus K)\big\} = \varepsilon + \mu_*(A)$$

Letting $\varepsilon \downarrow 0$, we see that (1.31.2) holds.

(1.31.3): By (1.30.2) and Carathéodory's theorem (1.23), we know that $\mathcal{M}(\mu_*)$ is an algebra and μ_* is σ-additive on $\mathcal{M}(\mu_*)$. Let $M_n \in \mathcal{M}(\mu_*)$ so that $M_n \downarrow M$, and let $A \subseteq \Omega$ be given. Then we have

$$\mu(K) = \mu_*(K \cap M_n) + \mu_*(K \setminus M_n) \leq \mu_*(K \cap M_n) + \mu_*(A \setminus M)$$

for all $K \in \mathcal{K}$ with $K \subseteq A$. Letting $n \to \infty$ and using (1.31.2), we get

$$\mu(K) \leq \mu_*(K \cap M) + \mu_*(A \setminus M) \leq \mu_*(A \cap M) + \mu_*(A \setminus M)$$

because $\mu(K) < \infty$ by assumption. Taking supremum over all $K \in \mathcal{K}$ with $K \subseteq A$, then by (1.31.1) we find

$$\mu_*(A) \leq \mu_*(A \cap M) + \mu_*(A \setminus M) \qquad \forall A \subseteq \Omega$$

But the converse inequality follows from superadditivity of μ_* and so $M \in \mathcal{M}(\mu_*)$. Hence, $\mathcal{M}(\mu_*)$ is a $(\downarrow)$-stable algebra and, thus, a σ-algebra.

(1.31.4): Since $\mathcal{K} \subseteq \mathcal{M}(\mu_*)$, we have that $K \cap M \in \mathcal{M}(\mu_*)$ for all $K \in \mathcal{K}$ and all $M \in \mathcal{M}(\mu_*)$. So suppose that M is a subset of Ω such that $M \cap K \in \mathcal{M}(\mu_*)$ for all $K \in \mathcal{K}$, and let $A \subseteq \Omega$ be given. Let $K \in \mathcal{K}$ such that $K \subseteq A$; then

$$\begin{aligned}\mu(K) &= \mu_*(K \cap (M \cap K)) + \mu_*(K \setminus (M \cap K)) \\ &\leq \mu_*(A \cap M) + \mu_*(A \setminus M) \leq \mu_*(A)\end{aligned}$$

Hence, taking supremum over all $K \in \mathcal{K}$ with $K \subseteq A$, we find that $M \in \mathcal{M}(\mu_*)$. □

1.32 Lebesgue-Stieltjes Measures on R

Let $F : \mathbf{R} \to \mathbf{R}$ be an increasing function which is *right continuous*, that is,

(1.32.1) $\quad x_n \downarrow x \quad \Rightarrow \quad F(x) = \lim_{n\to\infty} F(x_n)$

where $x_n \downarrow x$ means that $x_1 \geq x_2 \geq \cdots$ and $\lim_{n\to\infty} x_n = x$. Let

$$\mathcal{F} = \{\,]a, b] \mid a \leq b \,\} \quad , \quad \mu(]a, b]) = F(b) - F(a)$$

Then $\mathcal{F}$ is a $(\cap f)$-stable paving and μ is a set function on $\mathcal{F}$. Let $I \in \mathcal{F}$ and let $I_1, \ldots, I_n$ be disjoint sets in $\mathcal{F}$ so that $I \supseteq I_1 \cup \cdots \cup I_n$. Renumerating the I_j 's if necessary we may assume that $I =]a_0, a_{2n+1}]$ and $I_j =]a_{2j-1}, a_{2j}]$ for $j = 1, \ldots n$, where $a_0 \leq a_1 \leq a_2 \leq \cdots \leq a_{2n+1}$. Since F is increasing, we have

$$\begin{aligned}\mu(I) = F(a_{2n+1}) - F(a_0) &= \sum_{j=0}^{2n} (F(a_{j+1}) - F(a_j)) \\ &\geq \sum_{j=1}^{n} (F(a_{2j}) - F(a_{2j-1})) = \sum_{j=1}^{n} \mu(I_j)\end{aligned}$$

Hence, μ is superadditive on $\mathcal{F}$. Moreover, if $I =]a, b]$ and $J =]c, d] \subseteq I$, then

$$I \setminus J =]a, c] \cup]d, b] = I_1 \cup I_2$$

$$\mu(I) = F(b) - F(a) = \mu(J) + \mu(I_1) + \mu(I_2)$$

So by (1.26) we have that μ is inner tight. Since F is right continuous, it follows that the paving $\mathcal{K} = \{ [a, b] \mid a \leq b\}$ approximates μ on $\mathcal{F}$. Moreover, $\mathcal{K}$ is compact by (1.28.9). So by (1.30), (1.27), (1.10.1) and (1.7) we find

(1.32.2) If $F : \mathbf{R} \to \mathbf{R}$ is increasing and right continuous, then there exists a unique measure, denoted λ_F, on $\mathcal{B}(\mathbf{R})$ satisfying

$$\lambda_F(\,]a, b]\,) = F(b) - F(a) \qquad \forall\, a \leq b$$

The measure λ_F is called the *Lebesgue-Stieltjes measure induced by* F. Since $]-n, n] \uparrow \mathbf{R}$, we have by (1.4.2) that

(1.32.3) $$\lambda_F(\mathbf{R}) = F(\infty) - F(-\infty)$$

where $F(\infty)$ and $F(-\infty)$ are defined by

(1.32.4) $$F(\infty) := \lim_{x \to \infty} F(x) \quad \text{and} \quad F(-\infty) = \lim_{x \to -\infty} F(x)$$

Setting $F(x) = x$, we get:

(1.32.5) There exists a unique measure, usually denoted λ or λ_1, on $\mathcal{B}(\mathbf{R})$ satisfying $\lambda(]a, b]) = b - a$ for all $a \leq b$

This measure is called the *Lebesgue measure* on $\mathbf{R}$. Note that $\lambda(I)$ equals the length of I whenever I is an interval, and so $\lambda(B)$ may be interpreted as *the length* of B whenever B is a Borel subset of $\mathbf{R}$.

1.33 Lebesgue-Stieltjes Measures on $\mathbf{R}^k$

Let $k \geq 1$, and let $F : \mathbf{R}^k \to \mathbf{R}$ be a *right continuous function*, that is,

(1.33.1) $$x_{1n} \downarrow x_1 \cdots x_{kn} \downarrow x_k \;\Rightarrow\; F(x_1, \ldots, x_k) = \lim_{n \to \infty} F(x_{1n}, \ldots, x_{kn})$$

If $a = (a_1, \ldots, a_k)$ and $b = (b_1, \ldots, b_k)$ we define

$$a \leq b \iff a_j \leq b_j \quad \forall j = 1, \ldots, k$$

$$a < b \iff a_j < b_j \quad \forall j = 1, \ldots, k$$

$$]a, b] = \{ x \in \mathbf{R}^k \mid a < x \leq b \}$$

and we define *the increment* of F over the box $]a, b]$ as

(1.33.2) $$\Delta_a^b F = \sum_{\varepsilon_1=0}^{1} \cdots \sum_{\varepsilon_k=0}^{1} (-1)^{\varepsilon_1 + \cdots + \varepsilon_k} F(c_1(\varepsilon_1), \ldots, c_k(\varepsilon_k))$$

where $c_j(0) = b_j$ and $c_j(1) = a_j$. We say that F has *positive increments* if

(1.33.3) $$\Delta_a^b F \geq 0 \qquad \forall a \leq b$$

Setting $\mathcal{F} = \{]a, b] \mid a \leq b \}$ and $\mu(]a, b]) = \Delta_a^b F$, then as above we get:

(1.33.4) If $F : \mathbf{R}^k \to \mathbf{R}$ is right continuous and has positive increments, then there exists a unique measure, denoted λ_F and called *the Lebesgue-Stieltjes measure induced by* F , on $\mathcal{B}(\mathbf{R}^k)$ satisfying

$$\lambda_F(]a, b]) = \Delta_a^b F \qquad \forall a \leq b$$

Setting $F(x_1, \ldots, x_k) = x_1 \cdot x_2 \cdot \cdots \cdot x_k$ we find:

(1.33.5) There exists a unique measure, denoted λ_k and called *the Lebesgue measure*, on $\mathcal{B}(\mathbf{R}^k)$ satisfying

$$\lambda_k(]a, b]) = (b_1 - a_1)(b_2 - a_2) \cdots (b_k - a_k)$$

whenever $a = (a_1, \ldots, a_k) \leq (b_1, \ldots, b_k) = b$

Note that $\lambda_k(F)$ equals the k -dimensional volume of B whenever B is a box, and so $\lambda_k(A)$ may be interpreted as *the* k *-dimensional volume* of A whenever A is a Borel subset of $\mathbf{R}^k$.

1.34 Measures on Metric Spaces

Let Ω *be a metric space and let* μ *be a finite Borel measure on* Ω . *Let*

$$\mathcal{F} = \{ F \subseteq \Omega \mid F \text{ is closed} \} \quad , \quad \mathcal{K} = \{ K \subseteq \Omega \mid K \text{ is compact} \}$$

$$\mathcal{C} = \{ C \subseteq \Omega \mid C \text{ is closed and precompact} \}$$

[see (P.5)]. Then we have

$$(1.34.1) \qquad \mu(B) = \sup\{\mu(F) \mid F \in \mathcal{F}\ ,\ F \subseteq B\} \qquad \forall B \in \mathcal{B}(\Omega)$$

That is, $\mathcal{F}$ *approximates* μ *on* $\mathcal{B}(\Omega)$ *. Moreover, if there exists a separable set* $\Omega_0 \in \mathcal{B}(\Omega)$ *such that* $\mu(\Omega \setminus \Omega_0) = 0$ *, then we have*

$$(1.34.2) \qquad \mu(B) = \sup\{\mu(C) \mid C \in \mathcal{C}\ ,\ C \subseteq B\} \qquad \forall B \in \mathcal{B}(\Omega)$$

That is, $\mathcal{C}$ *approximates* μ *on* $\mathcal{B}(\Omega)$ *. Finally, if for all* $\varepsilon > 0$ *there exists a complete separable set* $\Omega_0 \subseteq \Omega$ *such that* $\mu(\Omega \setminus \Omega_0) < \varepsilon$ *, then we have*

$$(1.34.3) \qquad \mu(B) = \sup\{\mu(K) \mid K \in \mathcal{K}\ ,\ K \subseteq B\} \qquad \forall B \in \mathcal{B}(\Omega)$$

Remark: If $\mu(\Omega \setminus \Omega_0) = 0$ for some separable Borel set Ω_0 , then we say that μ is *τ-smooth*; and if μ satisfies (1.34.3) then we say that μ is a *Radon measure*. It can be shown that it does not lead to any contradictions in mathematics if we assume that all finite Borel measures (on metric spaces!) are τ-smooth, nor does it lead to any contradiction in mathematics to assume that there exists a finite non-τ-smooth measure; that is, in some models of mathematics we have that all finite Borel measures on a metric space are τ-smooth, and in other models there exist non-τ-smooth finite Borel measures.

Proof: (1.34.1): Let $\mathcal{A}$ denote the paving of all sets $B \in \mathcal{B}(\Omega)$ such that B and its complement B^c satisfy (1.34.1). Then, clearly, $\mathcal{A}$ is (C)-stable. Let $B_1, B_2, \cdots \in \mathcal{A}$, let $B = \cup_{j=1}^{\infty} B_n$ and let $\varepsilon > 0$ be given. By assumption, there exist closed sets F_n and H_n such that $F_n \subseteq B_n$, $H_n \subseteq B_n^c$ and

$$\mu(B_n \setminus F_n) < \varepsilon 2^{-n} \quad , \quad \mu(B_n^c \setminus H_n) < \varepsilon 2^{-n}$$

Since $B \setminus \cup_{j=1}^{k} B_j \downarrow \emptyset$, there exists $k \geq 1$ such that

$$\mu\Big(B \setminus \bigcup_{j=1}^{k} B_j\Big) < \varepsilon$$

Set $F = F_1 \cup \cdots \cup F_k$ and $H = \cap_1^{\infty} H_n$, then F and H are closed and $F \subseteq B$ and $H \subseteq B^c$. Since

$$B \setminus F \subseteq \Big(B \setminus \bigcup_{j=1}^{k} B_j\Big) \cup \bigcup_{j=1}^{k} (B_j \setminus F_j) \quad , \quad B^c \setminus H \subseteq \bigcup_{j=1}^{\infty} (B_j^c \setminus H_j)$$

we have

$$\mu(B \setminus F) \leq \varepsilon + \sum_{j=1}^{k} \varepsilon\, 2^{-j} \leq 2\,\varepsilon \quad , \quad \mu(B^c \setminus H) \leq \sum_{j=1}^{\infty} \varepsilon\, 2^{-j} \leq \varepsilon$$

Hence, $B \in \mathcal{A}$ and thus, $\mathcal{A}$ is a σ-algebra. Let $F \in \mathcal{F}$ and let $\varphi(\omega) = d(\omega, F)$ be the distance from ω to F [see (P.5)]. Then by (P.5.20) and (P.5.22) we have that φ is continuous and $F = \{\varphi = 0\}$. Let $F_n = \{\varphi \geq 2^{-n}\}$, then F_n is closed and $F_n \uparrow F^c$. Hence, $\mu(F_n) \to \mu(F^c)$ and so $F \in \mathcal{A}$. Thus, $\mathcal{A}$ is a σ-algebra containing $\mathcal{F}$, but $\mathcal{B}(\Omega)$ is the smallest σ-algebra containing $\mathcal{F}$; so $\mathcal{B}(\Omega) \subseteq \mathcal{A}$. Thus, (1.34.1) holds for all $B \in \mathcal{B}(\Omega)$.

(1.34.2): Suppose that $\mu(\Omega \setminus \Omega_0) = 0$ where Ω_0 is separable. Let $\mathcal{D} = \{\omega_1, \omega_2, \cdots\}$ be a countable dense subset of Ω_0 . Let $p \in \mathbf{N}$ be given; then $\Omega_p \subseteq \Omega \setminus \Omega_0$ and $\Omega_{np} \downarrow \Omega_p$ as $n \uparrow \infty$, where

$$\Omega_{np} = \Omega \setminus \bigcup_{j=1}^{n} b\big[\omega_j, 2^{-p}\big] \quad \text{and} \quad \Omega_p = \Omega \setminus \bigcup_{j=1}^{\infty} b\big[\omega_j, 2^{-p}\big]$$

Hence, $\mu(\Omega_p) = 0$. Let $p \geq 1$ and $\varepsilon > 0$ be given. Then by (1.4.3) there exists an integer $n(p) \geq 1$ with $\mu(\Omega_{n(p),p}) < \varepsilon\, 2^{-p}$. Let $B \in \mathcal{B}(\Omega)$ be given, by (1.34.1) there exists a closed set $F \subseteq B$ such that $\mu(B \setminus F) \leq \varepsilon$. Set

$$C = F \cap \bigcap_{p=1}^{\infty} \bigcup_{j=1}^{n(p)} b[\omega_j, 2^{-p}]$$

Then C is closed because F and $b[\omega_j, 2^{-p}]$ are closed and since C is covered by $n(p)$ balls of radius 2^{-p} , we see that C is precompact. Moreover,

$$B \setminus C \subseteq (B \setminus F) \cup \bigcup_{p=1}^{\infty} \left(\Omega \setminus \bigcup_{j=1}^{n(p)} b[\omega_j, 2^{-p}] \right)$$

and so we get

$$\mu(B \setminus C) \leq \varepsilon + \sum_{p=1}^{\infty} \varepsilon 2^{-p} = 2\varepsilon$$

Hence, (1.34.2) holds for all $B \in \mathcal{B}(\Omega)$.

(1.34.3): Let $\varepsilon > 0$ and $B \in \mathcal{B}(\Omega)$ be given and suppose that $\mu(\Omega \setminus \Omega_0) < \varepsilon$ for some separable complete set Ω_0 . By (1.34.2) there exists a closed precompact set $C \subseteq B \cap \Omega_0$ such that $\mu((B \cap \Omega_0) \setminus C) < \varepsilon$. Since Ω_0 is complete and C is a closed subset of Ω_0 , we have by (P.5.7) that C is compact, and $C \subseteq B$. Moreover, since $B \setminus C \subseteq (\Omega \setminus \Omega_0) \cup ((B \cap \Omega_0) \setminus C)$, we see that $\mu(B \setminus C) < 2\varepsilon$, and so the theorem is proved. $\square$

1.35 Metric Outer Measures

Let (Ω, d) be a metric space; if A and B are subsets of S, then we define *the distance between* A and B, denoted $d(A,B)$, and the *diameter* of A, denoted $\operatorname{diam}(A)$, as follows:

$$(1.35.1) \qquad d(A,B) := \inf_{(x,y)\in A\times B} d(x,y) \quad , \quad \operatorname{diam}(A) := \sup_{x,y\in A} d(x,y)$$

with the conventions: $\inf\emptyset = \infty$ and $\sup\emptyset = 0$. We define a *metric outer measure* μ on Ω to be a set function $\mu : 2^\Omega \to [0,\infty]$ satisfying

$$(1.35.2) \qquad \mu \text{ is } \sigma\text{-subadditive on } 2^\Omega$$

$$(1.35.3) \qquad \mu(A\cup B) = \mu(A) + \mu(B) \quad \text{if} \quad d(A,B) > 0$$

Let $\mathcal{F}$ be a paving on Ω and set $\mathcal{F}_\varepsilon = \{F \in \mathcal{F} \mid \operatorname{diam}(F) \le \varepsilon\}$ for all $\varepsilon > 0$. If $\lambda : \mathcal{F} \to \overline{\mathbf{R}}_+$ is a set function, then we define

$$(1.35.4) \qquad \lambda^\varepsilon(A) = \inf\left\{ \sum_{j=1}^\infty \lambda(F_j) \;\middle|\; F_1, F_2 \ldots \in \mathcal{F}_\varepsilon \,,\, A \subseteq \bigcup_{j=1}^\infty F_j \right\} \qquad \forall \varepsilon > 0$$

with the convention: $\inf\emptyset = \infty$; that is, $\lambda^\varepsilon(A) = \infty$ if A cannot be covered by a sequence of $\mathcal{F}$-sets with diameter at most ε. Then $\lambda^\varepsilon(A)$ is decreasing in ε for every $A \subseteq \Omega$ fixed. Thus, we may define *the metric outer measure generated by* λ, denoted λ^0, as follows:

$$(1.35.5) \qquad \lambda^0(A) = \lim_{\varepsilon\downarrow 0} \lambda^\varepsilon(A) = \sup_{\varepsilon>0} \lambda^\varepsilon(A) \qquad \forall A \subseteq \Omega$$

Then λ^0 is a set function on 2^Ω, and below we shall see that λ^0 is a metric outer measure such that $\mathcal{B}(\Omega) \subseteq \mathcal{M}(\mu)$, and thus, by Carathéodory's theorem (1.23) we have that the restriction of λ^0 to the Borel σ-algebra $\mathcal{B}(\Omega)$ is a Borel measure on Ω. This method gives a very powerful tool for constructing Borel measures on metric spaces and in Chapter 8 of Volume II we shall see how this method can be used to construct the q-dimensional volume of a subset of $\mathbf{R}^n$ whenever q is an arbitrary non-negative real number.

1.36 Construction of Metric Outer Measures

Let (Ω, d) be a metric space and let μ be a set function on 2^Ω. Then μ is a metric outer measure if and only if μ is σ-subadditive on 2^Ω and $\mathcal{B}(\Omega) \subseteq \mathcal{M}(\mu)$.

Moreover, if $\mathcal{F}$ is a paving on Ω and λ^ε for $\varepsilon \geq 0$ is defined by (1.35.4+5), where $\lambda : \mathcal{F} \to \overline{\mathbf{R}}_+$ is a given set function, then λ^0 is a metric outer measure on Ω .

Proof: Suppose that μ is a metric outer measure, and let F be a closed subset of Ω . I shall then show that $F \in \mathcal{M}(\mu)$. Since μ is subadditive, this means that we have to show that

$$\text{(i)} \qquad \mu(A) \geq \mu(A \cap F) + \mu(A \setminus F) \qquad \forall A \subseteq \Omega$$

So let $A \subseteq \Omega$. If $\mu(A) = \infty$, then (i) is trivially satisfied. Hence, we may assume that $\mu(A) < \infty$. Let $f(\omega) = d(\omega, F)$; then $F = \{\omega \mid f(\omega) = 0\}$ by (P.5.22) because F is closed. Hence, if $B_n = \{\omega \in A \mid f(\omega) \geq 2^{-n}\}$ and $C_n = B_{n+1} \setminus B_n$, then $B_n \uparrow A \setminus F$, and by the triangle inequality we have that

$$d(C_n, C_{n+2}) \geq 2^{-n-1} - 2^{-n-2} = 2^{-n-2} > 0$$

for all $n \geq 1$. Thus, by (1.35.3) we have that

$$\sum_{j=1}^{n} \mu(C_{2j}) = \mu\left(\bigcup_{j=1}^{n} C_{2j}\right) \leq \mu(A) < \infty$$

$$\sum_{j=1}^{n} \mu(C_{2j-1}) = \mu\left(\bigcup_{j=1}^{n} C_{2j-1}\right) \leq \mu(A) < \infty$$

Letting $n \to \infty$, we see that $\sum_{j=1}^{\infty} \mu(C_j) \leq 2\mu(A) < \infty$. Note that $A \setminus F = B_n \cup \cup_{j=n}^{\infty} C_j$; thus, we find

$$\mu(B_n) \geq \mu(A \setminus F) - \sum_{j=n}^{\infty} \mu(C_j)$$

Moreover, since $d(A \cap F, B_n) \geq 2^{-n}$, we get

$$\begin{aligned} \mu(A) &\geq \mu((A \cap F) \cup B_n) = \mu(A \cap F) + \mu(B_n) \\ &\geq \mu(A \cap F) + \mu(A \setminus F) - \sum_{j=n}^{\infty} \mu(C_j) \end{aligned}$$

for all $n \geq 1$. Since $\sum \mu(C_j) < \infty$ we have that the last term converges to 0. Hence, letting $n \to \infty$, we see that (i) holds. This shows that every closed set belongs to $\mathcal{M}(\mu)$, and since $\mathcal{M}(\mu)$ is a σ-algebra we have that $\mathcal{B}(\Omega) \subseteq \mathcal{M}(\mu)$. Thus, the "only if" in the first part of the theorem is proved.

Conversely, if μ is σ-subadditive and $\mathcal{B}(\Omega) \subseteq \mathcal{M}(\mu)$, then by (1.23.1) it follows easily that μ satisfies (1.35.1). Thus, μ is a metric outer measure, and the "if" in the first part of the theorem is proved.

So suppose that λ^ε is given by (1.35.4+5) where λ is a set function on $\mathcal{F}$. Let $\{A_n\}$ be a sequence of subsets of Ω , and let $A = \cup A_n$ I shall then show that

$$\text{(ii)} \qquad \lambda^0(A) \leq \sum_{j=1}^{\infty} \lambda^0(A_n)$$

If $\sum \lambda^0(A_n) = \infty$, then (ii) holds trivially, so suppose that $\sum \lambda^0(A_n) < \infty$, and let $\varepsilon > 0$ be given. Then by definition of λ^ε there exist $F_{jn} \in \mathcal{F}_\varepsilon$ such that $A_n \subseteq \cup_j F_{jn}$ and

$$\lambda^\varepsilon(A_n) \leq \sum_{j=1}^{\infty} \lambda(F_{jn}) \leq \varepsilon 2^{-n} + \lambda^\varepsilon(A_n)$$

Since $A \subseteq \cup_{j,n} F_{jn}$, we have

$$\begin{aligned} \lambda^\varepsilon(A) &\leq \sum_{n=1}^{\infty}\sum_{j=1}^{\infty} \lambda(F_{nj}) \leq \varepsilon + \sum_{n=1}^{\infty} \lambda^\varepsilon(A_n) \\ &\leq \varepsilon + \sum_{n=1}^{\infty} \lambda^0(A_n) \end{aligned}$$

So letting $\varepsilon \to 0$, we see that (ii) holds. Next I shall show:

$$\text{(iii)} \qquad \lambda^0(A \cup B) = \lambda^0(A) + \lambda^0(B) \quad \text{if} \quad d(A,B) > 0$$

Let $\varepsilon > 0$ and suppose that $d(A,B) > \varepsilon$ and let $\{F_n\}$ be a sequence in $\mathcal{F}_\varepsilon$ such that $A \cup B \subseteq \cup_1^\infty F_n$. Since $d(A,B) > \varepsilon \geq \operatorname{diam}(F_n)$, we see that F_n intersects at most one of the sets A and B . Hence, if

$$\alpha = \{n \in \mathbf{N} \mid A \cap F_n \neq \emptyset\} \quad , \quad \beta = \{n \in \mathbf{N} \mid B \cap F_n \neq \emptyset\}$$

then α and β are disjoint subsets of $\mathbf{N}$ such that

$$A \subseteq \bigcup_{n\in\alpha} F_n \quad , \quad B \subseteq \bigcup_{n\in\beta} F_n$$

Hence, we have

$$\lambda^\varepsilon(A) + \lambda^\varepsilon(B) \leq \sum_{n\in\alpha} \lambda(F_n) + \sum_{n\in\beta} \lambda(F_n) \leq \sum_{n=1}^{\infty} \lambda(F_n)$$

and so taking infimum over all sequences $\{F_n\} \subseteq \mathcal{F}_\varepsilon$ with $A \cup B \subseteq \cup_1^\infty F_n$ we see that $\lambda^\varepsilon(A) + \lambda^\varepsilon(B) \leq \lambda^\varepsilon(A \cup B)$. Letting $\varepsilon \to 0$, we conclude that $\lambda^0(A) + \lambda^0(B) \leq \lambda^0(A \cup B)$, and since the converse inequality holds by (ii), we see that (iii) is proved. Observe that (ii) and (iii) show that λ^0 is a metric outer measure, and so the second part of the theorem is proved. □

1.37 The Uniqueness Theorem II

Let (S,d) be a metric space, and let μ and ν be two measures on $(S,\mathcal{B}(S))$ such that $\mu(S)=\nu(S)<\infty$. Suppose that $\mathcal{H}\subseteq\mathcal{B}(S)$ is a $(\cap f)$-stable paving satisfying

(1.37.1) $\mu(H)=\nu(H) \qquad \forall H\in\mathcal{H}$

Then $\mu=\nu$ if just one of the following three conditions are satisfied:

(1.37.2) *$\mathcal{H}$ contains all closed subsets of S*

(1.37.3) *$\mathcal{H}$ contains all open subsets of S*

(1.37.4) *μ is a Radon measure, ν is τ- smooth, and for all $x\neq y$ there exists disjoint open sets $U,V\in\sigma(\mathcal{H})$ such that $x\in U$ and $y\in V$*

Proof: By the uniqueness theorem (1.7) with $F_n=S$ for all n, we have that $\mu(B)=\nu(B)$ for all $B\in\sigma(\mathcal{H})$. If (1.37.2) or (1.37.3) holds, then $\sigma(\mathcal{H})=\mathcal{B}(S)$, and so $\mu=\nu$.

So suppose that (1.37.4) holds and let K be a compact set. Then for every $x\in K$ and every $y\in S\setminus K$, there exist disjoint open sets $U(x,y)$ and $V(x,y)$ belonging to $\sigma(\mathcal{H})$ such that $x\in U(x,y)$ and $y\in V(x,y)$. Let $y\in S\setminus K$ be fixed for a moment, then $\{U(x,y)\mid x\in K\}$ is an open cover of K so by compactness of K there exists a finite set $\tau(y)$ such that

$$K\subseteq\bigcup_{x\in\tau(y)}U(x,y) \qquad \forall y\in S\setminus K$$

Let us define

$$H(y)=\bigcup_{x\in\tau(y)}U(x,y) \quad , \quad G(y)=\bigcap_{x\in\tau(y)}V(x,y)$$

for all $y\in S\setminus K$. Since $\tau(y)$ is finite, we have that $H(y)$ and $G(y)$ are open and belong to $\sigma(\mathcal{H})$. Evidently, we have that $K\subseteq H(y)$ and $y\in G(y)$ for all $y\in S\setminus K$. Since $U(x,y)\cap V(x,y)=\emptyset$ for all $x\in\tau(y)$ we see that $H(y)$ and $G(y)$ are disjoint sets. By assumption we know that ν is τ-smooth, so let S_0 be a separable Borel subset of S such that $\nu(S\setminus S_0)=0$. Since $\{G(y)\mid y\in S\setminus K\}$ is an open cover of $S_0\setminus K$, then by (P.5.3) there exists $\{y_n\mid n\geq 1\}\subseteq S\setminus K$ such that $S_0\setminus K\subseteq\cup_{n=1}^{\infty}G(y_n)$. Let

$$H_0=\bigcap_{n=1}^{\infty}H(y_n) \quad , \quad G_0=\bigcup_{n=1}^{\infty}G(y_n)$$

Then $K \subseteq H_0$ and $S_0 \setminus K \subseteq G_0$. Moreover, $H_0 \cap G_0 = \emptyset$ because $H(y_n) \cap G(y_n) = \emptyset$ for all $n \geq 1$. Hence, we have

$$S_0 \cap K \subseteq S_0 \cap H_0 \subseteq S_0 \setminus G_0 \subseteq S_0 \setminus (S_0 \setminus K) = S_0 \cap K$$

and so $S_0 \cap K = S_0 \cap H_0$. Thus, we find

$$\nu(K) = \nu(S_0 \cap K) = \nu(S_0 \cap H_0) = \nu(H_0) = \mu(H_0) \geq \mu(K)$$

because $\nu(S \setminus S_0) = 0$ and $H_0 \in \sigma(\mathcal{H})$. This shows that $\mu(K) \leq \nu(K)$ for all compact sets $K \subseteq S$, and since μ is a Radon measure, we see that

$$\begin{aligned} \mu(B) &= \sup\{\mu(K) \mid K \subseteq B \ , \ K \ \text{ is compact } \} \\ &\leq \sup\{\nu(K) \mid K \subseteq B \ , \ K \ \text{ is compact } \} \leq \nu(B) \end{aligned}$$

for all $B \in \mathcal{B}(S)$. But if $B \in \mathcal{B}(S)$, then $B^c \in \mathcal{B}(S)$, and so we have

$$\mu(B) = \mu(S) - \mu(B^c) = \nu(S) - \mu(B^c) \geq \nu(S) - \nu(B^c) = \nu(B)$$

because $\mu(S) = \nu(S)$. Thus, we conclude that $\mu = \nu$. □

1.38 Measurable Functions

Recall that a function f is continuous if and only if $f^{-1}(G)$ is open whenever G is an open set. We shall now define a measurable function in much the same way. A *measurable space* $(\Omega, \mathcal{F})$ is a set Ω with a given σ-algebra $\mathcal{F}$. Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be two measurable spaces, and let $f : \Omega \to \Lambda$ be a function. Then we say that f is *measurable* if

(1.38.1) $\quad f^{-1}(G) \in \mathcal{F} \qquad \forall G \in \mathcal{G}$

If it is not clear from the context which σ-algebras on Ω and/or Λ we are considering, we can specify this by saying that f is $(\mathcal{F}, \mathcal{G})$-*measurable*, or $\mathcal{F}$-*measurable* or $\mathcal{G}$-*measurable*. Let μ be a measure on $(\Omega, \mathcal{F})$ and let $\mathcal{M}(\mu)$ be the σ-algebra of all μ-measurable sets [see (1.28)], then a $(\mathcal{M}(\mu), \mathcal{G})$-measurable function $f : \Omega \to \Lambda$ is said to be μ-*measurable.*

Let $f : \Omega \to \Lambda$ be a function and $\mathcal{H}$ be a paving on Λ if we define the paving $f^{-1}(\mathcal{H})$ on Ω as

(1.38.2) $\quad f^{-1}(\mathcal{H}) := \{f^{-1}(H) \mid H \in \mathcal{H}\}$

then we see that

(1.38.3) f is $(\mathcal{F},\mathcal{G})$-measurable if and only if $f^{-1}(\mathcal{G}) \subseteq \mathcal{F}$

Let $(\Omega,\mathcal{F})$, $(\Lambda,\mathcal{G})$ and $(\Gamma,\mathcal{H})$ be measurable spaces, and let $f : \Omega \to \Lambda$ and $g : \Lambda \to \Gamma$ be measurable. Then $h = g \circ f$ maps Ω into Γ , and since $h^{-1}(H) = f^{-1}(g^{-1}(H))$, we see that h is measurable; that is, in analogy with continuous functions, we have

(1.38.4) The composition $h = g \circ f$ of two measurable functions f and g is itself measurable

Let $f : \Omega \to \Lambda$ be a function and let $\mathcal{H}$ be a paving on Λ , then I claim that we have

(1.38.5) $\sigma(f^{-1}(\mathcal{H})) = f^{-1}(\sigma(\mathcal{H}))$

To see this, let $\mathcal{F}_0 = f^{-1}(\mathcal{H})$ and $\mathcal{F} = f^{-1}(\sigma(\mathcal{H}))$. Then $\mathcal{F}_0 \subseteq \mathcal{F}$ since $\mathcal{H} \subseteq \sigma(\mathcal{H})$; moreover, as the co-image respects all set operations [see (P.3.1)–(P.3.3)], we have that $\mathcal{F}$ is a σ-algebra. Hence $\mathcal{F} \supseteq \sigma(\mathcal{F}_0)$. To prove the converse inclusion, consider the paving $\mathcal{H}_1 = \{B \subseteq \Lambda \mid f^{-1}(B) \in \sigma(\mathcal{F}_0)\}$. Again from (P.3.1)–(P.3.3) it follows easily that $\mathcal{H}_1$ is a σ-algebra and $\mathcal{H}_1 \supseteq \mathcal{H}$ because $f^{-1}(\mathcal{H}) = \mathcal{F}_0 \subseteq \sigma(\mathcal{F}_0)$. Hence, $\mathcal{H}_1 \supseteq \sigma(\mathcal{H})$ and so $\mathcal{F} \subseteq \sigma(\mathcal{F}_0)$; that is, $\mathcal{F} = \sigma(\mathcal{F}_0)$ and so (1.38.5) follows. Moreover, by (1.38.3) and (1.38.5) we get

(1.38.6) Let $(\Omega,\mathcal{F})$ and $(\Lambda,\mathcal{G})$ be two measurable spaces and let $\mathcal{G}_0$ be a paving on Λ such that $\sigma(\mathcal{G}_0) = \mathcal{G}$. Then $f : \Omega \to \Lambda$ is measurable if and only if $f^{-1}(G_0) \in \mathcal{F}$ for all $G_0 \in \mathcal{G}_0$

This result is in spite of its simplicity extremely useful. It means that we do not have to verify (1.38.1) for all $G \in \mathcal{G}$, it suffices to verify (1.38.1) for all G in a paving which generates $\mathcal{G}$. This is important because σ-algebras usually are not given explicitly but as the σ-algebras generated by some explicitly given paving.

1.39 Borel Functions

If Ω is a metric space, then we have the Borel σ-algebra $\mathcal{B}(\Omega)$ on Ω , and so $(\Omega,\mathcal{B}(\Omega))$ becomes a measurable space. ***And from now on we shall always consider a metric space as a measurable space endowed with its Borel σ-algebra*** $\mathcal{B}(\Omega)$. If Ω and Λ are metric spaces, then a $(\mathcal{B}(\Omega),\mathcal{B}(\Lambda))$-measurable function $f : \Omega \to \Lambda$ is called a *Borel function*, and f is said to be *Borel measurable*. For instance, a Borel function $f : \mathbf{R} \to \mathbf{R}$ is a function from $\mathbf{R}$ into $\mathbf{R}$,

which is measurable with respect to the Borel σ-algebra $\mathcal{B}(\mathbf{R})$ on $\mathbf{R}$. Since the Borel σ-algebra is the σ-algebra generated by the open sets, by (1.38.6) we see that every continuous function is a Borel function.

1.40 Real-Valued Functions

Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \mathbf{R}$ be a real-valued function. Since $\mathcal{B}(\mathbf{R})$ is generated by the intervals of the form $]-\infty, a]$, we have by (1.38.6) and (1.10.1)

(1.40.1) f is measurable if and only if $\{f \le a\} \in \mathcal{F}$ for all $a \in \mathbf{R}$

[recall that $\{f \le a\} = f^{-1}(]-\infty, a]) = \{\omega \mid f(\omega) \le a\}$ in our shorthand notation introduced in (P.3)]. By (1.38.4) we have

(1.40.2) If $f : \Omega \to \mathbf{R}$ is measurable and $g : \mathbf{R} \to \mathbf{R}$ is a Borel function, then $g \circ f$ is measurable

So if f is measurable from Ω into $\mathbf{R}$, then so are $f(\omega)^2$, $\sin f(\omega)$, $e^{f(\omega)}$, etc. I claim that we have

(1.40.3) If $f_1, f_2, \ldots$ are measurable functions, then

$$\sup_{n\ge1} f_n \quad , \quad \inf_{n\ge1} f_n \quad , \quad \limsup_{n\to\infty} f_n \quad , \quad \liminf_{n\to\infty} f_n$$

are all measurable functions

where $\sup_n f_n$ is the function: $g(\omega) = \sup_n f_n(\omega)$ and similarly, for the others. To see this, let $g = \sup_n f_n$, then we have

$$\{g \le a\} = \bigcap_{n=1}^{\infty} \{f_n \le a\} \in \mathcal{F} \qquad \forall a \in \mathbf{R}$$

because $\mathcal{F}$ is $(\cap c)$-stable and $\{f_n \le a\} \in \mathcal{F}$. Hence, g is measurable by (1.40.1). In a similar way one shows easily that $\inf_n f_n$ is measurable, and since

$$\limsup_{n\to\infty} f_n(\omega) = \inf_{n\ge1} \left\{ \sup_{j\ge n} f_j(\omega) \right\} \quad , \quad \liminf_{n\to\infty} f_n(\omega) = \sup_{n\ge1} \left\{ \inf_{j\ge n} f_j(\omega) \right\}$$

we see that $\limsup f_n$ and $\liminf f_n$ are measurable. In particular, we have

(1.40.4) Suppose that $f_n : \Omega \to \mathbf{R}$ is measurable for all $n \ge 1$ and that $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ for all $\omega \in \Omega$, then f is measurable

1.41 The Measurable Limit Theorem

Let (M,d) be a metric space and let $f_1, f_2, \cdots$ be a sequence of measurable functions from a measurable space $(\Omega, \mathcal{F})$ into M. If the limit

$$f(\omega) = \lim_{n\to\infty} f_n(\omega)$$

exists for all $\omega \in \Omega$, then the function $f : \Omega \to M$ is measurable.

Remark: Suppose that (M,d) is a separable complete metric space and that $f_1, f_2, \ldots$ are measurable function from $(\Omega, \mathcal{F})$ into M. Then it is fairly easily to show that [see Exc.1.74]

(1.41.1) $$L = \{\omega \in \Omega \mid \{f_n(\omega)\} \text{ converges}\} \in \mathcal{F}$$

Proof: Let F be a closed set in M and let $\varphi(x)$ be the distance from the point x to F, that is,

$$\varphi(x) = \inf_{y\in F} d(x,y) = d(x,F)$$

Then by (P.5.20) we have that φ is continuous from M into $\mathbf{R}$. Hence if $g_n = \varphi \circ f_n$ and $g = \varphi \circ f$, then g_n is measurable from Ω into $\mathbf{R}$ and

$$g(\omega) = \varphi(f(\omega)) = \lim_{n\to\infty} \varphi(f_n(\omega)) = \lim_{n\to\infty} g_n(\omega) \quad \forall \omega \in \Omega$$

Hence, by (1.40.4) we have that g is measurable and so $g^{-1}(0) \in \mathcal{F}$. By (P.5.21) we have $F = \varphi^{-1}(0)$ and, thus, we conclude

$$f^{-1}(F) = f^{-1}(\varphi^{-1}(0)) = g^{-1}(0) \in \mathcal{F}$$

Hence, f is measurable by (1.38.6). □

1.42 The σ-Algebra Generated by Functions

Let Ω be a set and let $(\Lambda, \mathcal{G})$ be a measurable space. Suppose that f is a function from Ω into Λ if $\mathcal{F}$ is a σ-algebra on Ω such that f is measurable. Then by definition we have that $\mathcal{F}$ contains $f^{-1}(\mathcal{G})$. Moreover, $f^{-1}(\mathcal{G})$ is a σ-algebra on Ω by (1.38.5). Hence $f^{-1}(\mathcal{G})$ is the smallest σ-algebra on Ω making f measurable, and we set

(1.42.1) $$\sigma(f) := f^{-1}(\mathcal{G}) = \{f^{-1}(G) \mid G \in \mathcal{G}\}$$

Suppose that $(\Lambda_j, \mathcal{G}_j)$ for $j = 1, \cdots, n$ are measurable spaces and that $f_j : \Omega \to \Lambda_j$ for $j = 1, \cdots, n$ are functions; then we set

(1.42.2) $$\sigma(f_1, \cdots, f_n) := \sigma\{f_j^{-1}(G_j) \mid j \in \{1, \cdots, n\}\ ,\ G_j \in \mathcal{G}_j\}$$

As above, we see that $\sigma(f_1, \cdots, f_n)$ is the smallest σ-algebra on Ω making all the functions $f_1, \cdots, f_n$ measurable.

More generally, let $(\Lambda_j, \mathcal{G}_j)$ be a measurable space and let f_j be a function: $\Omega \to \Lambda_j$ for all j in some set J. If $\mathcal{W} = \{f_j \mid j \in J\}$, then we set

(1.42.3) $$\sigma(\mathcal{W}) = \sigma(f_j \mid j \in J) := \sigma\{f_j^{-1}(G_j) \mid j \in J\ ,\ G_j \in \mathcal{G}_j\}$$

As above, we see that $\sigma(f_j \mid j \in J)$ is the smallest σ-algebra on Ω making all the functions f_j for $j \in J$ measurable. Suppose that $(\Gamma, \mathcal{H})$ is a measurable space and $g : \Gamma \to \Omega$ is a $(\mathcal{H}, \mathcal{F})$-measurable function, where $\mathcal{F} = \sigma(f_j \mid j \in J)$; then $f_j \circ g$ is measurable for all $j \in J$ by (1.38.4). Conversely, if $f_j \circ g$ is measurable for all $j \in J$, then

$$g^{-1}(f_j^{-1}(G_j)) = (f_j \circ g)^{-1}(G_j) \in \mathcal{H}$$

for all $j \in J$ and all $G_j \in \mathcal{G}_j$. Hence, by (1.38.6) we have

(1.42.4) A function $g : (\Gamma, \mathcal{H}) \to (\Omega, \mathcal{F})$, where $\mathcal{F} = \sigma(f_j \mid j \in J)$, is measurable if and only if $f_j \circ g$ is measurable for all $j \in J$

1.43 The Product σ-Algebra

Let $(\Omega_j, \mathcal{F}_j)$ be a measurable space for $j = 1, \cdots, n$, and consider *the product set*

$$\Omega_1 \times \cdots \times \Omega_n := \{(\omega_1, \cdots, \omega_n) \mid \omega_1 \in \Omega_1, \ldots, \omega_n \in \Omega_n\}$$

and *the projections* $\wp_j : \Omega_1 \times \cdots \times \Omega_n \to \Omega_j$:

$$\wp_j(\omega_1, \ldots, \omega_n) := \omega_j \qquad \forall 1 \le j \le n \quad \forall (\omega_1, \ldots, \omega_n)$$

We can then define a σ-algebra, called *the product σ-algebra* on $\Omega_1 \times \cdots \times \Omega_n$ as follows:

(1.43.1) $$\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n := \sigma(\wp_1, \ldots, \wp_n)$$

Observe that

$$\wp_j^{-1}(A_j) = \Omega_1 \times \cdots \times \Omega_{j-1} \times A_j \times \cdots \times \Omega_n$$

$$\bigcap_{j=1}^{n} \wp_j^{-1}(A_j) = A_1 \times \cdots \times A_n$$

whenever $A_1 \subseteq \Omega_1, \ldots, A_n \subseteq \Omega_n$. Hence we have

(1.43.2) $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n = \sigma\{F_1 \times \cdots \times F_n \mid F_1 \in \mathcal{F}_1, \ldots, F_n \in \mathcal{F}_n\}$

that is, $\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$ is the smallest σ-algebra on the product set which contains all "cubes" with measurable "sides". Let f be a function from a measurable space $(\Omega, \mathcal{F})$ into $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n)$, and let $f_1, \ldots, f_n$ be *the coordinate functions* of f ; that is,

$$f(\omega) = (f_1(\omega), \ldots, f_n(\omega)) \qquad \forall \omega \in \Omega$$

Then f_j is a function from Ω into Ω_j , and $f_j = \wp_j \circ f$. Hence, by (1.42.4) we have

(1.43.3) $f : (\Omega, \mathcal{F}) \to (\Omega_1 \times \cdots \times \Omega_n, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n)$ is measurable if and only if all its coordinate functions $f_1, \ldots, f_n$ are measurable

Suppose that (Ω_j, d_j) is a metric space for $j = 1, 2, \ldots, n$. Then we define *the product metric* d on the product space $\Omega_1 \times \cdots \times \Omega_n$ as follows:

$$d(\omega^*, \omega) = \sqrt{\sum_{j=1}^{n} d_j(\omega_j^*, \omega_j)^2}$$

whenever $\omega^* = (\omega_1^*, \ldots, \omega_n^*)$ and $\omega = (\omega_1, \ldots, \omega_n)$. Let $\omega = (\omega_1, \ldots, \omega_n)$ and $r > 0$ be given, then

$$b(\omega, r) \subseteq b(\omega_1, r) \times \cdots \times b(\omega_n, r) \subseteq b(\omega, r\sqrt{n})$$

Hence, if we set $r_k = n^{-k/2}$ for $k = 1, 2, \cdots$ and

$$\mathcal{H} = \{b(\omega_1, r) \times \cdots \times b(\omega_n, r) \mid r > 0 \ , \ \omega_1 \in \Omega_1, \ldots, \omega_n \in \Omega_n\}$$

then we see that (1.9.1) holds. So by (1.9) we find:

(1.43.4) If $\Omega_1, \ldots, \Omega_n$ are separable metric spaces, then $\mathcal{B}(\Omega_1 \times \cdots \times \Omega_n) = \mathcal{B}(\Omega_1) \otimes \cdots \otimes \mathcal{B}(\Omega_n)$

Warning: The statement in (1.43.4) is false if we drop the separability condition.

In particular, setting $\Omega_j = \mathbf{R}$ for $j = 1, \ldots, n$ we find

(1.43.5) $\mathcal{B}(\mathbf{R}^n) = \mathcal{B}(\mathbf{R}) \otimes \cdots \otimes \mathcal{B}(\mathbf{R})$

Hence, we have the following propositions

(1.43.6) If $f_1, \ldots, f_n$ are measurable functions from $(\Omega, \mathcal{F})$ in $\mathbf{R}$, then $f(\omega) = (f_1(\omega), \ldots, f_n(\omega))$ is measurable: $(\Omega, \mathcal{F}) \to \mathbf{R}^n$

(1.43.7) If $f_1, \ldots, f_n$ are measurable functions from $(\Omega, \mathcal{F})$ in $\mathbf{R}$ and φ is a Borel function from $\mathbf{R}^n$ into $\mathbf{R}^k$, then $\varphi(f_1(\omega), \ldots, f_n(\omega))$ is measurable: $(\Omega, \mathcal{F}) \to \mathbf{R}^k$

In particular, if f and g are measurable then so are $f+g$, $f \cdot g$, $\sin(f+g^2)$, etc.

1.44 Image Measures

A *measure space* $(\Omega, \mathcal{F}, \mu)$ is a set Ω, a σ-algebra $\mathcal{F}$ on Ω, and a measure μ on $(\Omega, \mathcal{F})$. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f be a measurable function from $(\Omega, \mathcal{F})$ into a measurable space $(\Lambda, \mathcal{G})$. Then we define *the image measure* of μ under f, denoted μ_f, as

(1.44.1) $\mu_f(G) := \mu(f^{-1}(G)) \qquad \forall G \in \mathcal{G}$

Note that $f^{-1}(G) \in \mathcal{F}$ for all $G \in \mathcal{G}$ by measurability of f. Hence, μ_f is a well-defined function from $\mathcal{G}$ into $\overline{\mathbf{R}}_+$. From formulas (P.3.1)–(P.3.3) it follows easily that

(1.44.2) μ_f is a measure on $(\Lambda, \mathcal{G})$

Let $g : (\Lambda, \mathcal{G}) \to (\Gamma, \mathcal{H})$ be another measurable function. Then $h = g \circ f$ is a measurable function from $(\Omega, \mathcal{F})$ into $(\Gamma, \mathcal{H})$ by (1.38.4), and

$$h^{-1}(H) = (g \circ f)^{-1}(H) = f^{-1}(g^{-1}(H))$$

so we have

$$\mu_h(H) = \mu(f^{-1}(g^{-1}(H)) = \mu_f(g^{-1}(H)) = (\mu_f)_g(H)$$

Hence, we have proved the following simple, but very useful, proposition:

(1.44.3) If $h = g \circ f$ then $\mu_h = (\mu_f)_g$

1.45 Simple and Discrete Functions

Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \mathbf{R}$ be a real valued measurable function. If f only assumes finitely respectively countably many different values we say f is a *simple* respectively a *discrete function.* Let f be a simple function and let $a_1, \ldots, a_n \in \mathbf{R}$ be the possible values of f. Set $A_j = \{f = a_j\}$ for $j = 1, \ldots, n$, then clearly we have

(1.45.1) $\quad A_1, \ldots, A_n$ are disjoint sets in $\mathcal{F}$ and $\Omega = A_1 \cup \cdots \cup A_n$

(1.45.2) $$f = \sum_{j=1}^{n} a_j \, 1_{A_j}$$

Conversely, if $A_1, \ldots, A_n$ satisfies (1.45.1) and $a_1, \ldots, a_n \in \mathbf{R}$, then (1.45.2) defines a simple function. Similarly, if f is a discrete function, and $a_1, a_2, \cdots$ are the possible values of f and $A_j = \{f = a_j\}$ for $j \in \mathbf{N}$, then as above we find

(1.45.3) $\quad A_1, A_2 \ldots$ are disjoint sets in $\mathcal{F}$ and $\Omega = \bigcup_{j=1}^{\infty} A_j$

(1.45.4) $$f = \sum_{j=1}^{\infty} a_j \, 1_{A_j}$$

Conversely, if $A_1, A_2, \cdots$ satisfies (1.45.3) and $a_1, a_2, \cdots \in \mathbf{R}$, then (1.45.4) defines a discrete function.

1.46 Approximation of Measurable Functions

Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : (\Omega, \mathcal{F}) \to \mathbf{R}$ be an arbitrary measurable function. Then there exists discrete functions $f_1, f_2, \ldots : \Omega \to \mathbf{R}$ and simple functions $g_1, g_2, \ldots : \Omega \to \mathbf{R}$ satisfying

(1.46.1) $$f_n(\omega) \le f(\omega) \le f_n(\omega) + 2^{-n} \qquad \forall n = 1, 2, \cdots$$

(1.46.2) $$f_1(\omega) \le f_2(\omega) \le \cdots \quad \textit{and} \quad f(\omega) = \lim_{n \to \infty} f_n(\omega) \quad \forall \omega \in \Omega$$

(1.46.3) $$f(\omega) = \lim_{n \to \infty} g_n(\omega) \qquad \forall \omega \in \Omega$$

(1.46.4) $$\textit{If } f \ge 0 \textit{ then } 0 \le g_1(\omega) \le g_2(\omega) \le \cdots \qquad \forall \omega \in \Omega$$

Thus, any real-valued measurable function f is a uniform limit of an increasing sequence of discrete functions and a pointwise limit of a sequence of simple functions. Moreover, any non-negative measurable function f is a pointwise limit of an increasing sequence of non-negative simple functions.

Proof: Let $int(x)$ denote *the integer part* of x for $x \in \mathbf{R}$; that is, $int(x)$ is the unique integer k satisfying $k \leq x < k+1$. Then we define

$$f_n(\omega) = 2^{-n}\, int(2^n f(\omega)) \quad , \quad g_n(\omega) = \begin{cases} n & \text{if } f_n(\omega) > n \\ f_n(\omega) & \text{if } |f_n(\omega)| \leq n \\ -n & \text{if } f_n(\omega) < -n \end{cases}$$

Observe that

$$f_n(\omega) = k\,2^{-n} \quad \text{if} \quad k\,2^{-n} \leq f(\omega) < (k+1)\,2^{-n}$$

and so f_n only assumes the values $0, \pm 2^{-n}, \pm 2 \cdot 2^{-n}, \cdots$. Hence, we see that (1.45.1+2) hold. Moreover, g_n only assume the $(1 + n\,2^{n+1})$ values $0, \pm 2^{-n}, \pm 2 \cdot 2^{-n}, \ldots, \pm n \cdot 2^n \cdot 2^{-n}$ and $g_n(\omega) = f_n(\omega)$ if $|f_n(\omega)| \leq n$. Thus, we see that (1.45.3+4) hold. □

1.47 The Image Catastrophe

By the very definition of a measurable function, we have that the co-image of a measurable set under a measurable function is measurable. It is natural to ask if this also holds for images. In 1905, Henri Léon Lebesgue (1875 – 1941) published a paper *Sur les fonctions représentables analytiquement* which became one of the cornerstones in measure theory. One of the theorems in the paper is (in modern language)

> *Let B be a Borel subset of the plane, then its projection onto the first axis is a Borel subset of the line*

This theorem is false! In 1916, the young ingenious Russian mathematician M. M. Souslin came to his teacher Nikolai Nikolaievich Lusin (1883 – 1950) and claimed that there was a mistake in Lebesgue's paper. Lusin did not believe him at first, but he quickly became convinced, and in 1917 Souslin published a counterexample to Lebesgue's theorem. Actually, Souslin constructed a concrete sequence $U_1, U_2, \cdots$ of open subsets of the plane such that the projection of their intersection onto the first axis is not a Borel set. Souslin died a few years later in the war between "the reds" and "the whites" during the Russian revolution. He only left three pages of mathematics, but the three pages became the start of a new mathematical field (the theory of analytic sets), which today has found many important applications in various mathematical fields. The morale of this story is: ***Never trust your textbooks too much!***.

This means that there exist nice measurable spaces $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ and a nice measurable (even continuous) function $f : \Omega \to \Lambda$ and a set $F_0 \in \mathcal{F}$ such

that the image $f(F_0)$ does not belong to $\mathcal{G}$. This phenomenon has been called *the image catastrophe*. However, the catastrophe turned out to be not so catastrophic. Souslin actually showed that the image of a Borel set under a Borel function belongs to a σ-algebra which is only slightly larger than the Borel σ-algebra and which, for most purposes in measure theory, may replace the Borel σ-algebra.

1.48 Functions on Product Spaces

Let $(S,\mathcal{A})$, $(T,\mathcal{B})$ and $(M,\mathcal{M})$ be measurable spaces and consider the product space $S\times T$ with its product σ-algebra $\mathcal{A}\otimes\mathcal{B}$. Let f be a measurable function from $(S\times T,\mathcal{A}\otimes\mathcal{B})$ into $(M,\mathcal{M})$ and let $a\in S$ and $b\in T$ be two given points. Set

$$f_a(t)=f(a,t) \quad , \quad f^b(s)=f(s,b)$$

for all $s\in S$ and all $t\in T$. Then f^b is measurable from $(S,\mathcal{A})$ into $(M,\mathcal{M})$ and f_a is measurable from $(T,\mathcal{B})$ into $(M,\mathcal{M})$.

Proof: Let $p_a(t)=(a,t)$, then p_a is a function from T into $S\times T$ and

$$p_a^{-1}(A\times B)=\begin{cases} B & \text{if } a\in A \\ \emptyset & \text{if } a\notin A \end{cases}$$

Hence, by (1.38.6) and the definition of the product σ-algebra $\mathcal{A}\otimes\mathcal{B}$ we have that p_a is measurable from $(T,\mathcal{B})$ into $(S\times T,\mathcal{A}\otimes\mathcal{B})$; and since $f_a=f\circ p_a$, we see that f_a is measurable by (1.38.4). In a similar manner, we find that f^b is measurable. □

The functions f_a and f^b are called *the partial functions of* f, and we shall use the notation $f(a,\cdot):=f_a$ and $f(\cdot,b):=f^b$.

1.49 The Functional Sierpińsky Lemma

Let Ω be a set and let $\mathcal{W}\subseteq\mathcal{V}$ be two non-empty sets of functions from Ω into $\overline{\mathbf{R}}_+$ satisfying the following

(1.49.1) $\quad f+g\in\mathcal{V}$ *and* $af\in\mathcal{V} \quad \forall f,g\in\mathcal{V}\ \forall 0\le a<\infty$

(1.49.2) $\quad$ *If* $f_n\uparrow f$ *and* $f_n\in\mathcal{V} \quad \forall n\ge 1$ *then* $f\in\mathcal{V}$

(1.49.3) *If* $F \subseteq \Omega$ *and* $a + 1_F \in \mathcal{V}$ $\forall a > 0$ *then* $1_F \in \mathcal{V}$ *and* $1_{F^c} \in \mathcal{V}$

(1.49.4) $f \cdot g \in \mathcal{H} \qquad \forall f, g \in \mathcal{W}$

(1.49.5) $\forall g \in \mathcal{W}\ \exists b \in]0, \infty[$ *so that* $g(\omega) \leq b$ $\forall \omega$ *and* $b - g(\cdot) \in \mathcal{H}$

where $\mathcal{H} = con_\sigma(\mathcal{W})$ *is the smallest* $(\uparrow)$*-stable convex cone containing* $\mathcal{W}$ *[see the remark below]. Then* $\mathcal{V}$ *contains every non-negative* $\sigma(\mathcal{W})$ *-measurable function* $f : \Omega \to \overline{\mathbf{R}}_+$.

Remarks: (1): Let $\mathcal{C}$ be a non-empty set of functions from Ω into $\overline{\mathbf{R}}_+$. Then we say that $\mathcal{C}$ is a *convex cone* if $\mathcal{C}$ satisfies (1.49.1), that is, if $f + g \in \mathcal{C}$ and $af \in \mathcal{C}$ whenever $f, g \in \mathcal{C}$ and $a \in \mathbf{R}_+$. We say that $\mathcal{C}$ is $(\uparrow)$*-stable* if $\mathcal{C}$ satisfies (1.49.2), that is, if $f \in \mathcal{C}$ whenever $f_n \in \mathcal{C}$ and $f_n \uparrow f$.

(2): Let $\mathcal{C}$ be an arbitrary nonempty set of functions from Ω into $\overline{\mathbf{R}}_+$, then exactly as in (1.5) one shows, that there exists a *smallest* $(\uparrow)$*-stable, convex cone* containing $\mathcal{C}$, which we denote $con_\sigma(\mathcal{C})$.

(3): A set $\mathcal{C}$ of functions from Ω into $\overline{\mathbf{R}}_+$ is said to satisfy *Stone's condition* if $\mathcal{C}$ satisfies (1.49.4) and (1.49.5), that is, if $f \cdot g \in con_\sigma(\mathcal{C})$ whenever $f, g \in \mathcal{C}$, and if for every $g \in \mathcal{C}$, there exists $b > 0$ such that $g(\omega) \leq b$ and $b - g(\cdot) \in con_\sigma(\mathcal{C})$.

(4): Suppose that $\mathcal{W}_j$ is a set of functions from Ω into $\mathbf{R}_+$ satisfying *Stone's condition* for all j in the set J . Let

$$\mathcal{W} = \left\{ \prod_{j \in \gamma} g_j \;\middle|\; \gamma \text{ finite } \subseteq J \text{ and } g_j \in \mathcal{W}_j \quad \forall j \in \gamma \right\} \tag{1.49.6}$$

and let $\mathcal{H}_j = con_\sigma(\mathcal{W}_j)$ and $\mathcal{H} = con_\sigma(\mathcal{W})$. Then using the argument in the proof of (i) below and a simple induction in the number of elements of γ we find

$$\prod_{j \in \gamma} h_j \in \mathcal{H} \quad \text{if} \quad \gamma \subseteq J \quad \text{is finite and} \quad h_j \in \mathcal{H}_j \quad \forall j \in \gamma \tag{1.49.7}$$

Using (1.49.7) and the trivial identity

$$\prod_{j=1}^{n} b_j - \prod_{j=1}^{n} g_j(\omega) = \sum_{j=1}^{n} \left\{ b_1 \cdots b_{j-1} \cdot (b_j - g_j(\omega)) \cdot g_{j+1}(\omega) \cdots g_n(\omega) \right\}$$

we obtain the following useful observation

(1.49.8) If $\mathcal{W}_j$ satisfies Stone's condition for all $j \in J$, then so does the set $\mathcal{W}$ defined by (1.49.6)

Proof: Before going to the proof of (1.49), I shall first establish the following facts: Let $\mathcal{W}$ be a non-empty set of functions from $\Omega \to \mathbf{R}_+$ satisfying Stone's condition. Set $\mathcal{H} = con_\sigma(\mathcal{W})$ and

$$\mathcal{W}_0 = \{ f \in \mathcal{H} \mid \exists b \in]0,\infty[\ : \ f(\omega) \le b \ \ \forall \omega \text{ and } b - f(\cdot) \in \mathcal{H} \}$$

$$\mathbf{W}_0 = \{ f : \Omega \to \mathbf{R}_+ \mid \forall \varepsilon > 0 \ \exists g \in \mathcal{W}_0 \ : \ |g(\omega) - f(\omega)| < \varepsilon \ \ \forall \omega \in \Omega \}$$

Note that $\mathcal{W}_0 \subseteq B(\Omega)$ and that $\mathbf{W}_0$ is the closure of $\mathcal{W}_0$ in the metric space $(B(\Omega), \|\cdot\|_\Omega)$ [see (P.6)]. Then I claim that we have

(i) $\quad a\,1_\Omega \in \mathcal{H}$ and $f \cdot g \in \mathcal{H} \quad \forall f,g \in \mathcal{H} \ \forall a \in \mathbf{R}_+$

(ii) $\quad \varphi(f) \in \mathbf{W}_0 \quad \forall f \in \mathcal{W}_0 \ \forall \varphi$ continuous $: \mathbf{R}_+ \to \mathbf{R}_+$

(iii) $\quad a + f \in \mathcal{W}_0$ and $f \cdot g \in \mathbf{W}_0 \quad \forall f,g \in \mathbf{W}_0 \ \forall 0 < a < \infty$

(i): Let g be any function in $\mathcal{W}$ and choose $b > 0$ such that $b - g \in \mathcal{H}$. Then $a\,1_\Omega = \frac{a}{b}\,g + \frac{a}{b}\,(b\,1_\Omega - g) \in \mathcal{H}$ for all $a \ge 0$ because $\mathcal{H}$ is a convex cone. The second part of (i) is very similar to the first part of the proof of Sierpińsky's stability lemma (1.6): Set

$$\mathcal{H}_f = \{ g : \Omega \to \overline{\mathbf{R}}_+ \mid g \cdot f \in \mathcal{H} \} \qquad \forall f : \Omega \to \overline{\mathbf{R}}_+$$

Then $\mathcal{H}_f$ is a $(\uparrow)$-stable convex cone, and by assumption we have $\mathcal{W} \subseteq \mathcal{H}_f$ for all $f \in \mathcal{W}$. So by minimality of $\mathcal{H}$ we have $\mathcal{H} \subseteq \mathcal{H}_f$ for all $f \in \mathcal{W}$, and since $g \in \mathcal{H}_f$ if and only if $f \in \mathcal{H}_g$, we see that $\mathcal{W} \subseteq \mathcal{H}_g$ for all $g \in \mathcal{H}$. Hence, by minimality of $\mathcal{H}$ we have $\mathcal{H} \subseteq \mathcal{H}_g$ for all $g \in \mathcal{H}$, and so $fg \in \mathcal{H}$ for all $f,g \in \mathcal{H}$.

(ii): Let $f \in \mathcal{W}_0$ be given and choose $b > 0$ such that $b - f \in \mathcal{H}$. Then we choose $c > 0$ such that $\varphi(t) \le c$ for all $0 \le t \le b$. Let $B_n(t)$ be the nth Bernstein polynomial of φ on $[0,b]$ [see Exc.4.29]:

$$B_n(t) = b^{-n} \sum_{j=0}^{n} \binom{n}{j} \varphi\left(\tfrac{jb}{n}\right) t^j \, (b-t)^{n-j}$$

By Exc.4.29 we have that $B_n(t) \to \varphi(t)$ uniformly on $[0,b]$. Since $\varphi \ge 0$, then by (i) we have that $f_n = B_n(f)$ belongs to $\mathcal{H}$. Moreover, since

$$c - B_n(t) = b^{-n} \sum_{j=0}^{n} \binom{n}{j} \left(c - \varphi\left(\tfrac{jb}{n}\right)\right) t^j \, (b-t)^{n-j}$$

and $c - \varphi(t) \ge 0$ for all $t \in [0,b]$, we also have that $c - f_n$ belongs to $\mathcal{H}$. Thus, $f_n \in \mathcal{W}_0$, and since $b - f \in \mathcal{H}$, we have that $f(\omega) \in [0,b]$ for all $\omega \in \Omega$. Hence, we see that $f_n \to \varphi(f)$ uniformly on Ω, and so $\varphi(f) \in \mathbf{W}_0$.

(iii): Let $f \in \mathbf{W}_0$ and let $a > 0$ be given. Then there exists $f_k \in \mathcal{W}_0$ such that $| f(\omega) - f_k(\omega) | \leq \frac{1}{3} 2^{-k} a$ for all ω and all k. Let $g_k = f_k + (1 - 2^{-k}) a$, then $g_k \to a + f$ and

$$\begin{aligned} g_k(\omega) &= f_k(\omega) + (1 - 2^{-k})\, a \leq f(\omega) + \tfrac{1}{3}\, 2^{-k}\, a + (1 - 2^{-k})\, a \\ &\leq f_{k+1}(\omega) + \tfrac{1}{3}\, 2^{-k-1}\, a + \tfrac{1}{3}\, 2^{-k}\, a + (1 - 2^{-k})\, a \\ &\leq f_{k+1}(\omega) + (1 - 2^{-k-1})\, a = g_{k+1}(\omega) \end{aligned}$$

Hence $g_n \uparrow a + f$ and $g_k \in \mathcal{H}$ because $\mathcal{H}$ is a convex cone containing f_k and $(1 - 2^{-k}) a 1_\Omega$. Thus, $a + f \in \mathcal{H}$ by $(\uparrow)$ -stability of $\mathcal{H}$, and since $f \in \mathbf{W}_0 \subseteq B(\Omega)$, we have that f is bounded. Moreover, since $f_k \leq f + \frac{1}{3} 2^{-k} a$, we can find $c > 0$ such that $f_k(\omega) \leq c$ for all ω and all k. Applying (ii) with $\varphi(t) = (c - t)^+$, shows that $c - f_k \in \mathbf{W}_0$ and so $c - f \in \mathbf{W}_0$. Thus, by the argument above, we have that

$$(c + 2a) - (a + f) = a + (c - f) \in \mathcal{H}$$

Hence, we see that $a + f \in \mathcal{W}_0$.

Let $f, g \in \mathbf{W}_0$; then there exists $f_n, g_n \in \mathcal{W}_0$ such that $f_n \to f$ and $g_n \to g$ uniformly on Ω. But then $f_n g_n \to fg$ uniformly on Ω, and by (i) and the trivial identity $ab - \varphi\psi = (a - \varphi)\psi + a(b - \psi)$ we see that $f_n g_n \in \mathcal{W}_0$. Hence $fg \in \mathbf{W}_0$ for all $f, g \in \mathbf{W}_0$.

Thus, (i)-(iii) are established and we can proceed with the proof of (1.49). Let us define

$$\mathcal{G} = \{ G \subseteq \Omega \mid \exists g \in \mathcal{W}_0 \text{ so that } G = \{ g > 1 \} \}$$

$$\mathcal{D} = \{ F \subseteq \Omega \mid 1_F \in \mathcal{V} \}$$

If $\mathcal{D}_0$ is the smallest $(\uparrow, \downarrow)$-stable paving containing $\mathcal{G}$, then I claim that

(iv) $\mathcal{D}$ is $(C, \downarrow, \uparrow)$-stable

(v) $\mathcal{G}$ is a $(\cap f)$-stable paving such that $\mathcal{G} \subseteq \mathcal{D}$ and $\sigma(\mathcal{W}) \subseteq \sigma(\mathcal{G})$

(vi) $G^c \in \mathcal{D}_0 \qquad \forall G \in \mathcal{G}$

(iv): By (1.49.2) we have that $\mathcal{D}$ is $(\uparrow)$-stable, and if $F \in \mathcal{D}$, then $a + 1_F$ belongs to $\mathcal{V}$ for all $a \geq 0$ by (1.49.1), because $a 1_\Omega \in \mathcal{H} \subseteq \mathcal{V}$. Hence, by (1.49.3) we have that $\mathcal{D}$ is $(C, \uparrow)$ -stable and, thus, $(\uparrow, \downarrow)$-stable.

(v): Let $G, H \in \mathcal{G}$ be given, then there exist $g, h \in \mathcal{W}_0$ with $G = \{g > 1\}$ and $H = \{h > 1\}$. If we set $f = 1 + (g - 1)^+ (h - 1)^+$, then $f \in \mathcal{W}_0$

by (ii)+(iii), and $G \cap H = \{f > 1\}$. Hence $\mathcal{G}$ is $(\cap f)$-stable. Let us set $g_n = 1 \wedge (n(g-1)^+)$, then $a + g_n \in \mathcal{W}_0 \subseteq \mathcal{V}$ for all $a > 0$ by (ii)+(iii), and $a + g_n \uparrow a + 1_G$. Hence, $a + 1_G \in \mathcal{V}$ by (1.49.2) and so $G \in \mathcal{D}$ by (1.49.3); thus, $\mathcal{G} \subseteq \mathcal{D}$. Let $g \in \mathcal{W}$, then $c^{-1}g \in \mathcal{W}_0$ for all $c > 0$, and so $\{g > c\} \in \mathcal{G}$. Hence, we have that $\sigma(\mathcal{W}) \subseteq \sigma(\mathcal{G})$.

(vi): Let $G \in \mathcal{G}$ be given, and let $g \in \mathcal{W}_0$ and $b > 2$ be chosen such that $G = \{g > 1\}$ and $b - g \in \mathcal{H}$, and set

$$g_n(\omega) = \frac{b - g(\omega)}{b - 1 - \frac{1}{n}} \qquad \forall \omega \in \Omega \ \forall n \geq 1$$

Then $g_n \in \mathcal{W}_0$ and $\{g_n > 1\} = \{g < 1 + \frac{1}{n}\} \downarrow G^c$; hence, $G^c \in \mathcal{D}_0$.

Thus, (iv)–(vi) are proved, and by Sierpińsky's stability lemma (1.6), we conclude that $\sigma(\mathcal{W}) \subseteq \sigma(\mathcal{G}) \subseteq \mathcal{D}$. Thus, by (1.49.1) we see that every non-negative simple $\sigma(\mathcal{W})$-measurable function belongs to $\mathcal{V}$. Hence, by (1.44.4) and (1.49.2) we conclude that every non-negative $\sigma(\mathcal{W})$-measurable function belongs to $\mathcal{V}$. □

1.50 The Projective Limit Theorem

Let $(\Omega_\alpha, \mathcal{F}_\alpha, \mu_\alpha)$ be a finite measure space and let $\wp_\alpha : \Omega \to \Omega_\alpha$ be a function for all α in the given set Γ . Suppose that $\mathcal{G}_\alpha$ and $\mathcal{K}_\alpha$ are nonempty pavings on Ω_α and $\mathcal{K}$ is a paving on Ω satisfying

(1.50.1) $\mathcal{G}_\alpha$ *is $(\cap f)$-stable and* $\sigma(\mathcal{G}_\alpha) = \mathcal{F}_\alpha$ *for all* $\alpha \in \Gamma$

(1.50.2) $\forall \alpha \in \Gamma \ \forall G \in \mathcal{G}_\alpha \ \exists G_1, G_2, \ldots \in \mathcal{G}_\alpha$ *disjoint* : $\Omega_\alpha \setminus G = \bigcup_{j=1}^{\infty} G_j$

(1.50.3) $\mathcal{K}$ *is compact, and* $\mathcal{K}_\alpha$ *approximates* μ_α *on* $\mathcal{G}_\alpha$ *for all* $\alpha \in \Gamma$

(1.50.4) $\wp_\alpha^{-1}(K) \in \mathcal{K} \qquad \forall K \in \mathcal{K}_\alpha \ \forall \alpha \in \Gamma$

(1.50.5) $\forall \alpha, \beta \in \Gamma \ \exists \gamma \in \Gamma \quad : \quad \mathcal{G}^\alpha \cup \mathcal{G}^\beta \subseteq \mathcal{G}^\gamma$

(1.50.6) $\mu_\alpha(F) = \mu_\beta(H) \quad \forall F \in \mathcal{G}_\alpha \ \forall H \in \mathcal{G}_\beta$ *with* $\wp_\alpha^{-1}(F) = \wp_\beta^{-1}(H)$

where $\mathcal{G}^\alpha = \wp_\alpha^{-1}(\mathcal{G}_\alpha)$ for all $\alpha \in \Gamma$. Let $\mathcal{B} = \sigma(\wp_\alpha \mid \alpha \in \Gamma)$. Then $\mathcal{B}$ is a σ-algebra on Ω and there exists a unique measure μ on $(\Omega, \mathcal{B})$ satisfying

(1.50.7) $\mu_\alpha(F) = \mu(\wp_\alpha^{-1}(F)) \qquad \forall F \in \mathcal{F}_\alpha \ \forall \alpha \in \Gamma$

Remark: Let $(\Omega_\alpha, \mathcal{F}_\alpha, \mu_\alpha)$ be a finite measure space, let $\wp_\alpha : \Omega \to \Omega_\alpha$ be a function for all α in a given set Γ, and let $\mathcal{B} = \sigma(\wp_\alpha \mid \alpha \in \Gamma)$. Then a measure μ on $(\Omega, \mathcal{B})$ satisfying (1.50.7) is called a *projective limit* of $\{(\mu_\alpha, \wp_\alpha) \mid \alpha \in \Gamma\}$ and is denoted $\mu = \varprojlim (\mu_\alpha, \wp_\alpha)$.

Proof: Let $\mathcal{G} = \cup_{\alpha\in\Gamma} \mathcal{G}^\alpha$ By (1.38.5), (1.50.1) and (1.50.5) we see that $\mathcal{G}$ is $(\cap f)$-stable and $\sigma(\mathcal{G}) = \mathcal{B}$ and by (1.50.2) we have that $\mathcal{G}_\alpha$ and $\mathcal{G}$ contains the empty set $\emptyset$ for all $\alpha \in \Gamma$. Moreover, by (1.50.6) we have that

$$\nu(H) := \mu_\alpha(G) \quad \text{if} \quad H = \wp_\alpha^{-1}(G) \text{ for some } \alpha \in \Gamma \text{ and some } G \in \mathcal{G}_\alpha$$

is a well-defined set function on $\mathcal{G}$. If $H \in \mathcal{G}^\alpha$ and $H_1, H_2, \ldots$ are disjoint sets in $\mathcal{G}^\alpha$ such that $\cup_j H_j \subseteq H$, then there exists $G, G_1, G_2, \ldots \in \mathcal{G}_\alpha$ with $H = \wp_\alpha^{-1}(G)$ and $H_j = \wp_\alpha^{-1}(G_j)$. Then $\wp_\alpha^{-1}(G_i \cap G_j) = H_i \cap H_j = \emptyset$ for all $i \neq j$; hence by (1.50.6) we have that $G_i \cap G_j = \emptyset$ μ_α-a.s. for all $i \neq j$, and in the same manner we see that $\cup_j G_j \subseteq G$ μ_α-a.s. Thus, by (1.19.6) and (1.19.4) we find:

$$\sum_{j=1}^{\infty} \nu(H_j) = \sum_{j=1}^{\infty} \mu_\alpha(G_j) = \mu_\alpha\left(\bigcup_{j=1}^{\infty} G_j\right) \leq \mu_\alpha(G) = \nu(H)$$

Hence, ν is superadditive on $\mathcal{G}^\alpha$ and I claim that we have

(i) ν is inner tight on $\mathcal{G}$

(ii) $\mathcal{K}$ approximates ν on $\mathcal{G}$

(i): Let $G, H \in \mathcal{G}$. Then by (1.50.5) there exists $\alpha \in \Gamma$ such that $G, H \in \mathcal{G}^\alpha$, and by (1.50.2) there exist disjoint sets $G_1, G_2, \ldots \in \mathcal{G}^\alpha$ such that $H^c = G_1 \cup G_2 \cup \ldots$. Then $G \cap G_j \in \mathcal{G}^\alpha$ and

$$G \setminus H = \bigcup_{j=1}^{\infty} G \cap G_j$$

Since ν is σ-additive on $\mathcal{G}^\alpha$, we see that ν satisfies (1.26.3). Hence, (i) follows from (1.26).

(ii): Let $G \in \mathcal{G}$ and let $a < \nu(G)$. Then $G = \wp_\alpha^{-1}(G_0)$ for some $\alpha \in \Gamma$ and some $G_0 \in \mathcal{G}_\alpha$. Since $\mathcal{K}_\alpha$ approximates μ_α on $\mathcal{G}_\alpha$, there exist $C \in \mathcal{K}_\alpha$ and disjoint sets $G_1, \ldots, G_n \in \mathcal{G}_\alpha$ such that

$$\bigcup_{j=1}^{n} G_j \subseteq C \subseteq G_0 \quad \text{and} \quad a < \sum_{j=1}^{n} \mu_\alpha(G_j)$$

Let $K = \wp_\alpha^{-1}(C)$ and $H_j = \wp_\alpha^{-1}(G_j)$. Then $K \in \mathcal{K}$ by (1.50.4), and $H_1, \ldots, H_n$ are disjoint sets in $\mathcal{G}$ such that

$$\bigcup_{j=1}^{n} H_j = \wp_\alpha^{-1}\Big(\bigcup_{j=1}^{n} G_j\Big) \subseteq \wp_\alpha^{-1}(C) = K \subseteq \wp_\alpha^{-1}(G_0) = G$$

$$a < \sum_{j=1}^{n} \mu_\alpha(G_j) = \sum_{j=1}^{n} \nu(H_j) \leq \nu_*(K)$$

Hence, $\mathcal{K}$ approximates ν on $\mathcal{G}$.

By assumption, we have that $\mathcal{K}$ is a compact paving. Hence, by (i), (ii), (1.30) and the extension theorem (1.27) there exists a measure μ on $\sigma(\mathcal{G})$ such that μ is an extension of ν . Thus, we have

$$\mu(\wp_\alpha^{-1}(G)) = \mu_\alpha(G) \qquad \forall\, \alpha \in \Gamma \quad \forall\, G \in \mathcal{G}_\alpha$$

Let $\alpha \in \Gamma$ be given, let G be some fixed set in $\mathcal{G}_\alpha$, and let $G_1, G_2, \ldots$ be chosen according to (1.50.2). Then setting $F_1 = G$ and $F_n = G_{n-1}$ for $n \geq 2$, we see that (1.50.7) follows from (1.50.1) and the uniqueness theorem I (1.7). In the same manner we see that any two measures satisfying (1.50.7) must coincide on $\mathcal{B}$, and so μ is uniquely determined by (1.50.7). □

1.51 References to Chapter 1

Measure theory emerged with the works of R. L. Baire, E. Borel and H. Lebesgue, who constructed the Borel σ-algebra and the Lebesgue measure on $\mathbf{R}$ and $\mathbf{R}^k$ in the period 1895–1905. Later, the subject was extended to abstract sets and σ-algebras by C. Carathéodry, G. Fubini, L. Tonelli, W. Sierpińsky and many other. The Sierpińsky lemmas (1.6) and (1.49) was found by Sierpińsky. in 1930's, where he considered a more abstract version invlovong abstract operations on general sets. The lemmas has been reproved several times and special cases have got their own name, for instance, *the monotone class lemma* and *Dynkin's lemma.* The classical extension theorem states that a σ-additive setfunction on an algebra can be extended to a measure on the σ-algebra generated by the algebra, the modern extension theorem (1.27) is due to F. Topsøe (see [6]). The main advantage of the modern version is that it allows us to extend a setfunction defined on an arbitrary paving to a measure on the σ-algebra in one stroke. As explained in the Introduction, it was only with aid of the abstract measure theory that Kolmogorov in 1933 was able to state a rigorous mathematical model for probability. The main justification of Kolmogorov's model is given by the projective limit theorem (1.50) which was found by Kolmogorov in 1933 (see [3]), in case that Γ is the set of all

finite subsets of a given set T , $\Omega_\alpha = \mathbf{R}^\alpha$, $\Omega = \mathbf{R}^T$ and $\wp_\alpha$ is the projection of $\mathbf{R}^T$ onto $\mathbf{R}^\alpha$. In Chapter 8 of Volume 2, we shall see how the projective limit theorem (1.50) can be used to show that every stochastic phenomenon, specified by consistent conditions, can be modelled within Kolmogorov's model. Halmos' book [1] served for many years as the classical textbok of measure theory. In [2], [3] and [6] you will find a more modern approach to measure theory and its relation to modern analysis, and in [4] you will a systematic treatment of the theory of Souslin sets and analytic spaces, which I have introduced in Excirses 1.90–1.101.

[1] Halmos, P. R. (1950): *Measure Theory*, Princeton University Press, Princeton, New York and London

[2] Kelley, J. L. and Srinvasan, T. P. (1988): *Measure and Integral*, Springer-Verlag, Berlin and New York

[3] Kolmogorov, A. N. (1933): *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin and New York

[4] Rogers, C. A. (1980): *Analytic Sets*, Academic Press, London and New York

[5] Rudin, W. (1966): *Real and Complex Analysis*, McGraw-Hill Book Co., New York and London

[6] Topsøe, F. (1970): *Topology and Measure*, Springer-Verlag, Berlin and New York

1.52 Exercises to Chapter 1

Exc. 1.1: Prove the laws of de Morgan, that is, prove the propositions (P.2.1)–(P.2.4).

Exc. 1.2: Prove the distributive laws of set theory, that is, prove the propositions (P.2.5)–(P.2.15).

Exc. 1.3: Let A, B, A_j and B_j be subsets of a given set Ω for all j in the set J . Verify that the symmetric difference satisfies the following relations:

(i) $A \triangle B = A^c \triangle B^c$

(ii) $A \triangle \big(\bigcup_{j \in J} B_j \big) \subseteq \bigcup_{j \in J} (A \triangle B_j)$

(iii) $A\triangle(\bigcap_{j\in J} B_j) \supseteq \bigcap_{j\in J} (A\triangle B_j)$

(iv) $(\bigcup_{j\in J} A_j)\triangle(\bigcup_{j\in J} B_j) \subseteq \bigcup_{j\in J} (A_j\triangle B_j)$

(v) $(\bigcap_{j\in J} A_j)\triangle(\bigcap_{j\in J} B_j) \subseteq \bigcup_{j\in J} (A_j\triangle B_j)$

Exc. 1.4: Verify propositions (P.4.8)–(P.4.18) giving the relationship between the indicator functions and the operations of set theory.

Exc. 1.5: Let $A_1, A_2, \ldots$ be a sequence of subsets of the set Ω . Show that

$$\limsup\,(A_n \setminus A_{n+1}) = \limsup\,(A_{n+1} \setminus A_n) = (\limsup\, A_n) \setminus (\liminf\, A_n)$$

Hint: Let $\omega \in (\limsup A_n) \setminus (\liminf A_n)$ and set $M = \{n \in \mathbf{N} \mid \omega \in A_n\}$ and $L = \mathbf{N} \setminus M$. Then L and M are infinite sets of integers. Hence, we may define the integers: $\sigma(k)$ and $\tau(k)$ inductively as follows: $\sigma(0) = \tau(0) = 0$ and

$$\sigma(k) = \min\{n \in M \mid n > \tau(k-1)\} \quad , \quad \tau(k) = \min\{n \in L \mid n > \sigma(k)\}$$

for $k \geq 1$. Verify that $\omega \in A_{\sigma(k)} \setminus A_{\sigma(k)-1}$ for all $k \geq 2$, and $\omega \in A_{\tau(k)-1} \setminus A_{\tau(k)}$ for all $k \geq 1$. Justify, that this shows the following inclusions:

$$(\limsup\, A_n) \setminus (\liminf\, A_n) \subseteq \limsup\,(A_{n+1} \setminus A_n)$$

$$(\limsup\, A_n) \setminus (\liminf\, A_n) \subseteq \limsup\,(A_n \setminus A_{n+1})$$

The converse inclusions are easy.

Exc. 1.6: Let μ be a measure on $(\Omega, \mathcal{F})$, and let $\{A_n\} \subseteq \mathcal{F}$. Suppose that

(*) either $\sum_{n=1}^{\infty} \mu(A_n \setminus A_{n+1}) < \infty$ or $\sum_{n=1}^{\infty} \mu(A_{n+1} \setminus A_n) < \infty$

and show that $\mu((\limsup\, A_n) \setminus (\liminf\, A_n)) = 0$. [Apply Exc.1.5.] Suppose that (*) holds, and that $\liminf_{n\to\infty} \mu(A_n) = 0$, then show that $\mu(\limsup\, A_n) = 0$.

Exc. 1.7: *(Binomial coefficients)* Let $a \in \mathbf{R}$ be a real number, and let $n \in \mathbf{Z}$ be an integer. Then, as you certainly know, we define *the binomial coefficient*: $\binom{a}{n}$ in the following way:

$$\binom{a}{n} = \frac{a(a-1)\cdots(a-n+1)}{n!} \quad \text{if} \quad n \geq 1$$

$$\binom{a}{0} = 1 \quad , \quad \binom{a}{n} = 0 \quad \text{if} \quad n \leq -1$$

Show that

(i) $\binom{a}{n} = 0 \quad \forall a \in \{0, 1, 2, \ldots, n-1\} \; \forall n \in \mathbf{N}$

(ii) $\binom{a}{n-1} + \binom{a}{n} = \binom{a+1}{n} \quad \forall a \in \mathbf{R} \; \forall n \in \mathbf{Z}$

In the following formulas a, b, c denote arbitrary real numbers, and k, m, n denote arbitrary non-negative integers. Prove the following formulas:

(iii) $\sum_{j=0}^{n} \binom{n}{j} a^j = (1+a)^n$

(iv) $\sum_{j=0}^{n} j \binom{n}{j} a^j = a\, n\, (1+a)^{n-1}$

(v) $\sum_{j=0}^{n} (-1)^j \binom{n}{j} = 0 \quad \text{if} \quad n \geq 1$

(vi) $\sum_{j=0}^{n} \binom{n}{j} = 2^n$

(vii) $\sum_{j=0}^{n} j \binom{n}{j} = n\, 2^{n-1}$

(viii) $\sum_{j=0}^{n} (-1)^j j \binom{n}{j} = 0 \quad \text{if} \quad n \geq 2$

(ix) $\sum_{j=0}^{k} \binom{n}{j} \binom{n-j}{k-j} a^j = \binom{n}{k} (1+a)^k$

(x) $\binom{-a}{k} = (-1)^k \binom{a+k-1}{k}$

(xi) $2^{-2n} \binom{2n}{n} = (-1)^n \binom{-\frac{1}{2}}{n}$

(xii) $\sum_{j=m}^{n} \binom{a+j}{k} = \binom{a+n+1}{k+1} - \binom{a+m}{k+1} \quad \text{if} \quad n \geq m$

(xiii) $\sum_{j=0}^{n} (-1)^j \binom{a}{j} = (-1)^n \binom{a-1}{n}$

(xiv) $\sum_{j=0}^{n} \binom{j+k-1}{k-1} = \binom{n+k}{k} = \binom{n+k}{n}$

(xv) $\sum_{j=0}^{n} \binom{a}{j} \binom{b}{n-j} = \binom{a+b}{n}$

(xvi) $\sum_{j=0}^{n} \binom{n}{j}^2 = \binom{2n}{n}$

(xvii) $\sum_{j=1}^{n} (-1)^{n-j} \binom{n}{j} \binom{k+j}{k+1} = \binom{k}{n-1}$

(xviii) $\sum_{j=0}^{n} (-1)^{n-j} \binom{a}{j} = \binom{a-1}{n}$

(xix) $\sum_{j=0}^{n-k} (-1)^j \binom{a}{j} \binom{n-j}{k} = \binom{n-a}{n-k}$

(xx) $\sum_{j=0}^{n} \binom{a+n-j-1}{n-j} \binom{b+j-1}{j} = \binom{a+b+n-1}{n}$

(xxi) $\sum_{j=0}^{n} (-1)^{n-j} j^k \binom{n}{j} = \begin{cases} 0 & \text{if } k < n \\ n! & \text{if } k = n \end{cases}$

(xxii) $\sum_{j=1}^{n} (-1)^{j-1} \frac{1}{j} \binom{n}{j} = \sum_{j=1}^{n} \frac{1}{j}$

(xxiii) $n! = \theta_n \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}$ if $n \geq 1$, where θ_n satisfies:

$$\exp\left(\frac{1}{12n+1}\right) < \theta_n < \exp\left(\frac{1}{12n}\right)$$

(xxiv) $\binom{n}{k} = \alpha_{nk} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} \sqrt{\frac{n}{2\pi k(n-k)}}$ if $n > k \geq 1$, where

$$\exp\left(-\frac{n}{12k(n-k)}\right) < \alpha_{nk} < 1$$

Formulas (i)–(xxiv) presents some of the most important formulas for binomial coefficients. Formulas (i)–(xvi) are straightforward to prove; formulas (xvii)–(xxii) are a bit more complicated; formula (xxiii) is difficult and goes under the name: *Stirling's formula*, and formula (xxiv) follows easily from Stirling's formula.

Exc. 1.8: *(The inclusion-exclusion principle)* Let $A_1, \ldots, A_n$ be subsets of Ω, and let the functions $f_0, f_1, \ldots, f_n$ be defined by

$$f_0(\omega) = 1 \quad , \quad f_k(\omega) = \sum_{1 \le i_1 < \ldots < i_k \le n} 1_{A_{i_1}}(\omega)\, 1_{A_{i_2}}(\omega) \cdots 1_{A_{i_k}}(\omega)$$

for all $\omega \in \Omega$. Let the sets B_j and C_j be given by

$$B_j = \{\omega \in \Omega \,|\, \omega \text{ belongs to exactly } j \text{ of the set } A_1, \ldots, A_n\}$$

$$C_j = \{\omega \in \Omega \,|\, \omega \text{ belongs to at least } j \text{ of the set } A_1, \ldots, A_n\}$$

for $0 \le j \le n$. Show that for every $\omega \in \Omega$ we have

(i) $\quad f_k(\omega) = \binom{f_1(\omega)}{k} = \sum_{j=k}^{n} \binom{j}{k} 1_{B_j}(\omega) \qquad \forall 0 \le k \le n$

(ii) $\quad 1_{B_k}(\omega) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} f_j(\omega) \qquad \forall 0 \le k \le n$

(iii) $\quad 1_{C_k}(\omega) = \sum_{i=k}^{n} 1_{B_i}(\omega) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j-1}{k-1} f_j(\omega) \qquad \forall 1 \le k \le n$

Hint: (ii): Let $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ be the $(n+1) \times (n+1)$-matrices, where $a_{ij} = \binom{j}{i}$ and $b_{ij} = (-1)^{j-i}\binom{j}{i}$ for $0 \le i, j \le n$. Show that $\mathbf{AB} = I_{n+1}$ where I_{n+1} is the $(n+1) \times (n+1)$-identity matrix, and argue that (ii) follows from (i).

Exc. 1.9: Let μ be a content on $(\Omega, \mathcal{F})$, and let us define $d(A, B) = \mu(A \triangle B)$ for all $A, B \in \mathcal{F}$. Show that d is pseudo-metric on $\mathcal{F}$, that is, show that d satisfies (P.5.1) and (P.5.2).

Exc. 1.10: Let $\mathcal{F}$ be a paving on the set Ω. Show that if $A \in \sigma(\mathcal{F})$, then there exists a sequence $\{F_n\} \subseteq \mathcal{F}$ such that $A \in \sigma\{F_n \mid n \ge 1\}$.

Hint: Let $\mathcal{A}$ be the paving of all sets $A \subseteq \Omega$ for which there exists $\{F_n\} \subseteq \mathcal{F}$ such that $A \in \sigma\{F_n \mid n \ge 1\}$. Show that $\mathcal{A}$ is σ-algebra containing $\mathcal{F}$.

Exc. 1.11: Let us define $c[\mathcal{H}] := \{A \cup B^c \mid A, B \in \mathcal{H}\}$ whenever $\mathcal{H}$ is a paving on the set Ω. Let $\mathcal{F}$ be a given paving on Ω, and let us define $\mathcal{F}_n$ inductively as follows: $\mathcal{F}_1 = \mathcal{F} \cup \{\emptyset, \Omega\}$ and $\mathcal{F}_{n+1} = c[\mathcal{F}_n]$. Show the following [see also (1.16)]

(i) $\quad \mathcal{F} \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \cdots$

(ii) $\alpha(\mathcal{F}) = \bigcup_{n=1}^{\infty} \mathcal{F}_n$

(iii) If $\mathcal{F}$ countable, then so is $\alpha(\mathcal{F})$

Hint: (i): Note that $\mathcal{H} \subseteq c[\mathcal{H}]$ whenever $\Omega \in \mathcal{H}$. (ii): Let $\mathcal{A} = \cup_{n=1}^{\infty} \mathcal{F}_n$. Show that

$$A \in \mathcal{F}_n \quad \Rightarrow \quad A^c \in \mathcal{F}_{n+1}$$

$$A, B \in \mathcal{F}_n \quad \Rightarrow \quad A \cup B \in \mathcal{F}_{n+2}$$

Conclude that $\mathcal{A}$ is an algebra containing $\mathcal{F}$ and, thus, $\alpha(\mathcal{F}) \subseteq \mathcal{A}$. The converse inclusion may be proved by showing inductively that $\mathcal{F}_n \subseteq \alpha(\mathcal{F})$. (iii): Note that if $\mathcal{H}$ is countable, then so is $c[\mathcal{H}]$.

Exc. 1.12: Let $\mathcal{F}$ be a paving on the set Ω and set $\mathcal{F}_0 = \mathcal{F} \cup \{\emptyset, \Omega\} \cup \{F^c \mid F \in \mathcal{F}\}$. Show that the algebra generated by $\mathcal{F}$ is given by [see also (1.16)]

(i) $$\alpha(\mathcal{F}) = \left\{ \bigcup_{i=1}^{n} \bigcap_{j=1}^{m} A_{ij} \;\middle|\; n, m \in \mathbf{N} \,,\; A_{ij} \in \mathcal{F}_0 \right\}$$

Hint: Let $\mathcal{A}$ denote the paving on the right-hand side of (i). Use (P.2.1) and (P.2.13) to show that $\mathcal{A}$ is an algebra containing $\mathcal{F}$. Conclude that $\alpha(\mathcal{F}) \subseteq \mathcal{A}$. The converse inclusion is easy.

Exc. 1.13: A σ-algebra $\mathcal{F}$ on the set Ω is said to be *separable* (some textbooks uses the terminology *countably generated*) if there exists a countable paving $\mathcal{C}$ such that $\mathcal{F} = \sigma(\mathcal{C})$. Suppose that $\mathcal{F}$ is a separable σ-algebra. Show that there exists a countable algebra $\mathcal{A}$ such that $\mathcal{F} = \sigma(\mathcal{A})$. [Apply Exc.1.11.]

Exc. 1.14: Let (Ω, d) be a separable metric space. Show that the Borel σ-algebra $\mathcal{B}(\Omega)$ is a separable σ-algebra in the sense defined in the previous exercise. [Apply (P.5.5).]

Exc. 1.15: Let Ω be an arbitrary set, and let $\mathcal{C}$ denote the paving of all sets $F \subseteq \Omega$ such that either F or F^c is countable [see (1.12)]. Show that $\mathcal{C}$ is a σ-algebra. Suppose that Ω is uncountable and show that $\mathcal{C}$ is **not** a separable σ-algebra in the sense defined in Exc.1.13

Hint: Suppose that $\mathcal{A}$ is a countable paving with $\mathcal{C} = \sigma(\mathcal{A})$. Then use Exc.1.13 to show that there exist a countable set $J \subseteq \Omega$ such that $\mathcal{C} = \sigma(\{\{\omega\} \mid \omega \in J\}$. Argue that this is impossible if Ω is uncountable.

Exc. 1.16: Let ν and μ be two contents on $(\Omega, \mathcal{F})$ such that $\nu(\Omega) = \mu(\Omega) < \infty$ and $\nu(F) \leq \mu(F)$ for all $F \in \mathcal{F}$. Show that $\nu(F) = \mu(F)$ for all $F \in \mathcal{F}$.

Exc. 1.17: Let ν be a finite content on $(\mathrm{N}, 2^{\mathrm{N}})$, and set $p_k = \nu(\{k\})$ for all $k \in \mathrm{N}$. Show that

(i) $\sum_{k \in A} p_k \leq \nu(A) < \infty \quad \forall A \subseteq \mathrm{N}$

Let μ be the measure on $(\mathrm{N}, 2^{\mathrm{N}})$ given by the left-hand side of (i) [see (1.14)]. Show that $\nu = \mu$ if and only if $\sum_1^\infty p_k = \nu(\mathrm{N})$. [Apply Exc.1.16.] And conclude that ν is a measure if and only if $\sum_1^\infty p_k = \nu(\mathrm{N})$.

Exc. 1.18: Let Ω be an infinite set, and let $\mathcal{C}_0$ be the paving of all $F \subseteq \Omega$ such that either F or F^c is finite. Set

$$P_0(F) = \begin{cases} 0 & \text{if } F \text{ is finite} \\ 1 & \text{if } F^c \text{ is finite} \end{cases}$$

Show that $\mathcal{C}_0$ is an algebra on Ω , and that P_0 is a content on $(\Omega, \mathcal{C}_0)$.

Exc. 1.19: Let (Ω, d) be a metric space and let F be a closed subset of Ω . Show that there exists a sequence of open sets $\{G_n\}$ such that $G_n \downarrow F$. Similarly, show that if G is an open subset of Ω , then there exists a sequence of closed sets $\{F_n\}$ such that $F_n \uparrow G$. Finally, show that the Borel σ-algebra $\mathcal{B}(\Omega)$ is the smallest $(\uparrow, \downarrow)$-stable paving containing all open (respectively all closed sets).

Hint: Let F be closed and consider the continuous function: $d(x, F)$ [see (P.5)]. Set $G_n = \{x \mid d(x, F) < \frac{1}{n}\}$ and apply (P.5.20) and (P.5.22).

Exc. 1.20: Let $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \ldots$ be an increasing sequence of algebras on the set Ω and let $\mathcal{A} = \cup_1^\infty \mathcal{A}_n$. Show that $\mathcal{A}$ is an algebra. Give an example showing that $\mathcal{A}_n$ may be a σ-algebra for all $n \geq 1$, but $\mathcal{A}$ is not a σ-algebra.

Hint: Let $\Omega = \mathrm{N}$, and let $E_n = \{2, 4, \ldots 2n\}$ be the set of all even integers $\leq 2n$. Let $\mathcal{A}_n$ be the paving of all sets $F \subseteq \mathrm{N}$ such that either $F \subseteq E_n$ or $F^c \subseteq E_n$. Show that $\{\mathcal{A}_n\}$ is an increasing sequence of σ-algebras, and note that $E_n \in \mathcal{A}$ for all $n \geq 1$, but $E = \cup_{n=1}^\infty E_n \notin \mathcal{A}$.

Exc. 1.21: A paving $\mathcal{R}$ on the set Ω is called a *ring* respectively a σ-*ring* if $\mathcal{R}$ is $(\setminus, \cup f)$-stable respectively $(\setminus, \cup c)$-stable. The smallest ring respectively σ-ring containing a given paving $\mathcal{F}$ is denoted $\alpha_0(\mathcal{F})$ respectively $\sigma_0(\mathcal{F})$. Show that a ring is $(\setminus, -, \cup f, \cap f, \Sigma f)$-stable and that a σ-ring is $(\setminus, -, \cup f, \cap f, \Sigma f, \cup c, \cap c, \Sigma c, \uparrow, \downarrow)$-stable. Thus, a ring is stable under all the finite set operations except possibly complementation (C), and a σ-ring is stable under all the set operations except possibly complementation (C). Moreover, show that

(i) $\mathcal{R}$ is an algebra (respectively a σ-algebra) if and only if $\Omega \in \mathcal{R}$ and $\mathcal{R}$ is a ring (respectively a σ-ring)

(ii) $\alpha_0(\mathcal{F}) \subseteq \alpha(\mathcal{F})$ with equality if and only if $\Omega \in \alpha_0(\mathcal{F})$

(iii) $\sigma_0(\mathcal{F}) \subseteq \sigma(\mathcal{F})$ with equality if and only if $\Omega \in \sigma_0(\mathcal{F})$

Exc. 1.22: Let $\mathcal{F}$ be paving on the set Ω, and let $\sigma_0(\mathcal{F})$ be the smallest σ-ring containing $\mathcal{F}$ [see Exc.1.21]. Show that for every $A \in \sigma_0(\mathcal{F})$, there exists a sequence $\{F_n\} \subseteq \mathcal{F}$ such that $A \subseteq \cup_{n=1}^{\infty} F_n$.

Hint; Let $\mathcal{A}$ be the paving of all sets $A \subseteq \Omega$ for which there exists a sequence $F_1, F_2, \ldots \in \mathcal{F}$ with $A \subseteq \cup_{n=1}^{\infty} F_n$. Show that $\mathcal{A}$ is a σ-ring containing $\mathcal{F}$.

Exc. 1.23: Let $\mathcal{F}$ be paving on the set Ω, and let $\alpha_0(\mathcal{F})$ be the smallest ring containing $\mathcal{F}$ [see Exc.1.21]. Show that for every $A \in \alpha_0(\mathcal{F})$, there exists finitely many sets $F_1, \ldots, F_m \in \mathcal{F}$ such that $A \subseteq \cup_{n=1}^{m} F_n$.

Hint; Let $\mathcal{A}$ be the paving of all sets $A \subseteq \Omega$ for which there exists finitely many sets $F_1, \ldots, F_m \in \mathcal{F}$ with $A \subseteq \cup_{n=1}^{m} F_n$. Show that $\mathcal{A}$ is a ring containing $\mathcal{F}$.

Exc. 1.24: Let $\mu : 2^{\Omega} \to [0, \infty]$ be a set function, and let $A \subseteq \Omega$ and $M \in \mathcal{M}(\mu)$. Show that $\mu(A) + \mu(M) = \mu(A \cap M) + \mu(A \cup M)$.

Hint: Apply (1.22.1) on A and $A \cup M$.

Exc. 1.25: Let μ be a set function on $(\Omega, \mathcal{F})$, and let $\bar{\mu}$ be the restriction of the outer measure μ^* to $\mathcal{M}(\mu^*)$. Show that $\bar{\mu}$ is a measure on $(\Omega, \mathcal{M}(\mu^*))$. If $\mathcal{F}$ is $(\setminus)$-stable and μ is superadditive on $\mathcal{F}$, then show that $\mathcal{F} \subseteq \mathcal{M}(\mu^*)$. If μ is σ-subadditive on $\mathcal{F}$, then show that μ^* is an extension of μ [cf. (1.28)].

Exc. 1.26: Let $f : \Omega \to \Lambda$ be a function, and let $\mathcal{K}$ and $\mathcal{L}$ be pavings on Ω and Λ, respectively. Suppose that f is surjective and $\mathcal{L}$ is a compact paving; show that $f^{-1}(\mathcal{L})$ is compact. Next suppose that $\mathcal{K}$ is compact and that f is injective; show that the paving $f(\mathcal{K}) := \{f(K) \mid K \in \mathcal{K}\}$ is compact.

Exc. 1.27: Let $\mathcal{K}$ be a $(\cap f)$-stable paving on Ω such that for every decreasing sequence $\{K_n\} \subseteq \mathcal{K}$ with $K_n \downarrow \emptyset$ there exists an integer $p \geq 1$ such that $K_p = \emptyset$. Show that $\mathcal{K}$ is a compact paving.

Exc. 1.28: Let $\mathcal{F}$ be a paving on the set Ω and let Ω_0 be a subset of Ω . Then we define *the trace* of $\mathcal{F}$ on Ω_0 , denoted $\operatorname{tr}(\mathcal{F} \mid \Omega_0)$, as

(*) $\quad \operatorname{tr}(\mathcal{F} \mid \Omega_0) = \{F \cap \Omega_0 \mid F \in \mathcal{F}\}$

Let γ be a list of some of the set operations in the list (1.1.1), and let $\gamma(\mathcal{M})$ denote the smallest γ-stable paving containing $\mathcal{M}$. Show the following:

(i) $\quad \gamma(\operatorname{tr}(\mathcal{F} \mid \Omega_0)) = \operatorname{tr}(\gamma(\mathcal{F}) \mid \Omega_0)$

(ii) If $\mathcal{F}$ is γ-stable, then so is $\operatorname{tr}(\mathcal{F} \mid \Omega_0)$

(iii) If $\mathcal{F}$ is an algebra (respectively a σ-algebra) on Ω , then $\operatorname{tr}(\mathcal{F} \mid \Omega_0)$ is an algebra (respectively a σ-algebra) on Ω_0

Exc. 1.29: Let $\mathcal{F}$ be a paving on the set Ω and let $A \subseteq \Omega$. Set $\mathcal{F}_A := \{B \subseteq \Omega \mid B \cap A \in \operatorname{tr}(\mathcal{F} \mid A)\}$ where $\operatorname{tr}(\mathcal{F} \mid A)$ denotes the trace of $\mathcal{F}$ on A [see Exc.1.28]. Let $f : A \to \Omega$ be *the natural injection*, that is, $f(\omega) = \omega$ for all $\omega \in A$, and let γ be a list of set operations from list (1.1.1). Show the following [apply Exc.1.28]:

(i) $\quad \mathcal{F} \subseteq \mathcal{F}_A$ and $B \in \mathcal{F}_A$ whenever either $B \supseteq A$ or $B \cap A = \emptyset$

(ii) $\quad \gamma(\mathcal{F}_A) = (\gamma(\mathcal{F}))_A$

(iii) If $\mathcal{F}$ is γ-stable, then so is $\mathcal{F}_A$

(iv) If $\mathcal{F}$ is an algebra respectively a σ-algebra, then so is $\mathcal{F}_A$

(v) $\quad \mathcal{F}_A \cap \mathcal{F}_{A^c} = \alpha(\mathcal{F} \cup \{A\})$ if $\mathcal{F}$ is an algebra

(vi) $\quad \mathcal{F}_A \cap \mathcal{F}_{A^c} = \sigma(\mathcal{F} \cup \{A\})$ if $\mathcal{F}$ is an σ-algebra.

Exc. 1.30: Let μ be a set function on $(\Omega, 2^\Omega)$, let $\mathcal{M} = \mathcal{M}(\mu)$ be the paving of all μ-measurable sets, and let $A \subseteq \Omega$. Set $\mu_A(B) := \mu(A \cap B)$ for all $B \subseteq \Omega$. Show that $\mathcal{M}_A \subseteq \mathcal{M}(\mu_A)$, where $\mathcal{M}_A$ is defined as in Exc.1.29.

Exc. 1.31: Let μ be an increasing set function on $(\Omega, \mathcal{F})$, and let us define the following two everywhere defined set functions:

$$\mu^{\triangle}(A) = \sup\{\mu(F) \mid F \in \mathcal{F}\,,\ F \subseteq A\} \qquad \forall A \subseteq \Omega$$

$$\mu^{\triangledown}(A) = \inf\{\mu(F) \mid F \in \mathcal{F}\,,\ F \supseteq A\} \qquad \forall A \subseteq \Omega$$

Let Ω_0 be a given subset of Ω and let $\mathcal{F}_0 = \mathrm{tr}(\mathcal{F} \mid \Omega_0)$ be the trace of $\mathcal{F}$ on Ω_0 [see Exc.1.28]. Show the following:

(i) $\mu^\triangle$ and μ^∇ are increasing extensions of μ such that

$$\mu^\triangle(A) \le \mu_*(A) \ , \ \mu^*(A) \le \mu^\nabla(A) \ , \ \mu^\triangle(A) \le \mu^\nabla(A) \quad \forall A \subseteq \Omega$$

(ii) If μ is superaddtive and $\mathcal{F}$ is (Σf)-stable, then $\mu^\triangle = \mu_*$ and $\mu^\triangle$ is superadditive on 2^Ω

(iii) If μ is superadditive and $\mathcal{F}$ is $(\setminus)$-stable, then μ^∇ is superadditive on $\mathcal{F}_0$

(iv) If μ is subadditive and $\mathcal{F}$ is $(\cup f)$-stable, then μ^∇ is subadditive on 2^Ω

(v) If μ is subadditive and $\mathcal{F}$ is $(\cap f)$-stable, then $\mu^\triangle$ is subadditive on $\mathcal{F}_0$

(vi) If μ is σ-subadditive and $\mathcal{F}$ is $(\cup c)$-stable, then $\mu^\nabla = \mu^*$ and μ^∇ is σ-subadditive on 2^Ω

(vii) If μ is σ-subadditive and $\mathcal{F}$ is $(\cap f)$-stable, then $\mu^\triangle$ is σ-subadditive on $\mathcal{F}_0$

Remark: Some textbooks uses the terms "inner and outer measure" for the set functions $\mu^\triangle$ and μ^∇ . Note that by (ii) we have that $\mu^\triangle = \mu_*$ if μ is superadditive and $\mathcal{F}$ is (Σf)-stable, and by (vi) we have that $\mu^\nabla = \mu^*$ if μ is σ-subadditive and $\mathcal{F}$ is $(\cup c)$-stable.

Hint: (ii)+(vi): Follows from (1.24.1), (1.24.5), (1.24.3) and (1.24.6). (iii): let $G_1, \ldots, G_k \in \mathcal{F}_0$ be disjoint sets, and let $G_0 \in \mathcal{F}_0$ such that $G_0 \supseteq G_1 \cup \ldots \cup G_k$. Choose $F_0, F_1, \ldots, F_k \in \mathcal{F}$ such that $G_j = F_j \cap \Omega_0$, and let $F \in \mathcal{F}$ be given such that $F \supseteq G_0$. Let $D_1, \ldots, D_k$ be the disjungation of $F_1, \ldots, F_k$:

$$D_1 = F_1 \quad , \quad D_j = F_j \setminus \bigcup_{i=1}^{j-1} F_i \quad \text{for} \quad 2 \le j \le k$$

Show that $D_j \in \mathcal{F}$ and $F \cap F_0 \cap D_j \supseteq G_j$ and conclude that

$$\mu(F) \ge \sum_{j=1}^{k} \mu(F \cap F_0 \cap D_j) \ge \sum_{j=1}^{k} \mu^\nabla(G_j)$$

Verify that this proves (iii). The remaining propositions are proved similarly.

Exc. 1.32: Suppose that μ is a content on $(\Omega, \mathcal{F})$, and let $\mu^\triangle$ and μ^∇ be defined as in Exc.1.31. Show the following:

(i) $\mathcal{F} \subseteq \mathcal{M}(\mu^\triangle) \cap \mathcal{M}(\mu^\nabla)$

(ii) $\mu^{\triangle}(B) \leq \mu^{\triangle}(B \cap A) + \mu^{\triangledown}(B \setminus A) \leq \mu^{\triangledown}(B) \quad \forall A, B \subseteq \Omega$

(iii) $\mu^{\triangle}(F \cap A) + \mu^{\triangledown}(F \setminus A) = \mu(F) \quad \forall F \in \mathcal{F}\ \forall A \subseteq \Omega$

(iv) $\mu^{\triangle}(A) + \mu^{\triangle}(B) \leq \mu^{\triangle}(A \cup B) + \mu^{\triangle}(A \cap B) \quad \forall A, B \subseteq \Omega$

(v) $\mu^{\triangledown}(A) + \mu^{\triangledown}(B) \geq \mu^{\triangledown}(A \cup B) + \mu^{\triangledown}(A \cap B) \quad \forall A, B \subseteq \Omega$

(vi) $\mu_* = \mu^{\triangle}$ and if μ is a measure, then $\mu^* = \mu^{\triangledown}$

Hint: (iii): Follows from (ii) and (i) in Exc.1.31. (ii): Let $G, F \in \mathcal{F}$ such that $F \supseteq B \setminus A$ and $G \subseteq B$. Then observe that $G \setminus F \subseteq B \cap A$ and $G \subseteq F \cup (G \setminus F)$. Hence,

$$\mu(G) \leq \mu(G \setminus F) + \mu(F) \leq \mu^{\triangle}(B \cap A) + \mu(F)$$

Verify that this proves the first inequality in (ii). The second inequality is proved similarly.

Exc. 1.33: *(The inner and outer restrictions)* Let μ be an increasing set function on $(\Omega, \mathcal{F})$, and let $\Omega_0 \subseteq \Omega$ be an arbitrary subset of Ω. Let $\mathcal{F}_0 = \mathrm{tr}(\mathcal{F} \mid \Omega_0)$ be the trace of $\mathcal{F}$ on Ω_0 [see Exc.1.28] and let $\mu^{\triangle}$ and $\mu^{\triangledown}$ be defined as in Exc.1.31. Let $\mu_\bullet$ and $\mu^\bullet$ be the restrictions of $\mu^{\triangle}$ and $\mu^{\triangledown}$ to $\mathcal{F}_0$ respectively. Show the following:

(i) If μ is a content (respectively a measure) on $(\Omega, \mathcal{F})$, then $\mu_\bullet$ and $\mu^\bullet$ are contents (respectively measures) on $(\Omega_0, \mathcal{F}_0)$

(ii) If μ is a content on $(\Omega, \mathcal{F})$ and $\mu(\Omega) = \mu^{\triangledown}(\Omega_0) < \infty$, then $\mu^\bullet(F \cap \Omega_0) = \mu(F)$ for all $F \in \mathcal{F}$

The set functions $\mu_\bullet$ and $\mu^\bullet$ are sometimes called *the inner* respectively *the outer restriction of* μ *to* Ω_0.

Hint: (i): Apply Exc.1.31. (ii): Apply Exc.1.32.

Exc. 1.34: Let μ be a content on $(\Omega, \mathcal{F})$ and let $\mu^{\triangle}$ and $\mu^{\triangledown}$ be defined as in Exc.1.31. Let $A \subseteq \Omega$ and let $\mathcal{F} \vee A := \alpha(\mathcal{F} \cup \{A\})$ be the smallest algebra containing A and the algebra $\mathcal{F}$. If $0 \leq t \leq 1$, then we set

$$\mu_t(G) := t\left(\mu^{\triangle}(G \cap A) + \mu^{\triangledown}(G \setminus A)\right) + (1-t)\left(\mu^{\triangle}(G \setminus A) + \mu^{\triangledown}(G \cap A)\right)$$

for all $G \in \mathcal{F} \vee A$. Let $\mathrm{tr}(\mathcal{F} \mid A)$ and $\mathcal{F}_A$ be defined as in Exc.1.28 and Exc.1.29. Show the following:

(i) $\mathcal{F} \vee A = \{G_1 \cup G_2 \mid G_1 \in \mathcal{F} \cap A,\ G_2 \in \mathcal{F} \cap A^c\} = \mathcal{F}_A \cap \mathcal{F}_{A^c}$

(ii) If $\mathcal{F}$ is a σ-algebra, then so is $\mathcal{F} \vee A$

(iii) $\mu_t(A) = t\,\mu^{\triangle}(A) + (1-t)\,\mu^{\triangledown}(A)$

(iv) μ_t is a content on $(\Omega, \mathcal{F} \vee A)$ such that $\mu_t(F) = \mu(F)$ for all $F \in \mathcal{F}$

(v) If μ is a measure on $\mathcal{F}$, then μ_t is a measure on $\mathcal{F} \vee A$

Thus, if μ is a content (respectively a measure) on $(\Omega, \mathcal{F})$, then μ_t is an extension of μ to $\mathcal{F} \vee A$ such that μ_t is a content (respectively a measure) on $\mathcal{F} \vee A$.

Hint: Note that (i) and (ii) follow from Exc.1.29; (iii) is evident; and (iv) and (v) follow from (i), (ii) and Exc.1.33 (i).

Exc. 1.35: *(Extensions of contents and measures)* Let μ be a content (respectively a measure) on $(\Omega, \mathcal{F})$ and let $A_1, \ldots, A_k$ be subsets of Ω . Let $\mathcal{A}$ be the smallest algebra (respectively σ-algebra) containing $\{A_1, \ldots, A_k\}$ and $\mathcal{F}$. Show that μ admits an additive (respectively a σ-additive) extension to $\mathcal{A}$. [Apply Exc.1.34 and induction in k .]

Exc. 1.36: *(Maximal extensions of contents)* Let μ be a content on $(\Omega, \mathcal{F})$ and show that there exists a content $\tilde{\mu}$ on $(\Omega, 2^{\Omega})$ such that $\tilde{\mu}$ is an extension of μ .

Hint: Let $\gimel$ be the set of all pairs $(\kappa, \mathcal{A})$ satisfying

(a) $\mathcal{A}$ is an algebra containing $\mathcal{F}$

(b) κ is a content on $(\Omega, \mathcal{A})$ such that $\kappa(F) = \mu(F)$ for all $F \in \mathcal{F}$

Let us order $\gimel$ as follows:

(c) $(\kappa', \mathcal{A}') \leq (\kappa'', \mathcal{A}'')$ if and only if $\mathcal{A}' \subseteq \mathcal{A}''$ and $\kappa'(A) = \kappa''(A)$ for all $A \in \mathcal{A}'$

Use Zorn's lemma [see (P.12) in the Appendix to Volume II] to show that $\gimel$ has a maximal element, say $(\tilde{\mu}, \tilde{\mathcal{A}})$, and use Exc.1.35 to show that $\tilde{\mathcal{A}} = 2^{\Omega}$. (The symbol $\gimel$ is called gimel, and it is the third letter in the Hebrew alphabet.)

Exc. 1.37: Show that there exists a probability content P on $(\mathbf{N}, 2^{\mathbf{N}})$ such that $P(F) = 0$ for every finite set $F \subseteq \mathbf{N}$. [Apply Exc.1.36 and Exc.1.18.] Show that the content P cannot possibly be a measure.

Exc. 1.38: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $A, B \subseteq \Omega$, and let $F \in \mathcal{F}$. Show the following:

(i) $\mu_*(A) \leq \mu_*(A \cap B) + \mu^*(A \setminus B) \leq \mu^*(A)$

(ii) $\mu_*(A) \leq \mu^*(A \cap B) + \mu_*(A \setminus B) \leq \mu^*(A)$

(iii) $\mu(F) = \mu_*(F \cap B) + \mu^*(F \setminus B) = \mu^*(F \cap B) + \mu_*(F \setminus B)$

(iv) $\exists\, F_0, F_1 \in \mathcal{F} \;:\; F_0 \subseteq A \subseteq F_1 \;,\; \mu(F_0) = \mu_*(A) \;,\; \mu(F_1) = \mu^*(A)$

(v) If $A_n \uparrow A$, then $\mu^*(A_n) \uparrow \mu^*(A)$

(vi) If $B_n \downarrow B$ and $\mu_*(B_1) < \infty$, then $\mu_*(B_n) \downarrow \mu_*(B)$

Hint: (i)–(iii): Let $\mu^\triangle$ and μ^∇ be defined as in Exc.1.31. Note that $\mu_* = \mu^\triangle$ and $\mu^* = \mu^\nabla$, and apply Exc.1.32. (iv): Choose two sequences $\{G_n\} \subseteq \mathcal{F}$ and $\{H_n\} \subseteq \mathcal{F}$ such that $\mu(G_n) \to \mu^\triangle(A)$, $\mu(H_n) \to \mu^\nabla(A)$ and $G_n \subseteq A \subseteq H_n$. Set $F_0 = \cup_1^\infty G_n$ and $F_1 = \cap_1^\infty H_n$. (v)+(vi): Apply (iv), (1.4.2) and (1.4.3).

Exc. 1.39: *(The Lebesgue μ-measurable sets)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $\mathcal{L}(\mu)$ be the paving of all Lebesgue μ-measurable sets [see (1.28)] and let $\tilde{\mu}$ be the restriction of μ^* to $\mathcal{L}(\mu)$. Prove (1.28.4)–(1.28.7). Moreover, show that

(i) $\mathcal{F} \subseteq \mathcal{L}(\mu) \subseteq \mathcal{M}(\mu^*) \cap \mathcal{M}(\mu_*) \quad , \quad \mathcal{N}(\mu) \subseteq \mathcal{L}(\mu)$

Hint: (i): Let $L \in \mathcal{L}(\mu)$ and choose $F_0, F_1 \in \mathcal{F}$ such that $F_0 \subseteq L \subseteq F_1$ and $\mu(F_1 \setminus F_0) = 0$. Note that $F_0, F_1 \in \mathcal{M}(\mu_*)$ and let C be an arbitrary subset of Ω . Show that

$$\begin{aligned}
\mu_*(C \cap L) + \mu_*(C \setminus L) \leq \mu_*(C) &= \mu_*(C \cap F_0) + \mu_*(C \setminus F_0) \\
&= \mu_*(C \cap F_0) + \mu_*(C \cap F_1 \setminus F_0) + \mu_*(C \setminus F_1) \\
&\leq \mu_*(C \cap F_0) + \mu_*(C \setminus F_1) + \mu(F_1 \setminus F_0) \\
&\leq \mu_*(C \cap L) + \mu_*(C \setminus L)
\end{aligned}$$

and conclude that $\mathcal{L}(\mu) \subseteq \mathcal{M}(\mu_*)$. The inclusion $\mathcal{L}(\mu) \subseteq \mathcal{M}(\mu^*)$ is proved similarly. (1.28.5): Apply (iv) in Exc.1.38.

Exc. 1.40: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $\mathcal{H} \subseteq \mathcal{F}$ be paving such that for every $F \in \mathcal{F}$ with $\mu(F) < \infty$ there exists $H \in \mathcal{H}$ with $H \supseteq F$ and $\mu(H) < \infty$. Show the following:

(i) $M \in \mathcal{M}(\mu^*) \quad \Leftrightarrow \quad \mu(H) \geq \mu^*(H \cap M) + \mu^*(H \setminus M) \quad \forall\, H \in \mathcal{H}$

(ii) $\mu^*(M) = \mu_*(M) \quad \forall M \in \mathcal{M}(\mu_*) \cup \mathcal{M}(\mu^*)$ with $\mu^*(M) < \infty$

(iii) $M \in \mathcal{M}(\mu^*) \quad \Leftrightarrow \quad M \cap H \in \mathcal{M}(\mu_*) \quad \forall H \in \mathcal{H}$

(iv) $\mathcal{F} \subseteq \mathcal{M}(\mu_*) \subseteq \mathcal{M}(\mu^*)$

(v) If $\mu(\Omega) < \infty$, then we have

$$M \in \mathcal{M}(\mu^*) \quad \Leftrightarrow \quad \mu(\Omega) \geq \mu^*(M) + \mu^*(M^c)$$

Recall that by (1.28) we have that $\mathcal{M}(\mu^*)$ is a σ-algebra and μ^* is σ-additive on $\mathcal{M}(\mu^*)$. By (1.23) we have that $\mathcal{M}(\mu_*)$ is a algebra and μ_* is additive on $\mathcal{M}(\mu_*)$. But I don't know if $\mathcal{M}(\mu_*)$ is a σ-algebra or if μ_* is σ-additive on $\mathcal{M}(\mu_*)$. In Exc.1.41 you will see that this is often so, and in Exc.1.42 you will see that we may have $\mathcal{M}(\mu^*) = \mathcal{M}(\mu_*)$, but $\mu^*(M) \neq \mu_*(M)$ for some $M \in \mathcal{M}(\mu_*)$.

Hint: (i): Suppose that M satisfies the condition on the right-hand side of the biimplication in (i), and let A be an arbitrary subset of Ω . If $\mu^*(A) = \infty$, then (1.22.1) simply follows from subadditivity of μ^* so suppose that $\mu^*(A) < \infty$. Then show that there exists $H \in \mathcal{H}$ such that $A \subseteq H$. Apply Exc.1.38 (iii) and (1.28.5) to show that $H \cap M \in \mathcal{M}(\mu^*)$, and note that $A \cap M = A \cap (H \cap M)$ and $A \setminus M = A \setminus (H \cap M)$. Show that this implies (1.22.1). The converse implication is evident.

Exc. 1.41: *(Semi-finite measures)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then we say that μ is *finitely founded* if for every $F \in \mathcal{F}$ with $\mu(F) > 0$ there exists $F_0 \in \mathcal{F}$ with $F_0 \subseteq F$ and $0 < \mu(F_0) < \infty$ [see (3.15+16)]. Show the following:

(i) μ finite $\Rightarrow$ μ σ-finite $\Rightarrow$ μ finitely founded

(ii) The Lebesgue measure on $\mathbf{R}^k$ is σ-finite, but not finite

(iii) The counting measure on an arbitrary set is finitely founded; but if the set is uncountable, then the counting measure is not σ-finite

(iv) If μ is finitely founded, then the paving $\mathcal{F}_0 = \{F \in \mathcal{F} \mid \mu(F) < \infty\}$ approximates μ on $\mathcal{F}$ [see (1.22.6)]

(v) If μ is σ-finite, then $\mathcal{L}(\mu) = \mathcal{M}(\mu^*) = \mathcal{M}(\mu_*)$ and $\mu_*(M) = \mu^*(M)$ for all $M \in \mathcal{M}(\mu^*)$, where $\mathcal{L}(\mu)$ is the paving of all Lebesgue μ-measurable sets [see (1.28)]

(vi) If μ is finitely founded, then $\mathcal{M}(\mu^*) = \mathcal{M}(\mu_*)$ and μ_* is σ-additive on $\mathcal{M}(\mu_*)$

(vii) If μ is σ-finite and $\{F_i \mid i \in I\} \subseteq \mathcal{F}$ is a disjoint family of sets, then the set $J = \{i \in I \mid \mu(F_i) > 0\}$ is at most countable

Exc. 1.42: Let Ω be an uncountable set and let $\mathcal{F}$ be σ-algebra on Ω containing all singletons $\{\omega\}$. Suppose that D is a subset of Ω such that $D \notin \mathcal{F}$. Let us define $\mu(F) = \#(F \cap D)$ for all $F \in \mathcal{F}$ and show that

(i) μ is a finitely founded measure on $(\Omega, \mathcal{F})$ [see Exc.1.41]

(ii) $\mu_*(A) = \#(A \cap D) \qquad \forall A \subseteq \Omega$

(iii) $\mu^*(A) = \begin{cases} \#(A \cap D) & \text{if } \exists F \in \mathcal{F} \text{ so that } A \setminus D \subseteq F \subseteq \Omega \setminus D \\ \infty & \text{otherwise} \end{cases}$

(iv) $\mathcal{M}(\mu^*) = \mathcal{M}(\mu_*) = 2^\Omega$ and μ_* and μ^* are both σ-additive on 2^Ω

(v) $\mu_*(D^c) = 0$ and $\mu^*(D^c) = \infty$

Exc. 1.43: Let $\mathcal{F}$ be a paving on the set Ω and let μ be a set function on $(\Omega, \mathcal{F})$ satisfying

(i) $\forall F, G \in \mathcal{F}\ \exists F_0, \ldots, F_k \in \mathcal{F}$ such that $F_j \setminus F_{j-1} \in \mathcal{F}$ for all $1 \le j \le k$ and $F \cap G = F_0 \subseteq F_1 \subseteq \ldots \subseteq F_k = F$

(ii) $\mu(F \cup G) = \mu(F) + \mu(G)$ if $F, G, F \cup G \in \mathcal{F}$ and $F \cap G = \emptyset$

Show that $\mathcal{F}$ is $(\cap f)$-stable and that μ is additive on $\mathcal{F}$.

Exc. 1.44: *(Semialgebras)* A paving $\mathcal{S}$ on the set Ω is called a *semialgebra* if $\mathcal{S}$ is $(\cap f)$-stable, and if for every $S \in \mathcal{S}$ there exists disjoint sets $S_1, \ldots S_k \in \mathcal{S}$ with $\Omega \setminus S = S_1 \cup \ldots \cup S_k$. Let $\mathcal{S}$ be a semialgebra on Ω and let μ be an additive set function on $(\Omega, \mathcal{S})$. Show the following [apply (3.6) and (1.27)]:

(i) $\alpha(\mathcal{S}) = \left\{ \bigcup_{j=1}^{k} S_j \;\middle|\; k \in \mathbf{N}\ ,\ S_1, \ldots, S_k \text{ disjoint } \in \mathcal{S} \right\}$

(ii) μ is inner tight on $\mathcal{S}$ and μ has an additive extension to $(\Omega, \alpha(\mathcal{S}))$

(iii) If μ is σ-additive on $\mathcal{S}$, then μ is outer tight on $\mathcal{S}$ and μ has a σ-additive extension to $(\Omega, \sigma(\mathcal{S}))$

Exc. 1.45: *(σ-smoothness)* Let $\mathcal{F}$ be a paving on the set Ω and let μ be a set function on $(\Omega, \mathcal{F})$. If $F \in \mathcal{F}$ and $\mathcal{G} \subseteq \mathcal{F}$ is a subpaving of $\mathcal{F}$, then we say that μ is *σ-smooth at F along $\mathcal{G}$* if $\mu(G_n) \to \mu(F)$ as $n \to \infty$ whenever $\{G_n\} \subseteq \mathcal{G}$ and $G_n \downarrow F$. If μ is σ-smooth at G along $\mathcal{G}$ for all $G \in \mathcal{G}$, then we say that μ is *σ-smooth on $\mathcal{G}$*. Let $\mathcal{G} \subseteq \mathcal{F}$ such that $\mu(G) < \infty$ for all $G \in \mathcal{G}$. Show the following:

(i) If $\mathcal{G}$ is $(-)$-stable and μ is σ-additive on $\mathcal{G}$, then μ is σ-smooth on $\mathcal{G}$

(ii) If $\mathcal{F}$ is $(\setminus)$-stable and μ is subadditive on $\mathcal{F}$ and σ-smooth at $\emptyset$ along $\mathcal{F}$, then μ is σ-subadditive on $\mathcal{F}$

Exc. 1.46: Let Ω be the set of all rational numbers in the unit interval $[0,1]$. Let $\mathcal{F}$ be the paving of all sets of the form $J \cap \Omega$, where J is an interval contained in $[0,1]$. Let μ be given by $\mu(J \cap \Omega) = b - a$ whenever $J \subseteq [0,1]$ is an interval with left end point a and right end point b. Show that

(i) $\mathcal{F}$ is a semialgebra [see Exc.1.44]

(ii) μ is additive, inner tight and σ-smooth on $\mathcal{F}$ [see Exc.1.45]

(iii) μ has no σ-additive extension to $\sigma(\mathcal{F})$

Why does this not contradict (3) in (1.27) or (iii) in Exc.1.44 ?

Hint: Apply Exc.1.44 to prove inner tightness of μ.

Exc. 1.47: Let μ be a measure on $(\Omega, \mathcal{F})$, let $d(F,G)$ be the pseudo-metric on $\mathcal{F}$ from Exc.1.9, and let $\mathcal{F}_0 = \{F \in \mathcal{F} \mid \mu(F) < \infty\}$. Show the following:

(i) $\lim_{n\to\infty} d(F_n, F) = 0$ if $\{F_n\} \subseteq \mathcal{F}$, $F_n \uparrow F$ and $\mu(F) < \infty$

(ii) $\lim_{n\to\infty} d(F_n, F) = 0$ if $\{F_n\} \subseteq \mathcal{F}$, $F_n \downarrow F$ and $\mu(F_1) < \infty$

(iii) $\mathcal{F}_0$ is both open and closed in $(\mathcal{F}, d)$

(iv) μ is a continuous function from $(\mathcal{F}, d)$ into $[0,\infty]$

(v) If $B \in \mathcal{F}$, then the functions $F \curvearrowright B \cap F$, $F \curvearrowright B \setminus F$ and $F \curvearrowright B \cup F$ are continuous from $(\mathcal{F}, d)$ into itself

(vi) If ν is a set function on $(\Omega, \mathcal{F})$, then ν is continuous with respect to the metric d at $\emptyset$ if and only if ν is μ-continuous [see (3.18)]

Let $\mathcal{H} \subseteq \mathcal{F}$ and let $\bar{\mathcal{H}}$ denote the closure of $\mathcal{H}$ in the pseudo-metric space $(\mathcal{F}, d)$. Suppose that $\{H_n\} \subseteq \bar{\mathcal{H}}$ is a sequence in $\bar{\mathcal{H}}$, and show the following:

(vii) If $H_n \uparrow F$ and $\mu(F) < \infty$, then $F \in \bar{\mathcal{H}}$

(viii) If $H_n \downarrow F$ and $\mu(H_1) < \infty$, then $F \in \bar{\mathcal{H}}$

(ix) If $\mathcal{H}$ is either $(\cap f)$-stable, or $(\cup f)$-stable, or $(\setminus)$-stable, then so is $\bar{\mathcal{H}}$

(x) If $\mathcal{H}$ is a ring [see Exc.1.21] and $\mathcal{H} \subseteq \mathcal{F}_0 \subseteq \sigma_0(\mathcal{H})$, then $\bar{\mathcal{H}} = \mathcal{F}_0$

Hint: (iii)+(iv): Note that $\mu(F) \leq d(F, G) + \mu(G)$ for all $F, G \in \mathcal{F}$. (ix): Apply Sierpińsky's stability lemma (1.6). (x): Let $H \in \mathcal{H}$ be fixed. Show that $\mathrm{tr}(\mathcal{H} \mid H)$ [see Exc.1.28] is an algebra on H and that $\mathrm{tr}(\bar{\mathcal{H}} \mid H)$ is a $(\downarrow, \uparrow)$-stable paving containing $\mathrm{tr}(\mathcal{H} \mid H)$. Conclude that $\mathrm{tr}(\mathcal{F}_0 \mid H) \subseteq \mathrm{tr}(\bar{\mathcal{H}} \mid H)$, and apply (vii) and Exc.1.23 to show that this implies $\mathcal{F}_0 \subseteq \bar{\mathcal{H}}$. The converse inclusion follows from (iii).

Exc. 1.48: Let μ and ν be two finite measures on $(\Omega, \mathcal{F})$ and let $\mathcal{G}$ be the paving of all sets $G \in \mathcal{F}$ satisfying $\mu(G) \leq \nu(G)$. Show that $\mathcal{G}$ is $(\uparrow, \downarrow, \sum c)$-stable. Now suppose that $\mathcal{A}$ is an algebra contained in $\mathcal{F}$ such that $\mu(A) \leq \nu(A)$ for all $A \in \mathcal{A}$ and show that

(i) $\mu(A) \leq \nu(A) \qquad \forall A \in \sigma(\mathcal{A})$

(ii) If $B \in \sigma(\mathcal{A})$ and $\mu(B) = \nu(B)$, then $\mu(G) = \nu(G)$ for all $G \in \sigma(\mathcal{A})$ with $G \subseteq B$

Hint: (ii): Apply (i) on G and $B \setminus G$.

Exc. 1.49: Let (S, d) be a metric space, and let μ and ν be two finite Borel measures on S such that $\mu(G) \leq \nu(G)$ for all open subsets G of S. Show that $\mu(B) \leq \nu(B)$ for all $B \in \mathcal{B}(S)$. Show that this result remains valid if we replace the open sets by the closed sets. [Apply Exc.1.48 and Exc.1.19.]

Exc. 1.50: Let $\mu = \#$ be the counting measure on $(\mathbf{Q}, 2^{\mathbf{Q}})$, where $\mathbf{Q}$ denotes the set of all rational numbers. Let $\nu = 2\mu$ and let $\mathcal{A}$ be the algebra generated by all sets of the form $]a, b] \cap \mathbf{Q}$. Show that

(i) $\mu(G) = \nu(G) \qquad \forall G \text{ open } \subseteq \mathbf{Q}$

(ii) $\mu(A) = \nu(A) \qquad \forall A \in \mathcal{A}$

Observe that $\mu \neq \nu$. Why doesn't this contradicts (1.7) or Exc.1.48 or Exc.1.49 ?

Exc. 1.51: *(The Cantor set)* The *Cantor set*, denoted $\mathbb{C}$, is defined to be the set of all real numbers $x \in [0,1]$ which admits an expansion of the form

$$x = \sum_{j=1}^{\infty} \varepsilon_j 3^{-j} \quad \text{where} \quad \varepsilon_j = 0 \text{ or } 2 \quad \forall j \geq 1$$

Show that the Cantor set $\mathbb{C}$ is compact and uncountable, and yet $\mathbb{C}$ is a λ-null set, where λ is the Lebesgue measure on $\mathbf{R}$

Hint: Let $\xi = (\xi_1, \ldots, \xi_n) \in \{0,2\}^n$ and set $a_n(\xi) = \xi_1 3^{-1} + \cdots + \xi_n 3^{-n}$ and

$$I_n(\xi) = \left[a_n(\xi), 3^{-n} + a(\xi)\right] \quad , \quad C_n = \bigcup_{\xi \in \{0,2\}^n} I_n(\xi)$$

for all $n \geq 1$. Show that C_n is compact, $\lambda(C_n) = (\frac{2}{3})^n$ and $C_n \downarrow \mathbb{C}$.

Exc. 1.52: Let λ_k denote the Lebesgue measure on $\mathbf{R}^k$ and let $A, B \in \mathcal{M}(\lambda_k)$ such that $\lambda_k(B) < \infty$. Show

(i) $f(x) := \lambda_k(A \cap (B + x))$ is continuous from $\mathbf{R}^k$ into $\mathbf{R}$

(ii) If $\lambda_k(B) > 0$, then $B - B$ contains an open neighborhood of 0

where $B + x = \{y + x \mid y \in B\}$ and $B - B = \{u - v \mid u, v \in B\}$.

Hint: (i): Let $\varepsilon > 0$ be given. Then by (1.34) there exists a compact set K and an open set G such that $K \subseteq B \subseteq G$ and $\lambda_k(G \setminus K) < \varepsilon$. Choose $\delta > 0$ such that $v + K \subseteq G$ for all $||v|| < \delta$. Suppose that $||x - y|| < \delta$. Then show that

$$(B + x) \triangle (B + y) \subseteq \{x + (G \setminus K)\} \cup \{y + (G \setminus K)\}$$

Conclude that the function $x \curvearrowright B + x$ is continuous from $\mathbf{R}^k$ into $(\mathcal{M}(\lambda_k), d)$, where d is the pseudo-metric on $\mathcal{M}(\lambda_k)$ defined in Exc.1.9. Then apply Exc.1.47. (ii): Set $g(x) = \lambda_k(B \cap (B + x))$ and show that $g(0) > 0$ and $\{g > 0\} \subseteq B - B$.

Exc. 1.53: Let $\mathbf{Q}$ denote the set of all rational numbers, and let us define an equivalence relation on $\mathbf{R}$ as follows:

$$x \equiv y \,(\text{mod}\, \mathbf{Q}) \quad \Leftrightarrow \quad x - y \in \mathbf{Q}$$

Show that $x + \mathbf{Q} = \{x + q \mid q \in \mathbf{Q}\}$ is the equivalence class containing x . Now let us pick exactly one point from each of the equivalence classes, and let M be the set of all the chosen points. Show that M is not Lebesgue-measurable, that is, if λ denotes the Lebesgue measure on $\mathbf{R}$, then you shall show that $M \notin \mathcal{M}(\lambda)$ [see (1.28.2)].

Hint: Suppose that M is Lebesgue-measurable, and let $\mathbf{Q} = \{q_1, q_2, \ldots\}$ be an enumeration of the rational numbers. Show that $(M - M) \cap \mathbf{Q} = \emptyset$ and $\mathbf{R} = (q_1 + M) \cup (q_2 + M) \cup \cdots$. Then apply Exc.1.52 to show $\lambda(M) = 0$, and argue that this impossible.

Exc. 1.54: *(Vector measures)* Let $\mathcal{F}$ be a σ-algebra on the set Ω. Then a function $\mu : \mathcal{F} \to \mathbf{R}^k$ is called a *k-dimensional vector measure* if

$$\mu\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \mu(F_n) \quad \forall \{F_n\} \text{ disjoint } \subseteq \mathcal{F}$$

in the sense that the infinite series converges to the vector $\mu(\cup_1^\infty F_n)$. If μ is a 1-dimensional vector measure, then μ is called a *signed measure*. Let μ be a k-dimensional vector measure and let $\nu : \mathcal{F} \to \mathbf{R}^k$ be a given function with coordinate functions $\nu_1, \ldots, \nu_k$. Show the following:

(i) $\mu(\emptyset) = 0$ and $\mu(F \setminus G) = \mu(F) - \mu(F \cap G) \quad \forall F, G \in \mathcal{F}$

(ii) ν is a vector measure if and only if $\nu_1, \ldots, \nu_k$ are signed measures

(iii) If $F_1, F_2 \ldots \in \mathcal{F}$ are disjoint, then $\sum_{n=1}^{\infty} ||\mu(F_n)|| < \infty$

(iv) Any finite linear combination of vector measures is a vector measure

(v) If $D \in \mathcal{F}$, then $\mu_D(F) := \mu(D \cap F)$ is a vector measure

(vi) If $\{F_n\} \subseteq \mathcal{F}$ and either $F_n \uparrow F$ or $F_n \downarrow F$, then $\mu(F_n) \to \mu(F)$

Remark: Suppose that μ and $\nu_1, \nu_2, \ldots$ are k-dimensional vector measures. Then it can be shown that we have the following;

(vii) If $\mathcal{R}$ is a σ-ring [see Exc.1.21], then the set $\{\mu(A) \mid A \in \mathcal{R}\}$ is a compact subset of $\mathbf{R}^k$

(viii) If the limit $\nu(F) = \lim_{n\to\infty} \nu_n(F)$ exists for all $F \in \mathcal{F}$, then ν is a k-dimensional vector measure

These propositions are very useful, but also difficult to prove.

Hint: (iii): First suppose that $k = 1$, and apply the well-known result from analysis: If every rearrangement of a series $\sum a_n$ converges to a finite value, then the series is absolutely convergent, that is, $\sum |a_n| < \infty$. Then apply (ii).

Exc. 1.55: *(The total variation)* Let μ be a k-dimensional vector measure on $(\Omega, \mathcal{F})$ [see Exc.1.54]. Then $\nu(F) = ||\mu(F)||$ is a set function on $\mathcal{F}$, and the inner measure ν_* is called *the total variation* of μ, and is denoted $|\mu| := \nu_*$. Show that

(i) $|\mu|(F) \le 2k \sup\left\{||\mu(G)|| \;\middle|\; G \in \mathcal{F} \text{ and } G \subseteq F\right\} < \infty \quad \forall F \in \mathcal{F}$

(ii) $|\mu|$ is σ-additive on $\mathcal{F}$

Hint: First note that the finiteness of the right-hand side of (i) follows from (vii) in Exc.1.54. Second show that (i) holds if $k = 1$, and then use (ii) in Exc.1.54 to prove (i) in the general case.

Exc. 1.56: *(The decomposition theorem)* Let μ be a k-dimensional vector measure on $(\Omega, \mathcal{F})$ [see Exc.1.54] and let $\mathcal{R}$ be a σ-ring [see Exc.1.21] such that $\mathcal{R} \subseteq \mathcal{F}$. We say that μ is *$\mathcal{R}$-regular* if $|\mu|(A) = 0$ for all $A \in \mathcal{R}$, where $|\mu|$ is the total variation of μ [see Exc.1.55]. We say that μ is *$\mathcal{R}$-singular* if there exists a set $B \in \mathcal{R}$ such that $|\mu|(\Omega \setminus B) = 0$. Show the following:

(i) If μ is both $\mathcal{R}$-regular and $\mathcal{R}$-singular, then $\mu \equiv 0$

(ii) A finite linear combination of $\mathcal{R}$-regular respectively $\mathcal{R}$-singular vector measures is $\mathcal{R}$-regular respectively $\mathcal{R}$-singular

(iii) If μ_0, ν_0 are $\mathcal{R}$-regular vector measures and μ_1, ν_1 are $\mathcal{R}$-singular vector measures such that $\mu_0 + \nu_1 = \nu_0 + \mu_1$, then $\mu_0 = \nu_0$ and $\mu_1 = \nu_1$

Show that there exists a set $S \in \mathcal{R}$ satisfying

(iv) $\mu_s(F) := \mu(F \cap S)$ for $F \in \mathcal{F}$ is an $\mathcal{R}$-singular vector measure, and $\mu_r(F \cap S) := \mu(F \setminus S)$ for $F \in \mathcal{F}$ is an $\mathcal{R}$-regular vector measure

(v) $\mu = \mu_r + \mu_s$

Thus, by (iii)–(v) we see that any vector measure admits a unique decomposition into a sum of an $\mathcal{R}$-regular vector measure and an $\mathcal{R}$-singular vector measure. The decomposition (v) is called *the $\mathcal{R}$-decomposition* of μ , and the decomposition $(S, \Omega \setminus S)$ is called *the $(\mu, \mathcal{R})$-decomposition* of Ω .

Hint: (iv)+(v): Choose $S_n \in \mathcal{R}$ such that $|\mu|(S_n) \to \alpha$ where $\alpha = \sup_{R \in \mathcal{R}} |\mu|(R)$, and set $S = \cup_1^\infty S_n$.

Exc. 1.57: *(The Jordan-Hahn decomposition)* Let μ be a signed measure [see Exc.1.54] on $(\Omega, \mathcal{F})$. Set

$$\mu^+(F) = \tfrac{1}{2}\left(|\mu|(F) + \mu(F)\right) \quad , \quad \mu^-(F) = \tfrac{1}{2}\left(|\mu|(F) - \mu(F)\right)$$

for all $F \in \mathcal{F}$, where $|\mu|$ is the total variation of μ [see Exc.1.55]. Show that

(i) μ^+ and μ^- are finite measures on $(\Omega, \mathcal{F})$

(ii) $|\mu|(F) = \mu^+(F) + \mu^-(F)$ and $\mu(F) = \mu^+(F) - \mu^-(F) \quad \forall F \in \mathcal{F}$

Thus, any signed measure can be decomposed into a difference of two positive finite measures. The decomposition (μ^+, μ^-) is called *the Jordan decomposition* of μ. Let $S \in \mathcal{F}$ be chosen such that $\mu(S) \geq \mu(F)$ for all $F \in \mathcal{F}$ [this is possible by (vii) in Exc.1.54]. Show that

(iii) $\mu^+(F) = \mu(F \cap S)$ and $\mu^-(F) = \mu(F \setminus S) \quad \forall F \in \mathcal{F}$

The decomposition $(S, \Omega \setminus S)$ is called *the Hahn decomposition* of Ω.

Hint: (iii): First observe that $\mu(S) = \mu(S \cap F) + \mu(S \setminus F) \geq \mu(S \setminus F)$ and conclude that $\mu(S \cap F) \geq 0$ for all $F \in \mathcal{F}$. Next observe that $\mu(F \setminus S) + \mu(S) = \mu(S \cup F) \leq \mu(S)$ and conclude that $\mu(F \setminus S) \leq 0$ for all $F \in \mathcal{F}$. Use these two facts to show $|\mu|(F) = \mu(F \cap S) - \mu(F \setminus S)$, and note that $\mu(F) = \mu(F \cap S) + \mu(F \setminus S)$. Argue that the latter two equalities proves (iii).

Exc. 1.58: *(Atoms)* Let μ be a k-dimensional vector measure on $(\Omega, \mathcal{F})$ [see Exc.1.54]. If $A \in \mathcal{F}$, then we say that A is a *μ-atom* if for every $F \in \mathcal{F}$ with $F \subseteq A$ we have that either $\mu(F) = 0$ or $\mu(F) = \mu(A)$; that is, a μ-atom cannot be divided into two disjoint set of non-zero measure. A μ-atom A is called a *positive μ-atom* if $|\mu|(A) > 0$, where $|\mu|$ is the total variation of μ [see Exc.1.55]. Let $\mu_1, \ldots, \mu_k$ be the coordinates functions of μ and let $A \in \mathcal{F}$ be given. Show the following:

(i) If $|\mu|(A) = 0$, then A is a μ-atom

(ii) A is a μ-atom if and only if A is a $|\mu|$-atom, and if and only if A is a μ_j-atom for all $j = 1, \ldots, k$

(iii) If A is a positive μ_j-atom for some $1 \leq j \leq k$, then A contains a positive μ-atom

(iv) If A is a μ-atom, then $F \cap A$ is a μ-atom for all $F \in \mathcal{F}$

Exc. 1.59: *(The atomic decomposition)* Let μ be a k-dimensional vector measure on $(\Omega, \mathcal{F})$ [see Exc.1.54]. Then we say that μ is *atom-free* if μ admits no positive atoms, that is, if $|\mu|(A) = 0$ for every μ-atom A [see Exc.1.58]. We say that μ is *atomic* if Ω is a countable union of μ-atoms. Let $\mathcal{R}$ denote the paving of all countable unions of μ-atoms. Show that $\mathcal{R}$ is a σ-ring and that

(i) μ is atom-free if and only if μ is $\mathcal{R}$-regular [see Exc.1.56]

(ii) μ is atomic if and only if μ is $\mathcal{R}$-singular [see Exc.1.56]

(iii) Any finite linear combination of atom-free respectively atomic vector measures is atom-free respectively atomic

And deduce from Exc.1.56, that there exists a unique pair (μ_a, μ_c) satisfying

(iv) μ_a is an atomic, μ_c is an atom-free and $\mu = \mu_a + \mu_c$

The decomposition (μ_a, μ_c) is called *the atomic decomposition* of μ . Show that there exists $S \in \mathcal{R}$ satisfying

(v) $\mu_a(F) = \mu(F \cap S)$ and $\mu_c(F) = \mu(F \setminus S) \quad \forall F \in \mathcal{F}$

The decomposition $(S, \Omega \setminus S)$ is called *the atomic decomposition* of Ω .

Remark: The so-called *Lyapounov's theorem* states that [cf. (vii) in Exc.1.54]

(vi) If μ is a k-dimensional atom-free vector measure, then the set $\{\mu(F) \mid F \in \mathcal{F}\}$ is a compact convex subset of $\mathbf{R}^k$

This theorem has significant applications to statistics and mathematical economy, but it is difficult to prove.

Exc. 1.60: *(The point spectral decomposition)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space such that $\{\omega\} \in \mathcal{F}$ for all $\omega \in \Omega$. If $\mu(\{\omega\}) = 0$ for all $\omega \in \Omega$, then we say that μ is *diffuse*, and if

$$\mu(A) = \sum_{\omega \in A} \mu(\{\omega\}) \quad \forall A \in \mathcal{F}$$

then we say that μ is *discrete* [cf. (4.14)]. Let μ be an arbitrary measure. Then we define the *point spectrum*, denoted $sp_+(\mu)$, of μ in the following way:

$$sp_+(\mu) := \{\omega \in \Omega \mid \mu(\{\omega\}) > 0\}$$

Show the following:

(i) $\mu(A) = \sum_{\omega \in A} \mu(\{\omega\}) \quad \forall A \in \mathcal{F}$ with $A \subseteq sp_+(\mu)$

(ii) If μ is σ-finite, then $sp_+(\mu)$ is at most countable

(iii) μ is diffuse if and only if $sp_+(\mu) = \emptyset$

(iv) If μ is discrete, then $\mu_*(\Omega \setminus sp_+(\mu)) = 0$

(v) If $\mu^*(\Omega \setminus sp_+(\mu)) = 0$, then μ is discrete

By (ii) we have that $sp_+(\mu) \in \mathcal{F}$ whenever μ is σ-finite. But if μ is not σ-finite, this need not be so. For instance, if μ is the measure defined in Exc.3.29, then $sp_+(\mu) = D \notin \mathcal{F}$, and we have $\mu_*(\Omega \setminus sp_+(\mu)) = 0$ and $\mu^*(\Omega \setminus sp_+(\mu)) = \infty$. Now let us set

$$\mu_0(A) = \mu_*(A \cap sp_+(\mu)) \quad \text{and} \quad \mu_1(A) = \mu^*(A \setminus sp_+(\mu)) \quad \forall A \in \mathcal{F}$$

Show that

(vi) μ_0 and μ_1 are measures on $(\Omega, \mathcal{F})$

(vii) μ_0 is discrete and μ_1 is diffuse

(viii) $\mu = \mu_0 + \mu_1$

Thus, any measure can be decomposed as a sum of a discrete and a diffuse measure. The decomposition (vii) is called the *point spectral decomposition* of μ.

Hint: Apply Exc.1.31 and Exc.1.38 to prove (vi) and (viii).

Exc. 1.61: *(The support of a measure)* Let Ω be a metric space and let μ be a Borel measure on Ω. Let $\{G_i \mid i \in I\}$ be collection of open subsets of Ω. Suppose that μ is τ-smooth [see the Remark to (1.34)] and apply (P.5.5) to show that

(i) $$\mu\left(\bigcup_{i\in I} G_i\right) = \sup_{\gamma\in\Gamma} \mu\left(\bigcup_{i\in\gamma} G_i\right)$$

where Γ is the set of all non-empty finite subsets of I. Let us say that a Borel measure μ is *τ^*-smooth* if (i) holds for every collection $\{G_i \mid i \in I\}$ of open subsets of Ω. Thus, we have that every τ-smooth measure is τ^*-smooth. Now suppose that μ is an arbitrary Borel measure on Ω. Show the following:

(ii) If μ is finite, then μ is τ-smooth if and only if μ is τ^*-smooth

(iii) If μ is a Radon measure [see the Remark to (1.34)], then μ is τ^*-smooth

If μ is a Borel measure, then we define *the support* (also called *the spectrum*) of μ, denoted $sp(\mu)$, as

$$sp(\mu) := \{\omega \in \Omega \mid \mu(b(\omega, r)) > 0 \quad \forall r > 0\}$$

Let μ be an arbitrary Borel measure on Ω. Show the following:

(iv) $sp(\mu)$ is a closed subset of Ω

(v) If F is a closed subset of Ω such that $\mu(\Omega \setminus F) = 0$, then $F \supseteq sp(\mu)$

(vi) If μ is τ^*-smooth, then $\mu(\Omega \setminus sp(\mu)) = 0$

(vii) $sp_+(\mu) \subseteq sp(\mu)$

where $sp_+(\mu)$ denotes the point spectrum of μ [see Exc.1.60]. Thus, if μ is τ^*-smooth (in particular, if μ is τ-smooth, or if μ is a Radon measure), then the support is the smallest closed subset of Ω , whose complement has μ-measure 0 . The difference set $sp(\mu) \setminus sp_+(\mu)$ is called *the continuous spectrum* of μ and is denoted $sp_c(\mu) := sp(\mu) \setminus sp_+(\mu)$

Hint: (ii): Suppose that μ is finite and τ^*-smooth. Applying (i) with $I = \Omega$ and $G_\omega = b(\omega, 2^{-n})$, where n is a fixed integer, we can find a countable set $\gamma_n \subseteq \Omega$ satisfying $\mu(\Omega) = \mu(\Gamma_n)$ where

$$\Gamma_n = \bigcup_{\omega \in \gamma_n} b(\omega, 2^{-n})$$

Set $\Omega_0 = \cap_1^\infty \Gamma_n$ and show that Ω_0 is separable and $\mu(\Omega \setminus \Omega_0) = 0$. Argue that this shows the "if" part. (vi): Note that for every $\omega \notin sp(\mu)$ there exists $r(\omega) > 0$ with $\mu(b(\omega, r(\omega))) = 0$, and apply (i) with $I = \Omega \setminus sp(\mu)$ and $G_\omega = b(\omega, r(\omega))$.

Exc. 1.62: *(Lattice infimum and supremum)* Let $\beth$ denote a set of measures on the measurable space $(\Omega, \mathcal{F})$, and let us define the set functions ν and ξ as follows

$$\nu(F) = \inf \left\{ \sum_{j=1}^{\infty} \mu_j(F_j) \;\middle|\; \begin{array}{l} F_1, F_2, \ldots \text{ disjoint } \in \mathcal{F} \\ \mu_1, \mu_2, \ldots \in \beth \;,\; F \supseteq \bigcup_{n=1}^{\infty} F_n \end{array} \right\}$$

$$\xi(F) = \sup \left\{ \sum_{j=1}^{\infty} \mu_j(F_j) \;\middle|\; \begin{array}{l} F_1, F_2, \ldots \in \mathcal{F} \text{ and} \\ \mu_1, \mu_2, \ldots \in \beth \;,\; F \subseteq \bigcup_{n=1}^{\infty} F_n \end{array} \right\}$$

for all $F \in \mathcal{F}$. Show the following:

(i) ν and ξ are measures on $(\Omega, \mathcal{F})$

(ii) $\xi(F) \leq \mu(F) \leq \nu(F) \qquad \forall F \in \mathcal{F}\ \forall \mu \in \beth$

(iii) If κ is a measure on $(\Omega, \mathcal{F})$ such that $\kappa \geq \mu$ for all $\mu \in \beth$, then $\kappa \geq \nu$

(iv) If κ is a measure on $(\Omega, \mathcal{F})$ such that $\kappa \leq \mu$ for all $\mu \in \beth$, then $\kappa \leq \xi$

Thus, ξ is the largest measure on $(\Omega, \mathcal{F})$ which is dominated by all the measures in $\beth$. The measure ξ is called *the lattice infimum* of $\beth$ and we use the notation $\xi = \wedge\beth$. Similarly, ν is the smallest measure on $(\Omega, \mathcal{F})$ which dominates all the measures in $\beth$. The measure ν is called *the lattice supremum* of $\beth$ and we use the notation $\nu = \vee\beth$. (The symbol $\beth$ is called beth, and it is the second letter in the Hebrew alphabet).

Exc. 1.63: *(Lattice difference)* Let μ and ν be two measures on the measurable space $(\Omega, \mathcal{F})$ such that $\mu(F) \leq \nu(F)$ for all $F \in \mathcal{F}$. Let us define $\mathcal{G} = \{F \in \mathcal{F} \mid \mu(F) < \infty\}$ and

$$\rho(F) = \sup\{\nu(G) - \mu(G) \mid G \in \mathcal{F} \text{ and } G \subseteq F\} \quad \forall F \in \mathcal{F}$$

Show that ρ is a measure on $(\Omega, \mathcal{F})$ satisfying

(i) $\nu = \mu + \rho$

(ii) $\rho(G) = \nu(G) - \mu(G) \quad \forall G \in \mathcal{G}$

Note that if $G \notin \mathcal{G}$, then the right-hand side of (ii) takes the form $\infty - \infty$, which is not defined. However, (i) shows that ρ is a sort of difference of ν and μ, and ρ is called *the lattice difference* of ν and μ, and we use the notation $\rho = \nu \ominus \mu$.

Exc. 1.64: Let $\mathcal{F}$ be a paving on the set Ω and let μ be a set function on $(\Omega, \mathcal{F})$. Show that $\mathcal{G}$ is an approximating paving for μ on $\mathcal{F}$ if and only if for every $F \in \mathcal{F}$ and every $a < \mu(F)$, there exists $G \in \mathcal{G}$ and finitely many disjoint sets $F_1, \ldots, F_n \in \mathcal{F}$ satisfying

(i) $a < \sum_{j=1}^{n} \mu(F_j)$ and $\bigcup_{j=1}^{n} F_j \subseteq G \subseteq F$

Exc. 1.65: Let $\mathcal{F}$ and $\mathcal{G}$ be pavings on the set Ω such that $\mathcal{G} \subseteq \mathcal{F}$ and let μ and ν be set functions on $\mathcal{F}$ respectively $\mathcal{G}$ such that $\mu(G) = \nu(G)$ for all $G \in \mathcal{G}$. Show that

(i) $\nu_*(A) \leq \mu_*(A)$ and $\mu^*(A) \leq \nu^*(A) \quad \forall A \subseteq \Omega$

Exc. 1.66: Let $\mathcal{K}$ be a $(\cap f)$-stable paving on the set Ω such that $\emptyset \in \mathcal{K}$. Set

$$\mathcal{F}_{\mathcal{K}} = \{F \subseteq \Omega | K \cap F \in \mathcal{K} \quad \forall K \in \mathcal{K}\}, \quad \mathcal{G}_{\mathcal{K}} = \{G \subseteq \Omega | K \setminus G \in \mathcal{K} \quad \forall K \in \mathcal{K}\}$$

Show the following:

(i) $\mathcal{F}_{\mathcal{K}}$ is a $(\cap f)$-stable paving, and $\mathcal{K} \cup \{\emptyset, \Omega\} \subseteq \mathcal{F}_{\mathcal{K}}$

(ii) $\mathcal{G}_{\mathcal{K}}$ is a $(\cup f)$-stable paving

(iii) $G \in \mathcal{G}_{\mathcal{K}} \quad \Leftrightarrow \quad G^c \in \mathcal{F}_{\mathcal{K}}$

(iv) $\mathcal{K}$ is a compact paving if and only if for every sequence $\{G_n\} \subseteq \mathcal{G}_{\mathcal{K}}$ and every set $K \in \mathcal{K}$ with $K \subseteq \cup_1^\infty G_n$, there exists an integer $m \geq 1$ such that $K \subseteq G_1 \cup \cdots \cup G_m$

(v) If Ω is a metric space and $\mathcal{K}$ is the paving of all compact subsets of Ω , then $\mathcal{F}_{\mathcal{K}}$ is the paving of all closed subsets of Ω , and $\mathcal{G}_{\mathcal{K}}$ is the paving of all open subsets of Ω

Exc. 1.67: Let $\mu : 2^\Omega \to [0, \infty]$ be a set function on Ω , and let $B \subseteq C \subseteq \Omega$ be two sets such that $C \setminus B \in \mathcal{M}(\mu)$. Show that $\mu(A \cap B) \leq \mu(A \cap C)$ for all $A \subseteq \Omega$.

Exc. 1.68: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f be a function from Ω into the measurable space $(M, \mathcal{A})$. Suppose that $\mathcal{A}$ is separable [see Exc.1.13] and let $\mathcal{L}(\mu)$ be the σ-algebra of all Lebesgue μ-measurable subsets of Ω . Show that f is $(\mathcal{L}(\mu), \mathcal{A})$-measurable if and only if $f = g$ μ-a.s. for some $(\mathcal{F}, \mathcal{A})$-measurable function $g : \Omega \to M$.

Hint: Suppose that f is $(\mathcal{L}(\mu), \mathcal{A})$-measurable, and let $A_1, A_2, \ldots \subseteq M$ be chosen such that $\mathcal{A} = \sigma\{A_1, A_2, \ldots\}$. Then choose $F_n, G_n \in \mathcal{F}$ such that $F_n \subseteq f^{-1}(A_n) \subseteq G_n$ and $\mu(N) = 0$, where $N = \cup_1^\infty (G_n \setminus F_n)$. Set

$$g(\omega) = \begin{cases} f(\omega) & \text{if } \omega \in \Omega \setminus N \\ a & \text{if } \omega \in N \end{cases}$$

and show that g works. The converse is easy.

Exc. 1.69: Prove the image and co-image formulas (P.3.1)–(P.3.7).

Exc. 1.70: Let $M = \{1, 2, \ldots n\}$, where $n \geq 1$ is a given integer. Let $1 \leq k \leq n$ be given and set

$$\Omega = \{(\omega_1, \ldots, \omega_k) \in M^k \mid \omega_i \neq \omega_j \quad \forall i \neq j\} \quad , \quad \Lambda = \{\gamma \in 2^M \mid \#\gamma = k\}$$

$$P(F) = \tfrac{(n-k)!}{n!} \#F \quad \forall F \subseteq \Omega \quad , \quad Q(G) = \binom{n}{k}^{-1} \#G \quad \forall G \subseteq \Lambda$$

$$T(\omega_1, \ldots, \omega_k) = \{\omega_1, \ldots, \omega_k\} \quad \forall (\omega_1, \ldots, \omega_k) \in \Omega$$

Show that P and Q are probability measures on $(\Omega, 2^\Omega)$ respectively $(\Lambda, 2^\Lambda)$, and that T is a measurable surjection of $(\Omega, 2^\Omega)$ onto $(\Lambda, 2^\Lambda)$ such that $Q = P_T$.

Exc. 1.71: Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be two measurable spaces, and let Ω_0 be a subset of Ω equipped with trace σ-algebra $\mathcal{F}_0 = \mathrm{tr}(\mathcal{F} \mid \Omega_0)$ [see Exc.1.28]. Let $g : \Omega \to \Lambda$ be a measurable function, and let g_0 be the restriction of g to Ω_0, and let $\iota(\omega) = \omega$ be the identity map from Ω_0 into Ω. Show that $\mathcal{F}_0 = \iota^{-1}(\mathcal{F})$, and that g_0 is $(\mathcal{F}_0, \mathcal{G})$-measurable from Ω_0 into Λ.

Exc. 1.72: Let $\mathcal{F}$ be a paving on the set Ω, and let $f : \Omega \to \Lambda$ be a given function. Then we define

$$f^{\leftarrow}(\mathcal{F}) := \{B \subseteq \Lambda \mid f^{-1}(B) \in \mathcal{F}\} \quad , \quad f(\mathcal{F}) := \{f(F) \mid F \in \mathcal{F}\}$$

Let $\mathcal{F}_0 = \mathcal{F} \cap f^{-1}(2^{\Lambda})$ and $\Lambda_0 = f(\Omega)$. Show that

(i) $\quad f^{\leftarrow}(\mathcal{F}) = \{B \subseteq \Lambda \mid B \cap \Lambda_0 \in f(\mathcal{F}_0)\}$

Let γ be a list of some of the set operations in list (1.1.1), and let $\gamma(\mathcal{A})$ denote the smallest γ-stable paving containing $\mathcal{A}$ [see Exc.1.28]. Suppose that $\mathcal{F}_0 = \mathcal{F} \cap f^{-1}(2^{\Lambda}) \neq \emptyset$, and show the following:

(ii) $\quad f(\mathcal{F}_0) \subseteq f^{\leftarrow}(\mathcal{F})$

(iii) If $\mathcal{F}$ is γ-stable, then so is $f^{\leftarrow}(\mathcal{F})$

(iv) $\quad \gamma(f^{\leftarrow}(\mathcal{F})) = f^{\leftarrow}(\gamma(\mathcal{F}_0))$

(v) If $\mathcal{F}$ is an algebra respectively a σ-algebra, then so is $f^{\leftarrow}(\mathcal{F})$

(vi) If $\mathcal{F}$ and $\mathcal{G}$ are σ-algebras on Ω and Λ, respectively, then f is measurable if and only if $\mathcal{G} \subseteq f^{\leftarrow}(\mathcal{F})$

Hint: The inclusion $\subseteq$ in (iv) follows easily from (iii), and the converse inclusion follows easily from (i) and the following equalities:

$$f\Big(\bigcap_{j \in J} A_j\Big) = \bigcap_{j \in J} f(A_j) \quad \text{if} \quad A_j \in f^{-1}(2^{\Lambda}) \quad \forall j \in J$$

$$f(A \setminus B) = f(A) \setminus f(B) \quad , \quad f(A^c) = \Lambda_0 \cap f(A)^c \quad \text{if} \quad A, B \in f^{-1}(2^{\Lambda})$$

Exc. 1.73: Let (S, d) be a separable metric space and let X and Y be measurable functions from the measurable space $(\Omega, \mathcal{F})$ into $(S, \mathcal{B}(S))$. Show that the function $F(\omega) = d(X(\omega), Y(\omega))$ is measurable: $\Omega \to \mathbf{R}$.

Hint: Apply (1.43.3) to show that $Z(\omega) = (X(\omega), Y(\omega))$ is measurable from $(\Omega, \mathcal{F})$ into $(S \times S, \mathcal{B}(S) \otimes \mathcal{B}(S))$. Note that $F = d \circ Z$ and that d is continuous on the product space $S \times S$ and, thus, $\mathcal{B}(S \times S)$-measurable. Then apply (1.43.4) and (1.38.4).

Warning: If (S,d) is not separable, then the function F need not be measurable. Actually if the cardinality of S exceeds the cardinality of $\mathbf{R}$, then the metric $d : S \times S \to \mathbf{R}$ is **not** measurable with respect to the product σ-algebra $\mathcal{B}(S) \otimes \mathcal{B}(S)$ [see Exc.1.87] but, of course, it is measurable with respect to the Borel σ-algebra $\mathcal{B}(S \times S)$. Let ℓ^∞ denote the set of all bounded sequences (x_n) of real numbers with the sup-metric

$$||x - y||_\infty := \sup_{n \geq 1} |x_n - y_n| \quad \forall x = (x_n) \in \ell^\infty \; \forall y = (y_n) \in \ell^\infty$$

Recently M. Talagrand has shown that the function: $(x,y) \curvearrowright ||x-y||_\infty$ from $\ell^\infty \times \ell^\infty$ into $\mathbf{R}$ is measurable with respect to the product σ-algebra $\mathcal{B}(\ell^\infty) \otimes \mathcal{B}(\ell^\infty)$ if and only if the continuum hypothesis holds [see (P.12.40) in the Appendix to Volume II]. This shows that one should be careful with measurability in non-separable metric spaces.

Exc. 1.74: Let (S,d) be a *polish space* (that is, a separable complete metric space), and let $Y, X_1, X_2, \ldots$ be a sequence of measurable functions from the measurable space $(\Omega, \mathcal{F})$ into S. Let

$$L = \{\omega \in \Omega \mid \{X_1(\omega), X_2(\omega), \ldots\} \text{ is convergent}\}$$

$$X(\omega) = \begin{cases} \lim\limits_{n\to\infty} X_n(\omega) & \text{if } \omega \in L \\ Y(\omega) & \text{if } \omega \notin L \end{cases}$$

Show that $L \in \mathcal{F}$ and that X is a measurable function. A metric space (S,d) is called *standard* if there exists a polish space (T,δ) such that (S,d) is imbedded in (T,δ) [see (P.5)] and $S \in \mathcal{B}(T)$. Show that if (S,d) is a standard metric space, then L and X defined as above are $\mathcal{F}$-measurable.

Hint: By completeness of S, we have that $\{X_n(\omega)\}$ converges if and only if $\{X_n(\omega)\}$ is a Cauchy sequence. Use this observation to show that

$$L = \bigcap_{i=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcap_{j=1}^{\infty} \{\omega \mid d(X_{k+n}(\omega), X_{j+n}(\omega)) < 2^{-i}\}$$

and then use Exc.1.73 to deduce that $L \in \mathcal{F}$. Let $Y_n(\omega) = X_n(\omega)$ for $\omega \in L$ and $Y_n(\omega) = Y(\omega)$ for $\omega \in \Omega \setminus L$. Show that Y_n is measurable and $Y_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$, and apply (1.41).

Warning: If S is not standard, then the set L need not belong to the σ-algebra $\mathcal{F}$ [see also the warning to Exc.1.73].

Exc. 1.75: Let $(\Omega, \mathcal{F})$ be a measurable space, and let (S, d) be a metric space. In analogy with (1.45) we say that a function $X : \Omega \to S$ is *simple* respectively *discrete* if X is measurable and $X(\Omega)$ is finite respectively countable. Let $X : \Omega \to S$ be a given function. Show that the following four statements are equivalent:

(i) X is measurable and $X(\Omega)$ is separable

(ii) $\exists\, X_1, X_2, \ldots$ discrete functions so that $X_n \rightrightarrows X$ uniformly on Ω

(iii) $\exists\, X_1, X_2, \ldots$ discrete functions so that $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$

(iv) $\exists\, X_1, X_2, \ldots$ simple functions so that $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$

A function $X : \Omega \to S$ satisfying one, and hence all, of conditions (i)–(iv) is said to be *Bochner measurable.*

Hint: Suppose that X is measurable and that $X(\Omega)$ is separable. Let $D = \{a_1, a_2, \ldots\}$ be countable dense subset of $X(\Omega)$ and set

$$S_{1j} = b\big(a_1, 2^{-j}\big) \ , \ S_{kj} = b\big(a_k, 2^{-j}\big) \setminus \bigcup_{i=1}^{k-1} b\big(a_i, 2^{-j}\big) \quad \text{if } k \geq 2$$

Replace the intervals $[k2^{-j}, (k+1)2^{-j}[$ in the proof of (1.46) with S_{kj} , and mimicking the argument of (1.46), show that (i) implies (ii)–(iv). The converse implications are easy.

Exc. 1.76: Let Ω be a set, then $\mathcal{T} = \{\emptyset, \Omega\}$ is the smallest σ-algebra on Ω , and it is called *the trivial σ-algebra.* Similarly, $\mathcal{D} = 2^{\Omega}$ is the largest σ-algebra on Ω , and it is called *the discrete σ-algebra.* Let $(\Lambda, \mathcal{G})$ be an arbitrary measurable space and let $f : \Omega \to \Lambda$ and $g : \Lambda \to \Omega$ be given functions. Show that f is $(\mathcal{D}, \mathcal{G})$-measurable and g is $(\mathcal{G}, \mathcal{T})$-measurable. Suppose that g is injective. Then show that g is $(\mathcal{G}, \mathcal{D})$-measurable if and only if $\mathcal{G} = 2^{\Lambda}$.

Suppose that the σ-algebra $\mathcal{G}$ *separates points*, that is, for all $\lambda' \neq \lambda''$ in Λ there exists a set $G \in \mathcal{G}$ such that $\lambda' \in G$ but $\lambda'' \notin G$. Then show that f is $(\mathcal{T}, \mathcal{G})$-measurable if and only if f is constant.

Exc. 1.77: Let $(\Omega, \mathcal{F})$ be a measurable space and let $A \subseteq \Omega$. Show that 1_A is measurable if and only if $A \in \mathcal{F}$.

Exc. 1.78: Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \mathbf{R}$ be a discrete function. Show that $f^{-1}(B) \in \mathcal{F}$ for all $B \subseteq \mathbf{R}$.

Exc. 1.79: Let $p \geq 2$ be a given integer and let $x \in [0,1[$. Then we define *the p-adic expansion* of x , denoted $x \sim 0.\, p_1(x)\, p_2(x)\, \ldots\, p_n(x)\, \ldots$, inductively as follows:

$$p_1(x) := int(p\,x) \quad , \quad p_n(x) := int\left(p^n\, x - \sum_{j=1}^{n-1} p_j(x)\, p^{n-j}\right)$$

where $int(y)$, as usual, denotes the integer part of y . Show that p_n is a Borel function and that

(i) $\quad 0 \leq x - \sum_{j=1}^{n} p_j(x)\, p^{-j} < p^{-n} \qquad \forall n \geq 1$

(ii) $\quad x = \sum_{n=1}^{\infty} p_n(x)\, p^{-n} \qquad \forall 0 \leq x < 1$

(iii) $\quad p_n(x) \in \{0, 1, \ldots, p-1\} \qquad \forall 0 \leq x < 1\ \forall n \geq 1$

Suppose that $x \in [0,1[$ admits an expansion of the form

(*) $\quad x = \sum_{n=1}^{\infty} a_n\, p^{-n} \quad$ where $\quad a_n \in \{0, 1, \ldots, p-1\} \quad \forall n \geq 1$

and set $m = \sup\{n \in \mathrm{N} \mid a_n \neq p-1\}$ with the convention: $\sup \emptyset = 0$. Show that $1 \leq m \leq \infty$, and

(iv) $\quad p_n(x) = a_n \qquad \forall n < m$

(v) $\quad$ If $m < \infty$, then $p_m(x) = 1 + a_m$ and $p_n(x) = 0$ for all $n > m$

Thus, if $m = \infty$, that is, if a_n is not equal to $p-1$ for all n from a certain point and onwards, then $0.\, a_1\, a_2\, \ldots$ is the p-adic expansion of x .

Remark: If $p = 10$, then $x \sim 0.\, 10_1(x)\, 10_2(x)\, \ldots$ is simply the usual *decimal expansion* of x , and $10_n(x)$ is simply the nth-decimal of the number x . Similarly, if $p = 2$, then $x \sim 0.\, 2_1(x)\, 2_2(x)\, \ldots$ is simply the usual *binary expansion* of x , and $2_n(x)$ is simply the nth-dyadic of the number x .

Hint: (i): Note that by definition of $int(x)$ we have

$$0 \leq \left(p^n\, x - \sum_{j=1}^{n-1} p_j(x)\, p^{n-j}\right) - p_n(x) < 1$$

Multiply these inequalities with p^{-n} . (ii): Follows easily from (i). (iii): Apply (i) with $n := n-1$ and multiply through with p^n . (iv)+(v): Let x be

given on the form (*) and suppose that $k \in \mathbf{N}$ such that $p_j(x) = a_j$ for all $1 \leq j < k$. Show that

$$p^k x - \sum_{j=1}^{k-1} p_j(x)\, p^{k-j} = a_k + \sum_{j=k+1}^{\infty} a_j\, p^{k-j}$$

Use this and the formula

$$\sum_{j=k+1}^{\infty} (p-1)\, p^{k-j} = 1$$

to prove (iv) and (v).

Exc. 1.80: *(The Marczewski function)* Let $\{A_n\}$ be a sequence of subsets of the set Ω . Then *the Marczewski function* induced by $\{A_n\}$ is defined as follows:

$$f(\omega) = \sum_{n=1}^{\infty} 10^{-n}\, 1_{A_n}(\omega) \qquad \forall \omega \in \Omega$$

Show that $f^{-1}(\mathcal{B}(\mathbf{R})) = \sigma\{A_n \mid n \in \mathbf{N}\}$, and show that f is injective if and only if $\{A_n\}$ satisfies

(*) $\quad \forall \omega' \neq \omega'' \; \exists n \geq 1$ so that $1_{A_n}(\omega') \neq 1_{A_n}(\omega'')$

Hint: Let $f_n(\omega) = 10_n(f(\omega))$, where $10_n(x)$ is the nth decimal function [see Exc.1.79]. Show that $A_n = f_n^{-1}(\{1\})$, and argue that this proves the inclusion: $\sigma\{A_n \mid n \in \mathbf{N}\} \subseteq f^{-1}(\mathcal{B}(\mathbf{R}))$. The converse inclusion follows from Exc.1.79, (1.43.7) and (1.41).

Exc. 1.81: Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Lambda \to \Omega$ be a function. Show that if $\mathcal{F}$ is a separable σ-algebra [see Exc.1.13], then so is $f^{-1}(\mathcal{F})$.

Exc. 1.82: Let $(\Omega, \mathcal{F})$ be a measurable space. Show that $\mathcal{F}$ is a separable σ-algebra [see Exc.1.13] if and only if there exists a measurable function $f : \Omega \to \mathbf{R}$ such that $\mathcal{F} = f^{-1}(\mathcal{B}(\mathbf{R}))$.

Hint: Apply Exc.1.81 and Exc.1.14 to prove the "if" part. Apply Exc.1.80 to prove the "only if" part.

Exc. 1.83: Let $(\Omega, \mathcal{F})$ be a measurable space. Then we say that $\mathcal{F}$ is *countably separating* if there exists a countable paving $\mathcal{G} \subseteq \mathcal{F}$ satisfying

(*) $\quad \forall \omega' \neq \omega'' \; \exists G \in \mathcal{G}$ so that $1_G(\omega') \neq 1_G(\omega'')$

Let $f : \Lambda \to \Omega$ be a injective function. Show that if $\mathcal{F}$ is countably separating, then so is $f^{-1}(\mathcal{F})$.

Exc. 1.84: Let (S, d) be a separable metric space. Show that $\mathcal{B}(S)$ is countably separating [see Exc.1.83].

Hint: Let $\{a_1, a_2, \ldots\}$ be a countable dense subset of S . Then show that $\mathcal{G} = \{b(a_j, 2^{-n}) \mid j, n \in \mathbf{N}\}$ satisfies (*) in Exc.1.83.

Exc. 1.85: Let $(\Omega, \mathcal{F})$ be a measurable space. Show that $\mathcal{F}$ is countably separating [see Exc.1.83] if and only if there exists an injective measurable function f from Ω into $\mathbf{R}$.

Hint: Apply Exc.1.83 and Exc.1.84 to prove the "if" part. Apply Exc.1.80 to prove the "only if" part.

Exc. 1.86: Let $(\Omega, \mathcal{F})$ be a measurable space and let $\Delta = \{(\omega, \omega) \mid \omega \in \Omega\}$ be the *diagonal* in $\Omega \times \Omega$. Show that $\mathcal{F}$ is countably separating [see Exc.1.83] if and only if $\Delta \in \mathcal{F} \otimes \mathcal{F}$.

Hint: If $f : \Omega \to \mathbf{R}$ is injective, then $\Delta = \{(\omega_1, \omega_2) \mid f(\omega_1) = f(\omega_2)\}$. Now apply Exc.1.85 to prove the "only if" part. Suppose that $\Delta \in \mathcal{F} \otimes \mathcal{F}$, then apply Exc1.10 to show that there exists a countable paving $\mathcal{G} \subseteq \mathcal{F}$ such that $\Delta \in \sigma\{G_1 \times G_2 \mid G_1, G_2 \in \mathcal{G}\}$, and verify that $\mathcal{G}$ satisfies (*) in Exc.1.83.

Exc. 1.87: Let $(\Omega, \mathcal{F})$ be a measurable space, and suppose that Ω has cardinality strictly larger than $\mathbf{R}$, that is, there exist no injective functions from Ω into $\mathbf{R}$. Show that the diagonal $\Delta = \{(\omega, \omega) \mid \omega \in \Omega\}$ does **not** belong to the product σ-algebra $\mathcal{F} \otimes \mathcal{F}$. [Apply Exc.1.86.] Deduce that if (S, d) is a metric space such that S has cardinality strictly larger than $\mathbf{R}$, then the metric $d(x, y)$ is **not** measurable with respect to the product σ-algebra $\mathcal{B}(S) \otimes \mathcal{B}(S)$.

Exc. 1.88: Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be two measurable spaces, let ν be a measure on $(\Lambda, \mathcal{G})$, and let $T : \Omega \to \Lambda$ be a given function such that $\mathcal{F} = T^{-1}(\mathcal{G})$ and $T(\Omega) \in \mathcal{G}$. Show that

(i) $\quad T(F) \in \mathcal{G} \qquad \forall F \in \mathcal{F}$

(ii) $\quad \exists \mu$ a measure on $(\Omega, \mathcal{F})$: $\nu = \mu_T \iff \nu(\Lambda \setminus T(\Omega)) = 0$

Note that (i) gives a partial solution to the image catastrophe [see (1.47)].

Hint: (i): Take a set $F \in \mathcal{F}$. Then by assumption there exists a set $G \in \mathcal{G}$ with $F = T^{-1}(G)$, and by (P.3.6) we have $T(F) = F \cap T(\Omega)$. (ii): Let $\mu(F) := \nu(T(F))$ for all $F \in \mathcal{F}$. Let $F_1, F_2 \in \mathcal{F}$ such that $F_1 \cap F_2 = \emptyset$ and apply (P.3.6) to show that $T(F_1) \cap T(F_2) = \emptyset$. Conclude that μ is a measure and show that $\nu = \mu_T$ if $\nu(\Lambda \setminus T(\Omega)) = 0$. The converse is easy.

Exc. 1.89: Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be measurable spaces and let $f : \Omega \to \Lambda$ be a measurable function. Let

$$\Gamma = \{(\omega, f(\omega)) \mid \omega \in \Omega\} \subseteq \Omega \times \Lambda$$

be *the graph* of the function f and suppose that $\mathcal{G}$ is countably separating [see Exc.1.83]. Show that the graph Γ belongs to the product σ-algebra $\mathcal{F} \otimes \mathcal{G}$.

Hint: By Exc.1.85 there exist an injective measurable function $\varphi : \Lambda \to \mathbf{R}$. Let $\Phi(\omega, \lambda) = (\varphi(f(\omega)), \varphi(\lambda))$. Show that Φ is a measurable function from $\Omega \times \Lambda$ into $\mathbf{R}^2$ and that $\Gamma = \Phi^{-1}(D)$, where $D = \{(x, y) \in \mathbf{R}^2 \mid x = y\}$ is the diagonal in $\mathbf{R}^2$.

Exc. 1.90: Let S be a separable metric space, and let $(T, \mathcal{A})$ be a measurable space. Suppose that $f : S \times T \to \mathbf{R}$ is a function such that $f(\cdot, t)$ is continuous for all $t \in T$ and $f(s, \cdot)$ is $\mathcal{A}$-measurable for all $s \in S$. Show that f is $(\mathcal{B}(S) \otimes \mathcal{A})$-measurable.

Hint: Let $\{a_1, a_2 \ldots\}$ be a countable dense sequence in S and set:

$$D_1^n = b(a_1, 2^n) \quad , \quad D_j^n = b(a_j, 2^n) \setminus \bigcup_{i=1}^{j-1} b(a_i, 2^n) \quad \text{for} \quad j = 2, 3, \ldots$$

$$f_n(s, t) = \sum_{j=1}^{\infty} f(a_j, t)\, 1_{D_j^n}(s) \qquad \forall n = 1, 2, \ldots$$

Show that $D_1^n, D_2^n, \ldots$ are disjoint and

$$S = \bigcup_{j=1}^{\infty} D_j^n \quad \forall n \geq 1 \quad \text{and} \quad f(s, t) = \lim_{n \to \infty} f_n(s, t) \quad \forall (s, t) \in S \times T$$

Then show that f_n is $(\mathcal{B}(S) \otimes \mathcal{A})$-measurable for all $n \geq 1$, and apply (1.40.4).

Exc. 1.91: *(The Souslin operation)* In (1.1.1) are listed the 11 most common set operations. In this exercise I shall add a 12th operation to the list. The operation is called *the Souslin operation*, and it is of a much more subtler type than the 11 simple operations in list (1.1.1). Let us set

$$\mathbf{N}^{(\mathbf{N})} := \{(n_1, \ldots, n_k) \mid k \in \mathbf{N} \ , \ n_1, \ldots, n_k \in \mathbf{N}\}$$

$$\mathbf{N}^{\mathbf{N}} := \{\sigma \mid \sigma \text{ is a function } \mathbf{N} \to \mathbf{N}\}$$

A k-tuple $(n_1, \ldots, n_k) \in \mathbf{N}^{(\mathbf{N})}$ is called a *multiindex of length* k . Let Ω be a given set. Then a *Souslin scheme* on Ω is a function $A : \mathbf{N}^{(\mathbf{N})} \to 2^{\Omega}$, that

is, $A(n_1,\ldots,n_k)$ is a subset of Ω for all $(n_1,\ldots,n_k) \in \mathbf{N}^{(\mathbf{N})}$. If $\mathcal{F}$ is a paving on Ω such that $A(n_1,\ldots,n_k) \in \mathcal{F}$ for all $(n_1,\ldots,n_k) \in \mathbf{N}^{(\mathbf{N})}$, then we say that A is an *$\mathcal{F}$-Souslin scheme*. If A is a Souslin scheme, then we define *the Souslin operation* as follows:

$$S(A) := \bigcup_{\sigma \in \mathbf{N}^{\mathbf{N}}} \bigcap_{k=1}^{\infty} A(\sigma(1),\ldots,\sigma(k))$$

Thus, the Souslin operation is a set operation, denoted (S) , just as the 11 other operations in list (1.1.1). The Souslin operation operates on Souslin schemes, just as $(\cup c)$ operates on sequences of sets, and $(\setminus)$ operates on pairs of sets. If $\mathcal{F}$ is paving on Ω , then we define the paving of all *$\mathcal{F}$-Souslin sets*, denoted $S(\mathcal{F})$, as follows:

$$S(\mathcal{F}) := \{S(A) \mid A \text{ is an } \mathcal{F}\text{-Souslin scheme}\}$$

We say that the paving is *(S)-stable* or *Souslin-stable* if $S(\mathcal{F}) \subseteq \mathcal{F}$. Show that

(i) $\mathcal{F} \subseteq S(\mathcal{F}) = S(S(\mathcal{F}))$

(ii) $S(\mathcal{F})$ is $(\cup c, \cap c)$-stable

The Souslin operation was invented by M.M. Souslin (1916). It is admittedly much harder to work with than the usual set operations, but it is a very powerful tool for establishing measurability of complicated sets. Note that $\mathbf{N}^{(\mathbf{N})}$ is countable [see (P.4.5) and (P.4.6)]. Thus, the Souslin operation operates on a countable number of sets, but it is **not** a countable operation since we take the union over the uncountable set $\mathbf{N}^{\mathbf{N}}$.

Hint: (i): Setting $A(n_1,\ldots,n_k) = B$ for all $(n_1,\ldots,n_k) \in \mathbf{N}^{(\mathbf{N})}$, we see that $S(A) = B$. So $\mathcal{F} \subseteq S(\mathcal{F}) \subseteq S(S(\mathcal{F}))$. Let $D \in S(S(\mathcal{F}))$. Then there exist an $S(\mathcal{F})$-Souslin scheme $C(\cdot) : \mathbf{N}^{(\mathbf{N})} \to S(\mathcal{F})$ with $D = S(C)$. For each $\lambda \in \mathbf{N}^{(\mathbf{N})}$ we have that $C(\lambda) \in S(\mathcal{F})$, and so there exists an $\mathcal{F}$-Souslin scheme $B_\lambda(\cdot) : \mathbf{N}^{(\mathbf{N})} \to \mathcal{F}$ with $C(\lambda) = S(B_\lambda)$. Observe that for every $k \in \mathbf{N}$ there exists **unique** integers $\alpha, \beta \in \mathbf{N}$ such that $k = 2^{\alpha-1}(2\beta - 1)$. Let $k \in \mathbf{N}$ and $\lambda = (\lambda_1,\ldots,\lambda_k) \in \mathbf{N}^k$ be given. Then there exists unique integers $p, q, \alpha(j), \beta(j) \in \mathbf{N}$ satisfying

$$k = 2^{p-1}(2q-1) \quad , \quad \lambda_j = 2^{\alpha(j)-1}(2\beta(j)-1) \quad \forall 1 \le j \le k$$

Set $\gamma(i) = \beta(2^{p-1}(2i-1))$ for $1 \le i \le q$ and

$$A(\lambda) = B_{\alpha(1),\ldots,\alpha(p)}(\gamma(1),\ldots,\gamma(q))$$

Then A is an $\mathcal{F}$-Souslin scheme. Show that $D = S(A)$, and conclude that this proves (i). (ii): Let $\{F_k\} \subseteq \mathcal{F}$, and set $I(n_1, \ldots, n_k) = F_k$ and $J(n_1, \ldots, n_k) = F_{n_1}$. Show that

$$S(I) = \bigcap_{k=1}^{\infty} F_k \quad , \quad S(J) = \bigcup_{k=1}^{\infty} F_k$$

Conclude that (ii) follows from (i).

Exc. 1.92: Let $f : \Omega \to \Lambda$ be a function and let $\mathcal{H}$ be a paving on Λ. Let (1.1.1)* denote the list of set operations from (1.1.1) supplemented by the Souslin operation (S) [see Exc.1.91] and let γ be a list of some of the set operations from (1.1.1)*. If $\mathcal{M}$ is a paving on some set M, then we let $\gamma(\mathcal{M})$ denote the smallest γ-stable paving on M containing $\mathcal{M}$. For instance, if $\sigma = (C, \cup c)$, then $\sigma(\mathcal{M})$ is the smallest σ-algebra containing $\mathcal{M}$. Show that $\gamma(f^{-1}(\mathcal{H})) = f^{-1}(\gamma(\mathcal{H}))$

Hint: Mimic the proof of (1.38.5).

Exc. 1.93: Let $\mathcal{F}$ be a $(\cup f, \cap f)$-stable paving on the set Ω, and let $A(n_1, \ldots, n_k)$ be an $\mathcal{F}$-Souslin scheme [see Exc.1.91]. Let us define

$$\mathbf{N}^*(n_1, \ldots, n_k) = \{\, \sigma \in \mathbf{N}^{\mathbf{N}} \mid \sigma(j) \le n_j \quad \forall j = 1, \ldots, k \}$$

$$A^*(n_1, \ldots, n_k) = \bigcup_{\sigma \in \mathbf{N}^*(n_1, \ldots, n_k)} \bigcap_{j=1}^{\infty} A(\sigma(1), \ldots, \sigma(j))$$

$$\mathbf{N}^k(n_1, \ldots, n_k) = \{\, m \in \mathbf{N}^k \mid m_j \le n_j \quad \forall j = 1, \ldots, k \}$$

$$A^\circ(n_1, \ldots, n_k) = \bigcup_{m \in \mathbf{N}^k(n_1, \ldots, n_k)} \bigcap_{j=1}^{k} A(m_1, \ldots, m_j)$$

whenever $(n_1, \ldots, n_k) \in \mathbf{N}^k$. Let $n_1, n_2 \ldots \in \mathbf{N}$ be given and show that

(i) $A^*(p) \uparrow S(A)$ and $A^*(n_1, \ldots, n_k, p) \uparrow A^*(n_1, \ldots, n_k)$ as $p \uparrow \infty$

(ii) $A^\circ(n_1, \ldots, n_k)$ is an $\mathcal{F}$-Souslin scheme

(iii) $A^\circ(n_1, \ldots, n_k, n_{k+1}) \subseteq A^\circ(n_1, \ldots, n_k) \quad \forall k \ge 1$

(iv) $A^*(n_1, \ldots, n_k) \subseteq A^\circ(n_1, \ldots, n_k) \quad \forall k \ge 1$

(v) $S(A) = S(A^\circ) = S(A^*)$

Hint: (v): Suppose that $\omega \in S(A^\circ)$. Then there exists $\sigma \in \mathbf{N}^{\mathbf{N}}$ such that $\omega \in A^\circ(\sigma(1), \ldots, \sigma(k))$ for all $k \geq 1$. Thus, for every $k \geq 1$ there exists $m^k = (m_1^k, \ldots, m_k^k) \in \mathbf{N}^k$ such that $\omega \in A(m_1^k, \ldots, m_j^k)$ and $m_j^k \leq \sigma(j)$ for all $k \geq j \geq 1$. Show that the set $N_1 = \{k \in \mathbf{N} \mid m_1^k = \tau(1)\}$ is infinite for some $\tau(1) \leq \sigma(1)$. Then show that the set $N_2 = \{k \in N_1 \mid k \geq 2 \, , \, m_2^k = \tau(1)\}$ is infinite for some $\tau(2) \leq \sigma(2)$. Continue in this way and show that there exists $\tau \in \mathbf{N}^{\mathbf{N}}$ such that for every fixed $j \in \mathbf{N}$ we have that $m_j^k = \tau(j)$ for infinitely many $k \geq j$. Show that $\omega \in A(\tau(1), \ldots, \tau(k))$ for all $k \geq 1$ and argue that this shows $S(A^\circ) \subseteq S(A)$. The rest is easy.

Exc. 1.94: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $\mathcal{M}(\mu)$ be the paving of all μ-measurable subsets of Ω [see (1.28)]. Show that $\mathcal{M}(\mu)$ is Souslin stable.

Hint: Let $\mathcal{H} := \{F \in \mathcal{F} \mid \mu(F) < \infty\}$, and let $L = S(A)$ where $A(n_1, \ldots, n_k)$ is a $\mathcal{M}(\mu)$-Souslin scheme. Let $H \in \mathcal{H}$ and $a < \mu^*(H \cap L)$ be given. Define the Souslin schemes A^* and A° as in Exc.1.93. Apply Exc.1.93.(i) and Exc.1.38.(v) to show that there exists $\tau \in \mathbf{N}^{\mathbf{N}}$ satisfying

$$a < \mu^*(H \cap A^*(\tau(1), \ldots, \tau(k))) \qquad \forall k \geq 1$$

Let $L^\circ = \cap_{j=1}^\infty A^\circ(\tau(1), \ldots, \tau(j))$, and apply Exc.1.93 and (1.4.3) to show that $L^\circ \subseteq L$ and $\mu^*(H \cap L^\circ) \geq a$. Deduce that

$$a + \mu^*(H \setminus L) \leq \mu^*(H \cap L^\circ) + \mu^*(H \setminus L^\circ) = \mu^*(H)$$

and apply Exc.1.40.(i).

Exc. 1.95: *(The bi-Souslin sets)* Let $\mathcal{F}$ be a paving on the set Ω . Then we define the paving of all *bi-Souslin $\mathcal{F}$-sets*, denoted $\mathcal{B}_{\mathcal{I}}(\mathcal{F})$, as follows:

$$\mathcal{B}_{\mathcal{I}}(\mathcal{F}) := \{A \subseteq \Omega \mid A \in S(\mathcal{F}) \text{ and } A^c \in S(\mathcal{F})\}$$

where $S(\mathcal{F})$ is the paving of all $\mathcal{F}$-Souslin sets [see Exc.1.91]. Show the following [apply Exc.1.91]

(i) If $\mathcal{B}_{\mathcal{I}}(\mathcal{F}) \neq \emptyset$, then $\mathcal{B}_{\mathcal{I}}(\mathcal{F})$ is a σ-algebra

(ii) If $F^c \in S(\mathcal{F})$ for all $F \in \mathcal{F}$, then $\mathcal{B}_{\mathcal{I}}(\mathcal{F})$ is a σ-algebra such that $\sigma(\mathcal{F}) \subseteq \mathcal{B}_{\mathcal{I}}(\mathcal{F})$

Exc. 1.96: Let (M, d) be metric space, and let $\mathcal{F}(M)$ be the paving of all closed subsets of M . Show that

(i) $\mathcal{B}(M) \subseteq \mathcal{B}_{\mathcal{I}}(\mathcal{F}(M))$

where $\mathcal{B}_{\mathcal{I}}(\mathcal{F}(M))$ is the paving of all bi-Souslin $\mathcal{F}(M)$-sets [see Exc.1.95]. It can be shown that if (M,d) is an analytic metric space [see (P.5)]. Then we have equality in (i), but this is much more difficult and requires the so-called *separation theorem.*

Hint: Apply Exc.1.19, and Exc.1.95(ii).

Exc. 1.97: Let $(T,d),(T_1,d_1),(T_2,d_2),\dots$ be polish spaces [see (P.5)], and let F be a closed subset of (T,d) . Show that F is polish and that the product space $\prod_n T_n$ is polish in its product metric. Let us define *the topological sum* of $(T_1,d_1),(T_2,d_2),\dots$ as follows:

$$\sum_{n=1}^{\infty} T_n := \bigcup_{n=1}^{\infty} T_n \times \{n\}$$

$$d_\sigma((x,n),(y,m)) = \begin{cases} \arctan d_n(x,y) & \text{if } n=m \\ \frac{\pi}{2} & \text{if } n \neq m \end{cases}$$

Show that $\left(\sum_{n=1}^{\infty} T_n, d_\sigma\right)$ is a polish space.

Hint: Recall that $\arctan(a+b) \leq \arctan a + \arctan b$ and $\arctan a \leq \frac{\pi}{2}$ for all $a,b \geq 0$.

Exc. 1.98: Let (M,d) be a metric space, and let $\mathcal{A}(M)$ denote the paving of all analytic subsets of M [see (P.5)]. Show the following

(i) If φ is a continuous function from M into the metric space (L,ρ) , then $\varphi(A)$ is analytic for all analytic sets $A \subseteq M$

(ii) $\mathcal{A}(M)$ is $(\cup c, \cap c)$-stable

(iii) $B \cap A \in \mathcal{A}(M) \quad \forall B \in \mathcal{B}(M)\ \forall A \in \mathcal{A}(M)$

(iv) If M is analytic, then $\mathcal{B}(M) \subseteq \mathcal{A}(M)$

Hint: (ii): Let A_n be analytic for all $n \geq 1$, and let T_n be polish and $f_n : T_n \to M$ be continuous such that $A_n = f(T_n)$. Let $T^\infty = \prod_1^\infty T_n$ be the product space and let $T_\sigma = \sum_1^\infty T_n$ be the topological sum [see Exc.1.97]. Set $g(\xi) = f_n(v)$ for all $\xi = (v,n) \in T_\sigma$, and $f(t) = f_1(t_1)$ for all $t = (t_j) \in T$, where

$$T = \{(t_n) \in T^\infty \mid f_n(t_n) = f_1(t_1) \quad \forall n \geq 1\}$$

Show that f and g are continuous functions such that $f(T) = \cap_{n=1}^\infty A_n$ and $g(T_\sigma) = \cup_{n=1}^\infty A_n$. Then apply Exc.1.97. (iii): Let

$$\mathcal{H} = \{\, B \subseteq M \mid B \cap A \in \mathcal{A}(M) \quad \forall A \in \mathcal{A}(M)\,\}$$

Use (ii) to show that $\mathcal{H}$ is $(\cup c, \cap c)$-stable. Use Exc.1.97 to show that $\mathcal{H}$ contains all closed subsets of M . Finally, use Exc.1.19 to prove (iii).

Exc. 1.99: Let $(M_1, d_1), (M_2, d_2), \ldots$ be a sequence of metric spaces, and let $M = \prod_n M_n$ be the product space. Let A_n be an analytic subset of M_n for all n, show that $\prod_n A_n$ is an analytic subset of M.

Exc. 1.100: Let (M, d) be a metric space and let $\mathcal{F}(M)$ and $\mathcal{A}(M)$ denote the pavings os all closed respectively analytic subsets of M [see (P.5)]. Show that

(i) $\mathcal{A}(M) \subseteq S(\mathcal{F}(M))$

(ii) $\mathcal{A}(M)$ is Souslin-stable

(iii) $H \cap A \in \mathcal{A}(M) \quad \forall H \in S(\mathcal{B}(M)) \, \forall A \in \mathcal{A}(M)$

(iv) If M is analytic, then $\mathcal{A}(M) = S(\mathcal{F}(M)) = S(\mathcal{B}(M))$

where $S(\mathcal{F}(M))$ is the paving of all $\mathcal{F}(M)$-Souslin sets [see Exc.1.91].

Hint: (i): Let A be an analytic subset of M, then by definition there exists a complete separable metric space (T, ρ) and a continuous function $f : T \to M$ with $A = f(T)$. Since T is separable, there exists sets $H(1), \ldots H(n), \ldots$ such that

$$T = \bigcup_{n=1}^{\infty} H(n) \quad \text{and} \quad \rho(t, u) < 2^{-1} \quad \forall n \geq 1 \, \forall t, u \in H(n)$$

Using the same argument on each of the sets $H(n)$, we can construct a Souslin scheme $H(n_1, \ldots, n_k)$ on T by induction such that

$$H(n_1, \ldots, n_k) = \bigcup_{q=1}^{\infty} H(n_1, \ldots, n_k, q) \quad \forall k \geq 1 \, \forall (n_1, \ldots, n_k) \in \mathbf{N}^k$$

$$\rho(t, u) < 2^{-k} \quad \forall k \geq 1 \, \forall (n_1, \ldots, n_k) \in \mathbf{N}^k \, \forall t, u \in H(n_1, \ldots, n_k)$$

Let $F(n_1, \ldots, n_k) = \mathrm{cl}\{f(H(n_1, \ldots, n_k))\}$, then $F(\cdot)$ is an $\mathcal{F}(M)$-Souslin scheme. Use the continuity of f and the completeness of T to show that $A = S(F)$, and argue that this proves (i). (ii): Let $A(\cdot) : \mathbf{N}^{(\mathbf{N})} \to \mathcal{A}(M)$ be an $\mathcal{A}(M)$-Souslin scheme. Then by Exc.1.98 and Exc.1.99 we have that the set

$$\Gamma = \bigcap_{k=1}^{\infty} \bigcup_{\lambda \in \mathbf{N}^k} A(\lambda_1, \ldots, \lambda_k) \times \{\sigma \in \mathbf{N}^{\mathbf{N}} \mid \sigma(j) = \lambda_j \quad 1 \leq j \leq k\}$$

is an analytic subset of $M \times \mathbf{N}^{\mathbf{N}}$. Show that $S(A) = \wp(\Gamma)$ where $\wp(x, \sigma) = x$ is the projection of $M \times \mathbf{N}^{\mathbf{N}}$ onto M. Conclude that this proves (ii). Finally, note that (iii) and (iv) follows from (i), (ii) and Exc.1.98 (iv).

Exc. 1.101: Let (M, d) and (L, ρ) be analytic metric spaces [see (P.5)] and let $f : M \to L$ be a Borel function. Show that $f(A)$ is an analytic subset of L, whenever A is an analytic subset of M.

Hint: Let Γ be the graph of f [see Exc.1.89] and let A be analytic subset of M. Set $C = \Gamma \cap (A \times L)$. Use Exc.1.89, Exc.1.98 (iii) and Exc.1.99 to show that C is an analytic subset of $M \times L$. Show that $f(A) = \wp(C)$ where $\wp(x, y) = y$ is the projection of $M \times L$ onto L.

Exc. 1.102: Let $f : \Omega \to \Lambda$ be a function, and let $\{A_n\}$ be a decreasing sequence of subsets of Ω such that $A_n \downarrow A$ and the paving: $\{A_n \cap f^{-1}(\lambda) \mid n \in \mathbf{N}\}$ is compact for every $\lambda \in \Lambda$. Show that

$$\text{(i)} \qquad f(A) = \bigcap_{n=1}^{\infty} f(A_n)$$

In his wrong proof of the false theorem "the projection of a Borel subset of the plane is a Borel subset of the line," Lebesgue actually claims that (i) is evidently true without any assumptions on the decreasing sequence $\{A_n\}$ when $\Omega = \mathbf{R}^2$, $\Lambda = \mathbf{R}$ and $f(x, y) = x$ is the projection [see (1.47)]. However, setting $A_n = \{(x, y) \mid y \geq n\}$ we see that the claim of Lebesgue is wrong because $A_n \downarrow \emptyset$, but $f(A_n) = \mathbf{R}$ for all $n \geq 1$. The result above is, in essence, due to M. M. Souslin, and the result is an essential tool in the solution to the image catastrophe [see (P.5.16) and (P.5.18)].

Hint: Let λ belong to the right-hand side of (i). Show that $A_n \cap f^{-1}(\lambda) \neq \emptyset$ for all $n \geq 1$ and conclude that $\bigcap_{n=1}^{\infty} \left(A_n \cap f^{-1}(\lambda)\right) \neq \emptyset$. Take an ω in this set and show that $\omega \in A$ and so $\lambda = f(\omega) \in f(A)$. The converse inclusion is trivial [see (P.3.5)].

Chapter 2

PROBABILITY MEASURES

2.1 Probability Spaces

After our small excursion into measure theory, let us return to probability theory to study consequences of the results we have obtained in Chapter 1. A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ where Ω is a set, $\mathcal{F}$ is a σ-algebra and P is a *probability measure* on $\mathcal{F}$, that is, $(\Omega, \mathcal{F}, P)$ satisfies the following three axioms:

(2.1.1) $\mathcal{F}$ is a σ-algebra of subsets of Ω

(2.1.2) P is a measure: $\mathcal{F} \to \mathbf{R}_+$

(2.1.3) $P(\Omega) = 1$

For reasons which I have explained in the introduction, these axioms have been taken as the model for intuitive probabilities by the vast majority of probabilists and statisticians since 1933. So let us consider a probability space $(\Omega, \mathcal{F}, P)$. Since P is a measure and a content, we know that all formulas (1.3.1)-(1.3.8), (1.4.1)-(1.4.6) and (1.19.2)-(1.19.6) hold with μ replaced by P. Moreover, since $P(\Omega) = 1$, by (1.3.2) and (1.3.3) we have

(2.1.4) $0 \leq P(F) \leq 1 \qquad \forall F \in \mathcal{F}$

(2.1.5) $P(F^c) = 1 - P(F) \qquad \forall F \in \mathcal{F}$

(2.1.6) $\lim_{n\to\infty} P(A_n) = 1 \quad \Rightarrow \quad \lim_{n\to\infty} P(F \cap A_n) = P(F) \quad \forall F \in \mathcal{F}$

For the rest of this book we will let $(\Omega, \mathcal{F}, P)$ denote a given probability space, and I will only rarely state that in the theorems, lemmas, examples, etc. to come.

2.2 Conditional Probabilities

Most real-life probabilities are conditioned on (often not stated) circumstances. For instance, when an engineer tells you that the probability that a safety valve in a nuclear reactor will fail is 10^{-5} per year, then this is conditioned on the assumption that the circumstances are normal, and you will find a quite different probability estimate under the extreme environment which may occur during a serious accident. Another example: Usually we assume that the probability of a "2" in one throw with a die equals $\frac{1}{6}$, but this is only true under the condition that the die is fair, and actual experiments may force us to reevaluate our opinion of the probability.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $A, B \in \mathcal{F}$. As you know from your elementary probability course, we define *the conditional probability of* A given B, denoted $P(A \mid B)$, as follows:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{provided that } P(B) > 0 \tag{2.2.1}$$

That is, $P(A \mid B)$ is the relative probability of $A \cap B$ compared to the probability of B. Intuitively $P(A \mid B)$ is interpreted as the probability that the event A will occur under the circumstances B. By (1.20.2) and (1.20.3) we have

(2.2.2) $F \curvearrowright P(F \mid B)$ is a probability measure on $\mathcal{F}$ for every fixed $B \in \mathcal{F}$ with $P(B) > 0$

In particular, $P(\cdot \mid B)$ have all the properties of a general probability measure. Let me recall some of the formulas for conditional probability which you have met in your elementary probability course

$$P\Big(\bigcap_{j=0}^{n} A_j \;\Big|\; A_0\Big) = \prod_{j=1}^{n} P(A_j | A_0 \cap \cdots \cap A_{j-1}) \tag{2.2.3}$$

$$P\Big(\bigcap_{j=0}^{n} A_j\Big) = P(A_0)P(A_1 \mid A_0) \cdots P(A_n \mid A_0 \cap \cdots \cap A_{n-1}) \tag{2.2.4}$$

These rules go under the name *the general multiplication laws*. Observing that $\cap_0^n A_j \downarrow \cap_0^\infty A_j$, we can extend these laws to infinitely many events:

$$P\left(\bigcap_{j=0}^{\infty} A_j \;\middle|\; A_0\right) = \prod_{j=1}^{\infty} P(A_j \mid A_0 \cap \cdots \cap A_{j-1}) \tag{2.2.5}$$

$$P\left(\bigcap_{j=0}^{\infty} A_j\right) = P(A_0) \prod_{j=1}^{\infty} P(A_j \mid A_0 \cap \cdots \cap A_{j-1}) \tag{2.2.6}$$

(Note that the infinite products exist since the factors are numbers between 0 and 1.) If $A_1, A_2, \cdots$ are disjoint and $B \subseteq \cup_j A_j$, we have

$$P(B) = \sum_j P(B \mid A_j)P(A_j) \tag{2.2.7}$$

$$P(A_n \mid B) = \frac{P(A_n)P(B \mid A_n)}{\sum_j P(B \mid A_j)P(A_j)} \tag{2.2.8}$$

Formula (2.2.7) goes under the name *the law of total probability*, and (2.2.8) is called *Bayes' rule*. Bayes' rule was not found by Thomas Bayes (1702 – 1761), but by Pierre-Simon de Laplace (1749 – 1827). However, Bayes did find, what nowadays usually is called *Laplace' deduction rule* - so they are even. The law of total probability (2.2.7) is useful in computing the probability of a "composite event," that is, an event B which depends on a series of "causes" $A_1, A_2, \cdots$. The formula tells us how to compute $P(B)$ when the probabilities of the causes $A_1, A_2, \cdots$ and the conditional probabilities of the event B given the cause A_j are known.

Bayes' rule (2.2.8) has been interpreted as a formula for "inverse probabilities": If $A_1, A_2, \cdots$ is a series of causes and B is a possible effect, then $P(B \mid A_j)$ is the probability of B when it is known that A_j is the cause, whereas $P(A_n \mid B)$ is the probability that A_n is the cause when it is known that B is the effect. Bayes has given name to *the Bayesian school of statisticians*. The Bayesians call $P(A_1), P(A_2), \cdots$ the *a priori probabilities*, that is, the probabilities of the causes before the experiment has taken place. Now perform the experiment and let B be the outcome of the experiment. Then on the basis of the experiment we reevaluate our a priori probabilities and obtain the *a posteriori probabilities* $P(A_1 \mid B), P(A_2 \mid B), \cdots$ which may be evaluated using (2.2.8).

2.3 Independence

It is well-known that two events A and B are called *independent* if $P(A \cap B) = P(A)P(B)$. This rule was the first rule ever to be established

in probability by Gerolamo Cardano (1501 – 1576). I shall now extend this notion. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A_1, \ldots, A_n \in \mathcal{F}$, then we say that $A_1, \ldots, A_n$ are *independent* if

$$(2.3.1) \qquad P\left(\bigcap_{j\in\alpha} A_j\right) = \prod_{j\in\alpha} P(A_j) \qquad \forall \alpha \subseteq \{1, \ldots, n\}\ ,\ \alpha \neq \emptyset$$

Since there are $2^n - 1$ non-empty subsets of $\{1, \ldots, n\}$, there are $2^n - 1$ equations in (2.3.1). However, if α is a singleton, for instance, $\alpha = \{3\}$, then (2.3.1) just claims that $P(A_3) = P(A_3)$, which is trivially true. Removing the singletons we are left with $2^n - n - 1$ nontrivial equations in (2.3.1), and it can be shown [see (2.16)] that not one single of these equations is a consequence of the others.

Let $\{A_t \mid t \in T\}$ be a finite or infinite family of events. Then we say that $\{A_t \mid t \in T\}$ are *independent* if $A_{t_1}, \ldots, A_{t_n}$ are independent whenever $t_1, \ldots, t_n$ are different elements in T and $n = 1, 2, \cdots$. Hence, $\{A_t \mid t \in T\}$ are independent if and only if

$$(2.3.2) \qquad P\left(\bigcap_{t\in\alpha} A_t\right) = \prod_{t\in\alpha} P(A_t) \qquad \forall \alpha \text{ finite } \subseteq T\ ,\ \alpha \neq \emptyset$$

Let $\mathcal{G}$ and $\mathcal{H}$ be two pavings contained in $\mathcal{F}$. Then we say that $\mathcal{G}$ and $\mathcal{H}$ are *independent* if G and H are independent whenever $G \in \mathcal{G}$ and $H \in \mathcal{H}$, that is, if

$$(2.3.3) \qquad P(G \cap H) = P(G)P(H) \qquad \forall G \in \mathcal{G} \quad \forall H \in \mathcal{H}$$

Similarly, the pavings $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are said to be *independent* if $G_1, \ldots, G_n$ are independent whenever $G_1 \in \mathcal{G}_1, \ldots, G_n \in \mathcal{G}_n$, that is, if

$$(2.3.4) \qquad P\left(\bigcap_{j\in\alpha} G_j\right) = \prod_{j\in\alpha} P(G_j)$$

whenever $\alpha \subseteq \{1, \ldots, n\}$, $\alpha \neq \emptyset$ and $G_j \in \mathcal{G}_j$ for all $j \in \alpha$. If $\{\mathcal{G}_t \mid t \in T\}$ is a family of pavings so that $\mathcal{G}_t \subseteq \mathcal{F}$ for all $t \in T$, we say that $\{\mathcal{G}_t \mid t \in T\}$ is *independent* if $\mathcal{G}_{t_1}, \ldots, \mathcal{G}_{t_n}$ are independent whenever $n \geq 1$ and $t_1, \ldots, t_n$ are different elements of T, that is, if

$$(2.3.5) \qquad P\left(\bigcap_{t\in\alpha} G_t\right) = \prod_{t\in\alpha} P(G_t)$$

whenever $\alpha \subseteq T$, α is finite and non-empty and $G_t \in \mathcal{G}_t$ for all $t \in \alpha$. Since $P(\Omega) = 1$ and $\Omega \cap A = A$ for all $A \subseteq \Omega$, we find

(2.3.6) Let $\mathcal{G}_1, \ldots, \mathcal{G}_n$ be pavings contained in $\mathcal{F}$ and set $\mathcal{G}_j^* = \{\Omega\} \cup \mathcal{G}_j$. Then $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are independent if and only if $\mathcal{G}_1^*, \ldots, \mathcal{G}_n^*$ are independent, and if and only if

$$P\left(\bigcap_{j=1}^{n} G_j\right) = \prod_{j=1}^{n} P(G_j) \quad \forall G_1 \in \mathcal{G}_1^* , \ldots , \forall G_n \in \mathcal{G}_n^*$$

If A and B are two independent events, we write $A \underline{\|} B$. Similarly, we write $\mathcal{G} \underline{\|} \mathcal{H}$ if $\mathcal{G}$ and $\mathcal{H}$ are two independent pavings, and we write $A \underline{\|} \mathcal{G}$ if A and G are independent for all $G \in \mathcal{G}$.

2.4 Independence Criterion for σ-Algebras I

Let $\mathcal{G}_1, \ldots, \mathcal{G}_n$ be $(\cap f)$-stable pavings all contained in $\mathcal{F}$ and let Q_j be a map from $\mathcal{G}_j$ into $\mathbf{R}$ for $1 \leq j \leq n$ satisfying

(2.4.1) $$P\left(\bigcap_{j=1}^{n} G_j\right) = \prod_{j=1}^{n} Q_j(G_j) \qquad \forall G_1 \in \mathcal{G}_1, \ldots, G_n \in \mathcal{G}_n$$

(2.4.2) $$\forall \varepsilon > 0 \; \exists H_1 \in \mathcal{G}_1, \cdots, H_n \in \mathcal{G}_n \;\; : \;\; P\left(\bigcap_{j=1}^{n} H_j\right) > 1 - \varepsilon$$

Then the σ-algebras $\sigma(\mathcal{G}_1), \ldots, \sigma(\mathcal{G}_n)$ are independent, and there exist real numbers $a_1, \ldots, a_n \in \mathbf{R}$ satisfying

(2.4.3) $$P(G) = a_j Q_j(G) \qquad \forall G \in \mathcal{G}_j \quad \forall j = 1, \ldots, n$$

(2.4.4) $$\prod_{j=1}^{n} a_j = 1$$

Proof: Let $p \in \mathbf{N}$ be a given integer. Then by (2.4.1) and (2.4.2) with $\varepsilon = 2^{-p}$ there exists $H_{1p} \in \mathcal{G}_1, \ldots, H_{np} \in \mathcal{G}_n$ satisfying

$$P(H_p) = \prod_{j=1}^{n} Q_j(H_{jp}) > 1 - 2^{-p}$$

where $H_p = H_{1p} \cap \cdots \cap H_{np}$. Let

$$b_p = \prod_{j=2}^{n} Q_j(H_{jp}) \quad , \quad c_p = Q_1(H_{1p}) \quad , \quad a_1 = \frac{b_1 \, P(H_{11})}{P(H_1)}$$

Since $b_p c_p > 0$ we have that $b_p \neq 0$. Let $G \in \mathcal{G}_1$; then by (2.4.1) and $(\cap f)$-stability of $\mathcal{G}_1$ we have

$$P(H_p \cap G) = P((H_{1p} \cap G) \cap H_{2p} \cap \cdots \cap H_{np}) = b_p Q_1(H_{1p} \cap G)$$

Setting $G = H_{1q}$, we find

$$\text{(i)} \qquad P(H_p \cap H_{1q}) = b_p Q_1(H_{1p} \cap H_{1q}) \qquad \forall p, q \in \mathbf{N}$$

Note that $P(H_{1q}) \to 1$ as $q \to \infty$; hence, setting $p = 1$ in (i), and letting $q \to \infty$, then by (2.1.6) we get

$$\lim_{q \to \infty} Q_1(H_{11} \cap H_{1q}) = \frac{P(H_1)}{b_1}$$

Moreover, setting $q = 1$ in (i) and letting $p \to \infty$, then by (2.1.6) we find

$$\lim_{p \to \infty} b_p = \lim_{p \to \infty} \frac{P(H_p \cap H_{11})}{Q_1(H_{1p} \cap H_{11})} = \frac{b_1 P(H_{11})}{P(H_1)} = a_1$$

Let $G \in \mathcal{G}_1$ be given and let $H^p = H_{2p} \cap \cdots \cap H_{np}$. Then by (2.1.6) we have that $P(G \cap H^p) \to P(G)$ as $p \to \infty$, and by (2.4.1) we have $P(G \cap H^p) = b_p Q_1(G)$. Letting $p \to \infty$, we see that $P(G) = a_1 Q_1(G)$ for all $G \in \mathcal{G}_1$. In exactly the same way, one shows that there exists $a_j \in \mathbf{R}$ such that $P(G) = a_j Q_j(G)$ for all $G \in \mathcal{G}_j$ and all $j = 1, \ldots, n$. Then we have

$$P(H_p) \prod_{j=1}^{n} a_j = \prod_{j=1}^{n} a_j Q_j(H_{jp}) = \prod_{j=1}^{n} P(H_{jp})$$

and since $P(H_p) \to 1$ and $P(H_{jp}) \to 1$ as $p \to \infty$ for all $j = 1, \ldots, n$, we see that (2.4.3) and (2.4.4) hold.

Let α be a non-empty subset of $\{1, \ldots, n\}$ and let $G_j \in \mathcal{G}_j$ for $j \in \alpha$. Set $G_{jp} = G_j$ if $j \in \alpha$ and $G_{jp} = H_{jp}$ if $j \notin \alpha$; then by (2.4.3), (2.4.4) and (2.4.1) we have

$$\prod_{j=1}^{n} P(G_{jp}) = \prod_{j=1}^{n} a_j Q_j(G_{jp}) = \prod_{j=1}^{n} Q_j(G_{jp}) = P\left(\bigcap_{j=1}^{n} G_{jp}\right)$$

And so inserting the G_{jp} 's, we find

$$\prod_{j \in \alpha} P(G_j) \prod_{j \notin \alpha} P(H_{jp}) = P\left(\bigcap_{j \in \alpha} G_j \cap \bigcap_{j \notin \alpha} H_{jp}\right)$$

and since

$$\lim_{p\to\infty} P(H_{jp}) = \lim_{p\to\infty} P\Big(\bigcap_{j\notin\alpha} H_{jp}\Big) = 1$$

then by (2.1.6) we find

$$\prod_{j\in\alpha} P(G_j) = P\Big(\bigcap_{j\in\alpha} G_j\Big)$$

for all non-empty sets $\alpha \subseteq \{1,\dots,n\}$ and all $G_j \in \mathcal{G}_j$ for $j \in \alpha$. Hence, $\mathcal{G}_1,\dots,\mathcal{G}_n$ are independent, and if we set $\mathcal{G}_j^* = \{\Omega\} \cup \mathcal{G}_j$, then by (2.3.6) we have that $\mathcal{G}_1^*,\dots\mathcal{G}_n^*$ are independent.

Let $G_2 \in \mathcal{G}_2^*,\dots,G_n \in \mathcal{G}_n^*$ be fixed for a moment and let

$$\mu(F) = P(F \cap G_2 \cap \cdots \cap G_n) \qquad \forall F \in \mathcal{F}$$
$$\nu(F) = P(F)\,P(G_2)\cdots P(G_n) \qquad \forall F \in \mathcal{F}$$

Then μ and ν are finite measures on $\mathcal{F}$ by (1.20.2) and (1.20.3), and $\mu(G) = \nu(G)$ for all $G \in \mathcal{G}_1^*$ because $\mathcal{G}_1^*,\dots,\mathcal{G}_n^*$ are independent. Hence, by the uniqueness theorem I (1.7) with $F_n = \Omega$ for all $n \geq 1$, we have $\mu(F) = \nu(F)$ for all $F \in \sigma(\mathcal{G}_1)$, that is,

$$P(F \cap G_2 \cap \cdots \cap G_n) = P(F)P(G_2)\cdots P(G_n)$$

for all $F \in \sigma(\mathcal{G}_1)$ and all $G_2 \in \mathcal{G}_2^*, \dots, G_n \in \mathcal{G}_n$. Thus, by (2.3.6) we have that $\sigma(\mathcal{G}_1), \mathcal{G}_2^*,\dots,\mathcal{G}_n^*$ are independent. Repeating this argument n times we see that $\sigma(\mathcal{G}_1),\sigma(\mathcal{G}_2),\dots,\sigma(\mathcal{G}_n)$ are independent and so the theorem is proved. □

2.5 Independence Criterion for σ-Algebras II

Let $\{\mathcal{G}_t \mid t \in T\}$ be a family of independent $(\cap f)$-stable pavings. Then the σ-algebras $\{\sigma(\mathcal{G}_t) \mid t \in T\}$ are independent. Moreover, if $\{T_q \mid q \in Q\}$ is a family of disjoint subsets of T and if

$$\mathcal{H}_q = \bigcup_{t\in T_q} \mathcal{G}_t, \qquad \mathcal{F}_q = \sigma(\mathcal{H}_q) \qquad \forall q \in Q$$

Then $\{\mathcal{F}_q \mid q \in Q\}$ are independent.

Proof: Let $\mathcal{G}_t^* = \mathcal{G}_t \cup \{\Omega\}$, then evidently $\mathcal{G}_t^*$ is $(\cap f)$-stable. By (2.3.6) we have that $\{\mathcal{G}_t^* \mid t \in T\}$ are independent. Hence, by (2.4) we have that $\{\sigma(\mathcal{G}_t) \mid t \in T\}$ are independent. Let

$$\hat{\mathcal{H}}_q = \{H_1 \cap \cdots \cap H_n \mid n \geq 1 \ , \ H_j \in \mathcal{H}_q \quad \forall j = 1, \cdots, n\}$$

for all $q \in Q$. Let α be a finite non-empty subset of Q and let $\hat{H}_q \in \hat{\mathcal{H}}_q$ for each $q \in \alpha$. Since $\mathcal{G}_t$ is $(\cap f)$-stable for all t , there exist finite sets $\alpha_q \subseteq T_q$ and $G_t \in \mathcal{G}_t$ for $t \in \alpha_q$ and $q \in \alpha$ such that

$$\hat{H}_q = \bigcap_{t \in \alpha_q} G_t$$

Let $\beta = \cup_{q \in \alpha} \alpha_q$; then β is a finite subset of T and $G_t \in \mathcal{G}_t$ for $t \in \beta$. Hence, by independence of $\{\mathcal{G}_t\}$ we have

$$\begin{aligned} P\Big(\bigcap_{q \in \alpha} \hat{H}_q\Big) &= P\Big(\bigcap_{t \in \beta} G_t\Big) = \prod_{t \in \beta} P(G_t) \\ &= \prod_{q \in \alpha} \prod_{t \in \alpha_q} P(G_t) = \prod_{q \in \alpha} P(\hat{H}_q) \end{aligned}$$

Thus, $\{\hat{\mathcal{H}}_q \mid q \in Q\}$ are independent and clearly $\hat{\mathcal{H}}_q$ is $(\cap f)$-stable for all $q \in Q$. Thus, the second part follows from the first part. □

2.6 Independence Criterion for σ-Algebras III

Let $\{A_t \mid t \in T\}$ be a family of independent events and let $\mathcal{A}_t = \{\emptyset, A_t, A_t^c, \Omega\}$. Then $\{\mathcal{A}_t \mid t \in T\}$ are independent. Moreover, if $\{T_q \mid q \in Q\}$ is a family of disjoint subsets of T , and if $\mathcal{F}_q = \sigma(A_t \mid t \in T_q)$, then $\{\mathcal{F}_q \mid q \in Q\}$ are independent.

Proof: Apply (2.5) with $\mathcal{G}_t = \{A_t\}$. □

The two criteria above are very useful in establishing independence of composite events. For instance,

(2.6.1) If $A_1, A_2, \cdots$ are independent, then so are $A_1^c, A_2^c, \cdots$

and if $A_1, A_2, \cdots$ are independent, then for instance, $(A_1 \setminus A_2) \cup A_3$, $A_4 \cup A_5$, A_9 and $(A_6 \cup A_7^c) \cap A_8$ are independent.

2.7 Random Variables, Vectors and Functions

The notion of random variables and functions is the cardinal point, around which all probability theory is turning. It is fair to say that probability is devoted to the study of random variables and functions. So it is about time to introduce them. As usual, let $(\Omega, \mathcal{F}, P)$ be a given probability space. Then a *random variable* is a measurable function X from Ω into $\mathbf{R}$ and a *random vector* is a measurable function $X = (X_1, \ldots, X_k)$ from Ω into $\mathbf{R}^k$. Moreover, if $(M, \mathcal{B})$ is a measurable space, then an *M-valued random function* is a measurable function from Ω into M . In Chapter 1 we have studied measurable functions and, of course, we have all the results from that chapter available for random variables, vectors and functions. For instance, by (1.43.3) we have

(2.7.1) $X = (X_1, \ldots, X_k) : \Omega \to \mathbf{R}^k$ is a random vector if and only if $X_1, \ldots, X_k$ are random variables

By (1.38.6) and (1.10.1), the following five statements are equivalent:

(2.7.2) X is a random variable

(2.7.3) $\{X \leq a\} \in \mathcal{F} \qquad \forall a \in \mathbf{R}$

(2.7.4) $\{X < a\} \in \mathcal{F} \qquad \forall a \in \mathbf{R}$

(2.7.5) $\{X \geq a\} \in \mathcal{F} \qquad \forall a \in \mathbf{R}$

(2.7.6) $\{X > a\} \in \mathcal{F} \qquad \forall a \in \mathbf{R}$

If the σ-algebra $\mathcal{F}$ is "small," then measurability of a function $X : \Omega \to \mathbf{R}$ is a restrictive assumption. For instance, $\mathcal{F} = \{\emptyset, \Omega\}$; then only constant functions will be measurable. However, the σ-algebra $\mathcal{F}$ will usually be so large that it takes a weird fantasy to imagine a function from Ω into $\mathbf{R}$ which is not a random variable. For all practical purposes you may assume that all the real-valued functions which you ever will meet are random variables.

2.8 Distribution Functions and Laws

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let X be a random function from Ω into the measurable space $(M, \mathcal{B})$. In (1.44.1), I defined the image measure P_X by

$$P_X(B) = P(X^{-1}(B)) = P(X \in B) \qquad \forall B \in \mathcal{B} \tag{2.8.1}$$

and in (1.44.2), I showed that P_X is a probability measure on $(M, \mathcal{B})$. The image measure P_X is called *the distribution* or *the distribution law* of X. Note that if X is a random variable, then P_X is a Borel probability measure on $\mathbf{R}$, and if X is a k-dimensional random vector, then P_X is a Borel probability measure on $\mathbf{R}^k$. Let X be a random variable; then we define *the distribution function* F_X of X as follows:

$$F_X(v) = P(X \leq v) \qquad \forall v \in \mathbf{R} \tag{2.8.2}$$

The distribution function F_X is a function: $\mathbf{R} \to \mathbf{R}$ satisfying

(2.8.3) $\quad F_X$ is increasing and right continuous

$$\lim_{v \to \infty} F_X(v) = 1 \quad , \quad \lim_{v \to -\infty} F_X(v) = 0 \tag{2.8.4}$$

$$P_X(B) = P(X \in B) = \lambda_{F_X}(B) \qquad \forall B \in \mathcal{B}(\mathbf{R}) \tag{2.8.5}$$

where λ_{F_X} is the Lebesgue-Stieltjes measure induced by F_X [see (1.32)].

(2.8.3): By (1.3.2) we have that F_X is increasing, and if $v_n \downarrow v$, then $\{X \leq v_n\} \downarrow \{X \leq v\}$ and so $F_X(v_n) \to F_X(v)$ by (1.4.3).

(2.8.4): Follows from (1.4.2) and (1.4.3) because $\{X \leq v_n\} \uparrow \Omega$ if $v_n \uparrow \infty$ and $\{X \leq v_n\} \downarrow \emptyset$ if $v_n \downarrow -\infty$.

(2.8.5): Note that by (1.3.3) we have

$$\begin{aligned} P_X(\,]a,b]\,) &= P(a < X \leq b) = P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \end{aligned}$$

Hence, P_X is a Borel measure on $\mathbf{R}$ satisfying (1.32.2), and since the Lebesgue-Stieltjes measure induced by F_X is the only measure with this property we see that (2.8.5) holds.

Let us turn things upside down and suppose that $F : \mathbf{R} \to \mathbf{R}$ is a given function satisfying (2.8.3) and (2.8.4); that is, F is increasing, right continuous, $F(\infty) = 1$ and $F(-\infty) = 0$ where $F(\infty)$ and $F(-\infty)$ are defined in (1.32.4). Then the Lebesgue-Stieltjes measure λ_F is a probability measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ by (1.32.3). Hence, $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \lambda_F)$ is a probability space, and if we

define $X(u) = u$, then X is a random variable such that $F_X = F$. Hence, we have proved the following:

(2.8.6) If $F : \mathbf{R} \to \mathbf{R}$ satisfies (2.8.3) and (2.8.4), then F is the distribution function of some random variable X on some probability space

Let $X = (X_1, \ldots, X_k)$ be a k-dimensional random vector. Then we define *the distribution function* F_X of X in a similar manner

(2.8.7) $$F_X(v_1, \ldots, v_k) = P(X_1 \le v_1, \ldots, X_k \le v_k)$$

for all $(v_1, \ldots, v_k) \in \mathbf{R}^k$. Then F_X is a function: $\mathbf{R}^k \to \mathbf{R}$ satisfying

(2.8.8) F_X is right continuous and has positive increments

(2.8.9) $$F_X(\infty, \infty, \ldots, \infty) = 1 \qquad F_X(v_1, \ldots, v_{j-1}, -\infty, v_{j+1}, \ldots, v_k) = 0$$

(2.8.10) $$P_X(B) = P(X \in B) = \lambda_{F_X}(B) \qquad \forall B \in \mathcal{B}(\mathbf{R}^k)$$

This is proved in essentially the same way as above, except that it is a bit more involved to prove that F_X has positive increments: Let $(a_1, \ldots, a_k) \le (b_1, \ldots, b_k)$ and set

$$B_j = \{X_j \le b_j\} \quad , \quad B = \bigcap_{j=1}^{n} B_j \quad , \quad A_j = \{X_j \le a_j\}$$

Then by the inclusion-exclusion principle [see Exc.2.25] we find

$$\begin{aligned} P(a < X \le b) &= P\Big(\bigcap_{j=1}^{k} (B_j \setminus A_j)\Big) = P\Big(B \setminus \bigcup_{j=1}^{k} (B \cap A_j)\Big) \\ &= P(B) - P\Big(\bigcup_{j=1}^{k} (B \cap A_j)\Big) \\ &= P(B) + \sum_{j=1}^{k} \sum_{1 \le \nu_1 < \cdots < \nu_j \le k} (-1)^j P\big(B \cap A_{\nu_1} \cap \cdots \cap A_{\nu_j}\big) \end{aligned}$$

and inserting $A_1, \ldots A_k$ and B , we get

(2.8.11) $$P(a_1 < X_1 \le b_1, \ldots, a_k < X_k \le b_k) = \Delta_a^b F_X$$

Thus, F_X has positive increments. Exactly as above, we find

(2.8.12) If $F : \mathbf{R}^k \to \mathbf{R}$ satisfies (2.8.8) and (2.8.9), then F is the distribution function of some random vector $X = (X_1, \ldots, X_k)$ on some probability space

This justifies the following definition: A function $F : \mathbf{R}^k \to \mathbf{R}$ is called a *distribution function* if F satisfies (2.8.8) and (2.8.9). Thus, F is a distribution function if and only if F is the distribution function of some random vector.

Let $X = (X_1, \ldots, X_k)$ and let $1 \leq \alpha(1) < \cdots < \alpha(m) \leq k$ be given. Then $Y = (X_{\alpha(1)}, \ldots, X_{\alpha(m)})$ is an m-dimensional random vector, and since $\{X_j \leq v\} \uparrow \Omega$ as $v \uparrow \infty$, by (1.4.2) we conclude that

$$F_Y(u_1, \ldots, u_m) = F_X(\infty, \ldots, u_1, \infty, \ldots, u_2, \infty, \cdots, u_m, \infty \cdots) \tag{2.8.13}$$

where u_1 sits on the $\alpha(1)^{\text{th}}$ coordinate, u_2 sits on the $\alpha(2)^{\text{th}}$ coordinate, etc., and we have taken the limit as all the other coordinates tend to $+\infty$. The distribution function F_Y is called *the* $(\alpha(1), \ldots, \alpha(m))^{\text{th}}$ *marginal distribution function of* X. Often we call F_X *the simultaneous distribution function of* X in order to distinguish it from the $2^k - 2$ possible marginal distribution functions.

Let X and Y be two random functions taking values in the same measurable space $(M, \mathcal{B})$. If $P_X = P_Y$ we say that X and Y are *identically distributed*, or that Y is a *version* of X and we write $X \sim Y$. Similarly, if μ is a probability measure on $(M, \mathcal{B})$, then we say that μ is *the law* of X if $P_X = \mu$ and we write $X \sim \mu$. Note that if f is a measurable function from (M, B) into $(L, \mathcal{A})$, then by (1.44.3) we have

(2.8.14) If $X \sim \mu$, then $f(X) \sim \mu_f$

(2.8.15) If $X \sim Y$, then $f(X) \sim f(Y)$

Similarly, if $F : \mathbf{R}^k \to \mathbf{R}$ is a distribution function and $X = (X_1, \ldots, X_k)$ is a k-dimensional random vector, then we write

$$X \sim F \iff F_X = F \tag{2.8.16}$$

Note that by (2.8.10) we have

$$X \sim Y \iff F_X = F_Y \tag{2.8.17}$$

whenever X and Y are k-dimensional random vectors.

2.9 Independent Random Functions

Let X and Y be two random functions defined on the same probability space $(\Omega, \mathcal{F}, P)$ but with values in possibly different measurable spaces, say $(M, \mathcal{A})$ and $(L, \mathcal{B})$. Then we say that X and Y are *independent*, and write $X \perp\!\!\!\perp Y$ if the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent, that is, if

$$(2.9.1) \qquad P(X \in A\,,\ Y \in B) = P(X \in A)P(Y \in B) \quad \forall A \in \mathcal{A}\ \forall B \in \mathcal{B}$$

Similarly, if $\mathcal{G}$ is a paving on Ω contained in $\mathcal{F}$, we say that X and $\mathcal{G}$ are *independent*, and write $X \perp\!\!\!\perp \mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent, that is, if

$$(2.9.2) \qquad P(G \cap \{X \in A\}) = P(G)\,P(X \in A) \qquad \forall G \in \mathcal{G}\ \forall A \in \mathcal{A}$$

More generally if X_t is a random function from Ω into a measurable space $(M_t, \mathcal{B}_t)$ for all t in the set T, then we say that $\{X_t \mid t \in T\}$ are *independent* if the σ-algebras $\{\sigma(X_t) \mid t \in T\}$ are independent, that is, if

$$(2.9.3) \qquad P(X_t \in B_t \ \ \forall t \in \alpha) = \prod_{t \in \alpha} P(X_t \in B_t)$$

whenever α is a non-empty finite subset of T and $B_t \in \mathcal{B}_t$ for all $t \in \alpha$. From (2.4) and (2.5) we obtain the following corollary.

2.10 Independence Criterion for Random Functions

Let $X = (X_1, \ldots, X_k)$ be a k-dimensional random vector. If F is the distribution function of X and F_j is the marginal distribution function of X_j for $j = 1, \ldots, k$, then the following four statements are equivalent:

(2.10.1) $X_1, \ldots, X_k$ *are independent*

$$(2.10.2) \qquad P\left(\bigcap_{j=1}^{k} \{X_j \in B_j\}\right) = \prod_{j=1}^{k} P(X_j \in B_j) \quad \forall B_1, \ldots, B_k \in \mathcal{B}(\mathbf{R})$$

$$(2.10.3) \qquad F(v_1, \ldots, v_k) = \prod_{j=1}^{k} F_j(v_j) \qquad \forall (v_1, \ldots, v_k) \in \mathbf{R}^k$$

(2.10.4) *There exist functions $H_j : \mathbf{R} \to \mathbf{R}$ for $j = 1, \ldots, k$ satisfying*

$$F(v_1, \ldots, v_k) = \prod_{j=1}^{k} H_j(v_j) \qquad \forall (v_1, \ldots, v_k) \in \mathbf{R}^k$$

Moreover, suppose that X_t *is a random function from* Ω *into the measurable space* $(M_t, \mathcal{B}_t)$ *for all* t *in the set* T *. Let* $\mathcal{G}_t \subseteq \mathcal{B}_t$ *be a* $(\cap f)$*-stable paving on* M_t *such that* $\sigma(\mathcal{G}_t) = \mathcal{B}_t$ *for all* $t \in T$ *, and suppose the pavings* $\{X_t^{-1}(\mathcal{G}_t) \mid t \in T\}$ *are independent. Then we have*

(2.10.5) *The random functions* $\{X_t \mid t \in T\}$ *are independent*

(2.10.6) *If* $Y_t = f_t(X_t)$ *where* $f_t : (M_t, \mathcal{B}_t) \to (L_t, \mathcal{A}_t)$ *is measurable for all* $t \in T$ *, then* $\{Y_t \mid t \in T\}$ *are independent*

(2.10.7) *If* $\{T_q \mid q \in Q\}$ *are disjoint subsets of* T *and* $\mathcal{F}_q = \sigma(X_t \mid t \in T_q)$ *for* $q \in Q$ *, then the* σ*-algebras* $\{\mathcal{F}_q \mid q \in Q\}$ *are independent*

□

2.11 The Two Borel-Cantelli Lemmas

Let $A_1, A_2, \cdots$ *be a sequence of events. Then we have*

(2.11.1) $P(\limsup A_n) = 0$ *if* $\sum_{n=1}^{\infty} P(A_n) < \infty$

(2.11.2) $P(\limsup A_n) = 1$ *if* $\sum_{n=1}^{\infty} P(A_n) = \infty$ *and* $\{A_n\}$ *are independent*

These two results are called *The first* and *the second Borel-Cantelli lemma*, respectively.

Proof: (2.11.1) is simply a restatement of (1.4.6). So suppose that $A_1, A_2, \cdots$ are independent and that $\sum P(A_n) = \infty$. Let $1 \le k \le m$ be given. Since $A_1^c, A_2^c, \cdots$ are independent by (2.6.1), we have

$$P\left(\bigcup_{j=k}^{m} A_j\right) = 1 - P\left(\bigcap_{j=k}^{m} A_j^c\right) = 1 - \prod_{j=k}^{m} P(A_j^c)$$
$$= 1 - \prod_{j=k}^{m} (1 - P(A_j)) \ge 1 - \exp\left(-\sum_{j=k}^{m} P(A_j)\right)$$

where we have used the inequality $1 - x \le e^{-x}$. Let us fix k . Then by assumption we know that the sum under the exp converges to infinity as $m \to \infty$. Hence if $A^k = A_k \cup A_{k+1} \cup \cdots$, then letting $m \to \infty$ we see that $P(A^k) = 1$ for all $k \ge 1$ and since $A^k \downarrow \limsup A_n$, we see that (2.11.2) holds. □

In particular, we see that if $A_1, A_2, \cdots$ are independent, then $\limsup A_n$ either have probability 0 or 1. This is a special case of the so-called 0–1 law which, loosely speaking, states that all events in the infinitely distant future have probability either 0 or 1.

2.12 The 0–1 Law for Pavings

Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$ be an increasing sequence of $(\cap f)$-stable pavings on Ω and let $\mathcal{F}^\infty$ be a $(\cap f)$-stable paving satisfying

(2.12.1) $\mathcal{F}_n$ *and* $\mathcal{F}^\infty$ *are independent for all* $n \geq 1$

(2.12.2) $$\mathcal{F}^\infty \subseteq \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n\right)$$

Then for every $F \in \sigma(\mathcal{F}^\infty)$ we have either $P(F) = 0$ or $P(F) = 1$.

Proof: Let $\mathcal{F}_\infty = \cup_1^\infty \mathcal{F}_n$. Then $\mathcal{F}_\infty$ is $(\cap f)$-stable, and by (2.12.1) we have that $\mathcal{F}^\infty$ and $\mathcal{F}_\infty$ are independent. Hence, by (2.5) we have that $\sigma(\mathcal{F}^\infty)$ and $\sigma(\mathcal{F}_\infty)$ are independent. Let $F \in \sigma(\mathcal{F}^\infty)$, then by (2.12.2) we have that $F \in \sigma(\mathcal{F}_\infty)$, and so F is independent of itself. This means that $P(F) = P(F \cap F) = P(F)^2$. But the equation: $x^2 = x$ only has the two solutions $x = 0$ and $x = 1$. Hence, the theorem is proved. □

2.13 The 0–1 Law for Random Functions

Let $X_1, X_2, \cdots$ be a sequence of independent random functions and set

$$\mathcal{F}^\infty = \bigcap_{n=1}^{\infty} \sigma(X_j \mid j \geq n)$$

Then for every $F \in \mathcal{F}^\infty$, we have either $P(F) = 0$ or $P(F) = 1$. Moreover, if Y is a random variable such that Y is $\mathcal{F}^\infty$-measurable, then Y is degenerated; that is, there exist $a \in \mathbf{R}$ such that $Y = a$ P-a.s.

Remarks: (1): A random function $Y : \Omega \to (M, \mathcal{B})$ is called *degenerated* if it is constant a.s., that is, if there exist $a \in M$ such that $P(Y = a) = 1$.

(2): Let $X_1, X_2, \cdots$ be independent random variables. Then the 0–1 law states that any event, which only depends on the infinitely distant future either is almost sure or almost impossible. For instance, all the events

$$F_1 = \left\{ \lim_{n\to\infty} X_n \text{ exists} \right\} \quad , \quad F_2 = \left\{ \sup_n |X_n| < \infty \right\}$$

$$F_3 = \left\{ \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j \text{ exists} \right\} \quad , \quad F_4 = \left\{ \sum_{1}^{\infty} X_n \text{ converges} \right\}$$

have probability either 0 or 1 . Similarly, any function of the X_n 's which only depends on the infinitely far away X_n 's will be constant a.s. For instance, all the random variables:

$$Y_1 = \limsup_{n\to\infty} X_n \quad , \quad Y_2 = \liminf_{n\to\infty} X_n$$

$$Y_3 = \limsup_{n\to\infty} \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j \right\} \quad , \quad Y_4 = \liminf_{n\to\infty} \left\{ \frac{1}{n} \sum_{j=1}^{n} X_j \right\}$$

will be degenerated, that is, constant a.s. Intuitively, this means that in the long run randomness will disappear and, after an infinitely long time, randomness will be replaced by certainty.

Proof: Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and $\mathcal{F}^n = \sigma(X_j \mid j \geq n+1)$. Then $\mathcal{F}_n$ and $\mathcal{F}^n$ are independent by (2.10), and since $\mathcal{F}^\infty \subseteq \mathcal{F}^n$, we see that $\mathcal{F}^\infty$ and $\mathcal{F}_n$ are independent for all $n \geq 1$. Moreover,

$$\mathcal{F}^\infty \subseteq \mathcal{F}^0 = \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{F}_n \right)$$

Hence, by (2.12) we have $P(F) = 0$ or 1 for all $F \in \mathcal{F}^\infty$. Let Y be an $\mathcal{F}^\infty$-measurable random variable and let F be the distribution function of Y . Since $\{Y \leq v\} \in \mathcal{F}^\infty$, we see that $F(v) = 0$ or 1 for all $v \in \mathbf{R}$ and since $F(-\infty) = 0$, $F(\infty) = 1$ and F is increasing and right continuous, we see that there exists $a \in \mathbf{R}$ such that

$$F(v) = \begin{cases} 1 & \text{if } v \geq a \\ 0 & \text{if } v < a \end{cases}$$

Hence, $P(a - \varepsilon < Y \leq a) = 1$ for all $\varepsilon > 0$. Letting $\varepsilon \to 0$, we see that $P(Y = a) = 1$. Thus, Y is degenerated. □

2.14 Probability Mass Functions

Let X be a random function with values in the measurable space $(M, \mathcal{B})$, and suppose that $\mathcal{B}$ contains all singletons. Then we define *the probability mass function* p_X of X by

(2.14.1) $$p_X(v) = P(X = v) \qquad \forall v \in M$$

Then p_X is a function from M into $[0,1]$. By σ-additivity we have

(2.14.2) $$P(X \in Q) = \sum_{q \in Q} p_X(q) \qquad \forall Q \text{ countable } \subseteq M$$

In particular, $\{v \in M \mid p_X(v) \geq \frac{1}{n}\}$ has at most n elements for all $n = 1, 2, \cdots$. Hence, if we define the *discrete support* of X, denoted $sp_+(X)$, by

$$sp_+(X) = \{v \in M \mid P(X = v) > 0\}$$

Then we have

(2.14.3) $$sp_+(X) \text{ is countable}$$

(2.14.4) $$\sum_{v \in B} p_X(v) = P(X \in B \cap sp_+(X)) \leq 1 \qquad \forall B \in \mathcal{B}$$

Recall that X is called *degenerated* if X is constant a.s., that is, if $P(X = a) = 1$ for some $a \in M$. Clearly, we have that the following three statements are equivalent:

(2.14.5) $$P(X = a) = 1$$

(2.14.6) $$p_X(a) = 1 \text{ and } p_X(v) = 0 \qquad \forall v \neq a$$

(2.14.7) $$P_X = \delta_a$$

where δ_a is the Dirac measure at a [see (1.15)].

As in (1.45) we say that the M-valued random function X is *simple* if there exists a finite set $F \subseteq M$ such that $P(X \in F) = 1$. Clearly, we have:

(2.14.8) $$X \text{ is simple} \iff sp_+(X) \text{ is finite and } \sum_{v \in M} p_X(v) = 1$$

Similarly, we say that X is *discrete* if there exists a countable set $Q \subseteq M$ such that $P(X \in Q) = 1$. By (2.14.1)-(2.14.4) we get

(2.14.9) $$X \text{ is discrete} \iff \sum_{v \in M} p_X(v) = 1$$

Note that if X is discrete, then by (2.14.1)-(2.14.4) we have

(2.14.10) $$P(X \in B) = P_X(B) = \sum_{v \in B} p_X(v) \qquad \forall B \in \mathcal{B}$$

Hence if X is discrete, then p_X determines uniquely the distribution law P_X by formula (2.14.10). We say that X is *diffuse* if $P(X = v) = 0$ for all $v \in M$, that is, if $p_X(v) = 0$ for all $v \in M$. Clearly, we have

(2.14.11) X is diffuse if and only if $sp_+(X) = \emptyset$

Suppose that X is a real random variable, and let $v \in \mathbf{R}$. Since $]v - 2^{-n}, v] \downarrow \{v\}$, we have

(2.14.12) $$p_X(v) = F_X(v) - F_X(v-)$$

where we define $F(v-)$ by

(2.14.13) $$F(v-) = \lim_{x \uparrow v} F(x) = \sup_{x<v} F(x)$$

whenever F is an increasing function. Note that (2.14.12) shows that $p_X(v)$ is *the jump* of F_X at v, and so we have

(2.14.14) $$sp_+(X) = \{v \in \mathbf{R} \mid F_X \text{ is discontinuous at } v\}$$

(2.14.15) X is diffuse if and only if F_X is continuous

If $X = (X_1, \ldots, X_k)$ is a k-dimensional random vector with $k \geq 2$, then the situation becomes more complicated [see Exc.2.35].

2.15 Lottery

The notion of independence as defined here goes back to Cardano (1501 – 1576), who found the multiplication law for finitely many events. The concept is of course modeled over our intuitive feeling for independent repetitions of an experiment. However, the mathematical definition (2.3.1) also makes sense for events defined only on the basis of one experiment, and this can sometimes give seemingly strange result which can be difficult to understand on an intuitive level. Let us look at an example.

We are joining a lottery with N lots enumerated $1,\dots,N$. The winning lot will be drawn in a fair way. Then $\Omega = \{1,\dots,N\}$, and if $A \subseteq \Omega$, the probability that the winning lot belongs to A is given by

$$(2.15.1) \qquad P(A) = \frac{\#A}{N} \qquad \forall A \subseteq \Omega$$

Let A and B be two sets with $1 \le k, m < N$, where $m = \#A$ and $k = \#B$, and set $n = \#(A \cap B)$. Then the events A and B are independent if and only if $n \cdot N = k \cdot m$. If N is a prime number, this means that k and m are divisors of n , but $n \le k$ and $n \le m$ so this is only possible if $k = n = m$; but then $N = k \le N - 1$ which is impossible. Hence, if N is a prime number and A and B are independent, then either A or B equals $\emptyset$ or Ω . So in this case there are no nontrivial independent events. If N is not a prime number, then $N = k \cdot m$ for some integers $k, m \ge 2$. Set $A = \{1,\dots,m\}$ and $B = \{m, 2m, 3m, \dots, km\}$. Then we have

$$P(A) = \tfrac{1}{k} \quad , \quad P(B) = \tfrac{1}{m} \quad , \quad P(A \cap B) = \tfrac{1}{N} = P(A)\,P(B)$$

so A and B are independent. Many other nontrivial independent pairs may be constructed in the same way. Hence if

$$N = 2^{127} - 1 = 170141183460469231731687303715884105727$$

then there are no nontrivial pairs of independent events, and if

$$N = 2^{127} = 170141183460469231731687303715884105728$$

then there are lots of nontrivial pairs of independent events.

This example does not imply that our definition of independence is wrong, but it shows that the mathematical definition of independence does not fully cover our intuitive concept of independence.

PS: In 1876, Lucas proved that the number $2^{127} - 1$ is a prime number (without a computer!).

2.16 Pairwise Independence

Let $A_1, A_2, \cdots$ be events; then we say that $A_1, A_2, \cdots$ are *pairwise independent* if A_i and A_j are independent for all $i \ne j$, that is, if

$$(2.16.1) \qquad P(A_i \cap A_j) = P(A_i)P(A_j) \qquad \forall i \ne j$$

and we define *pairwise independence* for pavings or random functions similarly. It is a common mistake to believe that pairwise independence implies independence. ***This is not the case!*** Let us consider the lottery model from (2.15), and set $N = 4$ and

$$A = \{1,2\}, \qquad B = \{2,3\}, \qquad C = \{1,3\}$$

Then A, B and C are pairwise independent but not independent. Another common mistake is to believe that $A_1, \ldots, A_n$ are independent if

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$$

This is also wrong! Set $N = 8$ in the lottery model and

$$A = \{1,2,3,4\},\ B = \{1,2,3,5\},\ C = \{1,6,7,8\}$$

Then $P(A) = P(B) = P(C) = \frac{1}{2}$ and $P(A \cap B \cap C) = \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$, but $P(A \cap B) = \frac{3}{8} \neq P(A)P(B)$. So A and B are not independent. By similar examples one can show that every one of the $2^n - n - 1$ nontrivial equations in (2.3.1) is needed, and not one of them can be left out.

2.17 Maximal Inequalities

Let $U_1, U_2, \cdots$ be a sequence of real random variables and consider the random variables

$$M_n = \max_{1 \le j \le n} U_j, \qquad M = \sup_{j \in \mathbf{N}} U_j = \lim_{n \to \infty} M_n$$

$$U = \limsup_{j \to \infty} U_j = \lim_{n \to \infty} \left(\sup_{j \ge n} U_j \right)$$

A *maximal inequality* is then an upper estimate for the tail probabilities $P(M_n > x)$ or $P(M > x)$ or $P(U > x)$. Since $\{U > x\} \subseteq \limsup\{U_n > x\}$ and

$$\{M_n > x\} = \bigcup_{j=1}^{n} \{U_j > x\} \ , \ \{M > x\} = \bigcup_{j=1}^{\infty} \{U_j > x\}$$

we obtain the following maximal inequalities from (1.4.1) and (1.4.6)

(2.17.1) $$P(M_n > x) \le \sum_{j=1}^{n} P(U_j > x) \qquad \forall x \in \mathbf{R}$$

(2.17.2) $$P(M > x) \le \sum_{j=1}^{\infty} P(U_j > x) \qquad \forall x \in \mathbf{R}$$

(2.17.3) $$P(U > x) = 0 \quad \text{if} \quad \sum_{j=1}^{\infty} P(U_j > x) < \infty$$

Maximal inequalities are often used to establish the a.s. convergence of a sequence $\{U_n\}$ of random variables in the following way: Suppose we can prove that for some sequence $\sigma(1) < \sigma(2) < \cdots$ of integers we have $U_{\sigma(j)} \to U$ a.s. Then a.s. convergence of $\{U_n\}$ to U can often be established by using a maximal inequality to estimate the probability:

$$P\left(\max_{\sigma(j-1)<n<\sigma(j)} U_n > x \right)$$

However, the general maximal inequalities (2.17.1)-(2.17.2) are too crude to be of much use. So let us impose some additional assumptions. By the second Borel-Cantelli lemma (2.11.2) it follows easily that we have

(2.17.4) If $U_1, U_2, \cdots$ are independent, then $\limsup_{n\to\infty} U_n = c$ a.s., where

$$c = \sup\left\{ x \in \mathbf{R} \;\middle|\; \sum_{j=1}^{\infty} P(U_n > x) = \infty \right\}$$

Moreover, we have the following maximal inequalities which will turn out be a crucial point in the proof of the law of large numbers [see (6.8)-(6.12)]:

2.18 Maximal Inequalities for Subadditive Schemes

Suppose that $\{U_j\}, \{V_j\}$ *and* $\{W_j\}$ *are three sequences of random variables satisfying*

(2.18.1) $U_j \le V_j + W_j \qquad \forall j \ge 1$

and let us set

$$M_n = \max\{U_1, \ldots, U_n\} \quad , \quad L_n = \max\{W_1, \ldots, W_n\} \quad , \quad M = \sup_{j\ge1} U_j$$

$$L = \sup_{j\ge1} W_j \quad , \quad U = \limsup_{j\to\infty} U_j \quad , \quad W = \limsup_{j\to\infty} W_j$$

If one of the following two conditions holds:

(2.18.2) V_j *and* $(U_1, \ldots, U_j)$ *are independent for all* $j \ge 1$

(2.18.3) V_j *and* $(U_j, U_{j+1} \ldots)$ *are independent for all* $j \ge 1$

then we have

$$P(M_n > x) \le \frac{P(L_n > x-y)}{\min\limits_{1\le j\le n} P(V_j \le y)} \qquad \forall x,y \in \mathbf{R} \tag{2.18.4}$$

$$P(M > x) \le \frac{P(L > x-y)}{\inf\limits_{j\ge 1} P(V_j \le y)} \qquad \forall x,y \in \mathbf{R} \tag{2.18.5}$$

$$P(U > x) \le \frac{P(W > x-y)}{\liminf\limits_{j\to\infty} P(V_j \le y)} \qquad \forall x,y \in \mathbf{R} \tag{2.18.6}$$

with the ***unusual*** *convention:* $\frac{0}{0} = 1$. *In particular, if* $a_1,\dots,a_n$ *are strictly positive numbers and* $S_n = X_1+\cdots+X_n$ *where* $X_1, X_2, \dots, X_n$ *are independent random variables, then for all* $x,y \in \mathbf{R}$, *we have*

$$P(M_n > x) \min_{1\le j\le n} P\left(|S_n - S_j| \le \frac{y}{a_j} \right) \le P\left(|S_n| > \frac{x-y}{m} \right) \tag{2.18.7}$$

where $M_n = \max\{a_1|S_1|,\dots,a_n|S_n|\}$ and $m = \max\{a_1,\dots,a_n\}$.

Remark: Loosely speaking, (2.18.7) tells us that the maximal partial sum of independent random variables is of the same order of magnitude as the last sum.

Proof: Suppose that (2.18.2) holds, set $A_j = \{U_j > x\}$, $B_j = \{V_j \le y\}$ and let $D_1, D_2, \dots$ be the disjungation of $A_1, A_2, \dots$:

$$D_1 = A_1, \qquad D_j = A_j \setminus (A_1 \cup \cdots \cup A_{j-1}) \qquad \forall j \ge 2$$

Then by (1.2.12) and (1.2.13), we have $\{M_n > x\} = D_1 \cup \cdots \cup D_n$ and $D_1, D_2, \dots$ are disjoint. Note that $A_j \cap B_j \subseteq \{W_j > x-y\}$ by (2.18.1) and that $D_j \in \sigma\{U_1,\dots,U_j\}$. Hence, D_j and B_j are independent by (2.18.2) and, thus, we have

$$\begin{aligned}
P(L_n > x-y) &= P\left(\bigcup_{j=1}^n \{W_j > x-y\} \right) \ge P\left(\bigcup_{j=1}^n (A_j \cap B_j) \right) \\
&\ge P\left(\bigcup_{j=1}^n (D_j \cap B_j) \right) = \sum_{j=1}^n P(D_j \cap B_j) \\
&= \sum_{j=1}^n P(D_j)P(B_j) \ge \min_{1\le j\le n} P(B_j) \sum_{i=1}^n P(D_i) \\
&= \min_{1\le j\le n} P(B_j)\, P\left(\bigcup_{i=1}^n D_i \right) = \min_{1\le j\le n} P(B_j)\, P(M_n > x)
\end{aligned}$$

and so (2.18.4) holds. Letting $n \to \infty$ in (2.18.4) we obtain (2.18.5), and by (2.18.5) we have

$$P(M^p > x) \inf_{j \geq p} P(V_j \leq y) \leq P(L^p > x - y)$$

where $M^p = \sup\{U_j \mid j \geq p\}$ and $L^p = \sup\{W_j \mid j \geq p\}$. Since $M^p \downarrow U$ and $L^p \downarrow W$, we see that (2.18.6) follows by letting p tend to infinity.

Thus, we have proved (2.18.4)-(2.18.6) under assumption (2.18.2). The proof of the case that (2.18.3) holds is exactly the same if we redefine the D_j 's as follows:

$$D_n = A_n\,, \qquad D_j = A_j \setminus (A_{j+1} \cup \cdots \cup A_n) \qquad \forall 1 \leq j \leq n-1$$

Moreover, setting $U_j = a_j|S_j|$, $V_j = a_j|S_n - S_j|$ and $W_j = m|S_n|$, we see that $U_j \leq V_j + W_j$ and since $\sigma(V_j) \subseteq \sigma(X_{j+1}, \ldots, X_n)$ and $\sigma(U_1, \ldots, U_j) \subseteq \sigma(X_1, \ldots, X_j)$ by independence of $X_1, \ldots, X_n$, we see that (2.18.2) holds by (2.10); thus, (2.18.7) follows from (2.18.4). □

2.19 A Factory Model

In a factory, two machines I and II produce screw bolts. From previous experience it is known that machine I has probability p_1 for producing a defective bolt and machine II has probability p_2 for producing a defective bolt, where $0 < p_2 < p_1 < 1$; that is, machine I produces more defect bolts than machine II. On a certain day, machine I has produced n_1 bolts and machine II has produced n_2 bolts. One bolt is picked at random from the total $n = n_1 + n_2$ bolts. An inspection shows that the chosen bolt is defective. What is the probability that the bolt has been produced by machine I?

Let A_1 and A_2 denote the events that the chosen bolt is produced by machines I and machine II, respectively, and let B denote the event that the bolt is defective. Then we have $P(A_j) = \frac{n_j}{n}$ and $P(B \mid A_j) = p_j$ for $j = 1, 2$. By Bayes' rule (2.2.8), we get

$$P(A_1 \mid B) = \frac{n_1 p_1}{n_1 p_1 + n_2 p_2}$$

and so

$$P(A_1 \mid B) - P(A_1) = \frac{n_1 n_2 (p_1 - p_2)}{n(n_1 p_1 + n_2 p_2)} > 0$$

Hence, knowing that the bolt is defective increases the probability that the bolt was produced by machine I.

2.20 Fake Infection

An insurance company has N insurance holders: $1, 2, \ldots, N$. An insurance holder is picked at random; that is, insurance holder k has a probability of $\frac{1}{N}$ of being chosen. The record of the chosen insurance holder is observed over a number of years (year $1, 2, \ldots, L$). Set

$$A_j = \{ \text{ the chosen insurance holder has an accident in year } j \}$$
$$B_k = \{ \text{ the chosen insurance holder has the number } k \}$$

Let q_k be the probability that insurance holder k has an accident in any given year, and let us suppose that the events that insurance holder k has an accident in years $1, 2, \cdots$ are independent; that is, we assume that

$$P(B_k) = \frac{1}{N} \quad , \quad P\left(\bigcap_{j=1}^{n} A_j \;\middle|\; B_k \right) = q_k^{\,m}$$

for $1 \le k \le N$ and $1 \le m \le L$. By the law of total probability (2.2.7) we find

$$P\left(\bigcap_{j=1}^{m} A_j \right) = \frac{1}{N} \sum_{k=1}^{n} q_k^{\,m} = P\left(\bigcap_{j=n+1}^{n+m} A_j \right)$$

$$P(A_1) = \frac{1}{N} \sum_{k=1}^{n} q_k \quad , \quad P(A_1 \cap A_2) = \frac{1}{N} \sum_{k=1}^{n} q_k^2$$

for all $n \ge 0$ and all $1 \le m \le L-n$. Setting $\alpha = P(A_1)$ and $\beta = P(A_1 \cap A_2)$ we find

$$P(A_2 \mid A_1) - P(A_2) = \frac{\beta}{\alpha} - \alpha = \frac{\beta - \alpha^2}{\alpha} = \frac{1}{\alpha N} \sum_{k=1}^{N} (q_k - \alpha)^2$$

Hence, $P(A_2 \mid A_1) \ge P(A_2)$, and if just two of the insurance holders have different accident probabilities, then $P(A_2 \mid A_1) > P(A_2)$; that is, knowing that the chosen insurance holder has an accident the first year, increases the probability that he will have an accident the second year. This seems to contradict the assumptions that the events that any given insurance holder has an accident in year $1, 2, \cdots$ are independent and have the same probability - but it doesn't. The reason being that the observed insurance holder is not a given one but a randomly chosen one. The phenomenon is usually called *fake infection* - the probability of an event being "infected" by knowing an ostensible independent event.

Let me illustrate this by a simpler example. Consider the two sequences of zeroes and ones:

(a) 000111101001101010111 01

(b) 00110000100100000000100

These two sequences are actually produced in such a way that the digits are chosen independently of each other, and the probability of "1" in series (a) is $\frac{1}{2}$ and in series (b) it is $\frac{1}{3}$. Let us toss a fair coin. If head comes up, we choose (a); if tail comes up we choose (b). In this way we get a series of zeroes and ones [either (a) or (b)]. The digits in this new series are no longer independent of each other; actually, if $A_j = \{$ the j-th digit is $1\}$, then we have

$$P(A_j) = \frac{5}{12} \quad , \quad P(A_1 \cap A_2) = \frac{13}{72} \neq \frac{5}{12} \cdot \frac{5}{12}$$

in spite of the fact that all the original digits were chosen completely independently of each other, and choosing one of the series (a) and (b), of course, does not change the digits.

This is only one of the many examples in probability theory showing that our intuition about probabilities, independence, conditioning, etc. is poor and often takes us on the wrong track. The moral of this tale is never to rely too much on your intuition when you evaluate probabilities, in particular, conditional probabilities, but rely on your computations.

2.21 Nuclear Safety

In the journal *Bulletin of the Atomic Scientist* (September 1975, pp. 25-28), Norman C. Rasmussen (main author of The Reactor Safety Study, WASH-1400) computes the probability of the worst possible accident on a nuclear power plant as follows:

$$\begin{aligned} P(A \cap B \cap C \cap D) &= P(A)P(B)P(C)P(D) \\ &= 10^{-3} \cdot 10^{-3} \cdot 10^{-1} \cdot 10^{-2} = 10^{-9} \end{aligned}$$

where A, B, C and D are the events

A: the accident initiates,

B: the safety systems fail,

C: the worst possible weather conditions,

D: the plant is situated near a large city.

I shall leave it to the reader to think about the following questions: Are A and B independent? What about B and C ? What has D to do with an accident on a nuclear power plant? Less than 15 miles from the center of Copenhagen there is a nuclear power plant (Barsebäck, Sweden); if the probability that this plant is near Copenhagen is 1 %, where is it the other 99 % of the time?

2.22 Waiting Times

Let X be the waiting time for a certain event to occur, for instance, the waiting time until the next airplane accident. Let us assume that the event has no memory and no ageing. This means that the probability for waiting more than t time units equals the conditional probability for waiting more than $(t+s)$ time units, given that we have already waited more than s time units. There are two possible models, depending on whether we measure time in discrete time units (such as days or months or years) or if we measure time continuously.

The discrete model: Suppose that time is measured in discrete time units. Then X assumes its values in $\mathbf{N}$. Let $F = F_X$ be the distribution function of X and $p = p_X$ be the probability mass function of X, and set

$$R(t) = P(X > t) = 1 - F(t)$$

Our assumption that the event is without memory can be expressed as

$$R(t) = P(X > t) = P(X > t + s \mid X > s) = \frac{R(t+s)}{R(s)}$$

for all $s, t \in \mathbf{N}$. Hence, $R(t+s) = R(s)R(t)$ for all $s, t \in \mathbf{N}$. Setting $q = R(1) = P(X > 1)$, we find $R(n) = q^n$, and so if $int(v)$ is the integer part of v, then we have

$$F(v) = F(int(v)) = 1 - q^{int(v)} \qquad \forall v \geq 0 \tag{2.22.1}$$

$$p_X(k) = q^{k-1}(1-q) \qquad \forall k \in \mathbf{N} \tag{2.22.2}$$

Recall that the *geometric distribution* $Ge(q)$ ($0 \leq q \leq 1$) is the distribution with probability mass function given by

$$p(n) = q^n (1-q) \qquad \forall n = 0, 1, 2, \ldots$$

Thus, we see that $X - 1 \sim Ge(q)$.

Let $X_1, \ldots, X_n$ be the waiting time until certain independent event occurs; that is, $X_1, \ldots, X_n$ are independent random variables such that $X_j - 1 \sim Ge(q_j)$ for $j = 1, \ldots, n$. Let X be the waiting time until one of the events occurs, that is, $X = X_1 \wedge \cdots \wedge X_n$ [recall that $a \wedge b = \min\{a, b\}$]. Then we have

$$P(X > t) = P(X_1 > t, \ldots, X_n > t) = \prod_{j=1}^{n} P(X_j > t) = \left(\prod_{j=1}^{n} q_j \right)^t$$

for all $t \in \mathbf{N}$. Hence, we have $X - 1 \sim Ge(q)$ where $q = q_1 \cdot q_2 \cdots q_n$.

According to *The Nuclear Safety Report - WASH 1400*, the probability of a meltdown on a given nuclear power plant equals $5 \cdot 10^{-5}$ per year. In 1987 there were at least 600 nuclear power plants in use. Hence if X is the waiting time until the next meltdown on a nuclear power plant, then $X - 1 \sim Ge(q)$ where $q = (1 - 5 \cdot 10^{-5})^{600} = 0.970$ and $F(k) = 1 - q^k$ is the probability of at least one melt-down in a nuclear plant within the next k years. In Table 2.1 you will find the probabilities of a nuclear accident within the next k years. For instance, the probability of at least one meltdown on some nuclear power plant within the next 25 years equals 53%, according to the official estimates. The mean of X equals $1/(1-q) = 33.8$ and the variance equals $q(1-q)^{-2} = 1110.97$. Hence, the expected waiting time until the next melt-down is 33.8 years with a standard deviation equal to 33.33.

Table 2.1

The probability $F(k)$ *for a nuclear accident within the next* k *years*									
k	1	2	5	10	15	20	25	30	50
$F(k)$	.03	.06	.14	.26	.36	.45	.53	.59	.78

The continuous model: If time is measured continuously and $R(t) = 1 - F(t) = P(X > t)$, then, as above, we find that R is decreasing and

$$R(0) = 1 \quad \text{and} \quad R(s+t) = R(s)\,R(t) \qquad \forall s,t \geq 0 \tag{2.22.3}$$

It is well known that the only decreasing solutions to (2.22.3) are of the form $R(t) = e^{-\lambda t}$ where λ is a positive constant. Hence, we have

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \tag{2.22.4}$$

Recall that the *exponential distribution* $E(\lambda)$ ($\lambda > 0$) is the distribution function with density

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Thus, $X \sim E(\lambda)$, and as before we find that if $X_1, \ldots, X_n$ are independent and $X_j \sim E(\lambda_j)$ for $j = 1, \ldots, n$. Then

$$X = X_1 \wedge \cdots \wedge X_n \sim E(\lambda_1 + \cdots + \lambda_n) \tag{2.22.5}$$

The mean in $E(\lambda)$ equals $\frac{1}{\lambda}$, and if $\varepsilon > 0$ is small, then

$$\frac{P(X \le \varepsilon)}{P(|X - \frac{1}{\lambda}| \le \frac{\varepsilon}{2})} = e^{1-\frac{1}{2}\varepsilon\lambda} \sim e = 2.718 \cdots$$

Hence, in an exponential distribution the probability of a "small" waiting time is approximately 2.7 times the probability of getting an "average" waiting time. This means independent random events (such as air plane accidents, nuclear melt-downs, deaths, etc.) have a tendency to come in clusters of a little less than 3. This offers a mathematical explanation of the proverb: "Accidents never come alone."

2.23 Order Statistics

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with a common distribution function F . We can then order $X_1, \ldots, X_n$ according to size: $X_{1n} \le \cdots \le X_{nn}$, that is,

$$\begin{aligned}
X_{1n} &= X_1 \wedge \cdots \wedge X_n = \text{ the smallest of the } X_j \text{ 's} \\
X_{2n} &= \text{ the second smallest of the } X_j \text{ 's} \\
X_{kn} &= \text{ the } k\text{-th smallest of the } X_j \text{ 's} \\
X_{nn} &= \text{ the largest of } X_j \text{ 's } = X_1 \vee \cdots \vee X_n
\end{aligned}$$

The random variables $X_{1n}, \ldots, X_{nn}$ are called *the order statistics* of $X_1, \ldots, X_n$ and we define *the empirical distribution function* $F_n(v, \omega)$ as

$$F_n(v,\omega) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{X_j \le v\}}(\omega) = \begin{cases} 0 & \text{if } v < X_{1n} \\ \frac{k}{n} & \text{if } X_{kn} \le v < X_{k+1,n} \\ 1 & \text{if } v \ge X_{nn} \end{cases} \tag{2.23.1}$$

Note that $F_n(v, \cdot)$ for fixed v is a random variable and that $F_n(\cdot, \omega)$ for fixed ω is a distribution function. We shall later see that $F_n(v)$ converges to the true distribution $F(v)$ as $n \to \infty$.

Let $v \in \mathbf{R}$ be given and set $A_j = \{X_j \le v\}$; then $A_1, \ldots, A_n$ are independent events with the same probability $P(A_j) = F(v)$, and $n\,F_n(v)$ is the number of occurrences among $A_1, \ldots, A_n$. Hence, we have

$$n\,F_n(v) \sim Bi(n, F(v)) \tag{2.23.2}$$

where $Bi(n,q)$ ($n \in \mathbf{N}$, $0 \le q \le 1$) is the *binomial distribution* with parameters (n,q) , that is, the distribution with probability mass function

$$p(j) = \binom{n}{j} q^j (1-q)^{n-j} \quad \text{for} \quad j = 0, 1, \ldots, n$$

Since $\{X_{kn} \le v\} = \{n F_n(v) \ge k\}$, we have that the distribution function of the kth order statistics is given by

$$(2.23.3) \qquad P(X_{kn} \le v) = \sum_{j=k}^{n} \binom{n}{j} F(v)^j (1 - F(v))^{n-j}$$

Suppose that the following condition holds:

(2.23.4) The limit $a = \lim_{x \downarrow 0} \frac{F(x)}{x}$ exists and $a > 0$

That is, $X_j > 0$ a.s., and the right-hand derivative of F at 0 exists and is positive. Let $k \ge 1$ be fixed and let $n \ge k$ and $v \ge 0$, then by (2.23.3) we find

$$\begin{aligned} P(n X_{kn} \le v) &= \sum_{j=k}^{n} \binom{n}{j} F\left(\tfrac{v}{n}\right)^j \left(1 - F\left(\tfrac{v}{n}\right)\right)^{n-j} \\ &= 1 - \sum_{j=0}^{k-1} \binom{n}{j} F\left(\tfrac{v}{n}\right)^j \left(1 - F\left(\tfrac{v}{n}\right)\right)^{n-j} \\ &= 1 - \sum_{j=0}^{k-1} \tfrac{1}{j!} \tfrac{n(n-1)\cdots(n-j+1)}{n^j} \left\{n F\left(\tfrac{v}{n}\right)\right\}^j \left(1 - F\left(\tfrac{v}{n}\right)\right)^{n-j} \end{aligned}$$

By (2.23.4) we have $n F(\frac{v}{n}) \to av$ as $n \to \infty$. Hence, using the well-known result from calculus:

$$(2.23.5) \qquad \lim_{n\to\infty} c_n = c \quad \Rightarrow \quad \lim_{n\to\infty} \left(1 - \frac{c_n}{n}\right)^{n-\alpha} = e^{-c} \quad \forall \alpha \in \mathbf{R}$$

we get

$$\lim_{n\to\infty} \left\{ n F\left(\tfrac{v}{n}\right) \right\}^j = (av)^j \quad , \quad \lim_{n\to\infty} \tfrac{n(n-1)\cdots(n-j+1)}{n^j} = 1$$

$$\lim_{n\to\infty} \left(1 - F\left(\tfrac{v}{n}\right)\right)^{n-j} = \lim_{n\to\infty} \left(1 - \frac{n F\left(\frac{v}{n}\right)}{n}\right)^{n-j} = e^{-av}$$

for all $j = 0, 1, \ldots$. Let $\Gamma(x|\alpha,\beta)$ be the distribution function of the *the gamma distribution* with parameters α and β [see (4.23)], that is,

$$\Gamma(x|\alpha,\beta) = \begin{cases} \dfrac{\beta^{-\alpha}}{\Gamma(\alpha)} \displaystyle\int_0^x t^{\alpha-1} e^{-t/\beta} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Then by performing a partial integration k-times, we get

(2.23.6) $$\lim_{n\to\infty} P(n X_{nk} \le v) = 1 - \sum_{j=0}^{k-1} \frac{(av)^j e^{-av}}{j!} = \Gamma\big(v \mid k, \tfrac{1}{a}\big) \quad \forall v \in \mathbf{R}$$

(2.23.7) $n X_{kn}$ is approximately $\Gamma(k, \frac{1}{a})$ -distributed when n is large and k is fixed

2.24 Uniform Distributions

Let I be an interval on $\mathbf{R}$ and let us try to construct a model for picking a random number in I. Let us require that the probabilities of picking the number in two subintervals of the same length are equal. Let X be the random number and let Q be the distribution law of X. Our requirement means that Q is a probability measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ such that there exists a function $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ satisfying

(2.24.1) $$Q(J) = \varphi(\lambda(J)) \qquad \forall J \text{ an interval } \subseteq I$$

(2.24.2) $$Q(I) = 1$$

where λ is the Lebesgue measure on $\mathbf{R}$. Since Q is increasing and finitely additive, we must have that φ is increasing on $[0, \lambda(I)]$ and

$$\varphi(x+y) = \varphi(x) + \varphi(y) \qquad \text{if} \qquad x, y \ge 0 \qquad \text{and} \qquad x + y \le \lambda(I)$$

However, the only increasing solutions to this equation is $\varphi(x) = ax$ where a is some positive number. Thus, we see that $Q(J) = a\,\lambda(J)$ for all intervals $J \subseteq I$. Hence, by the uniqueness theorem (1.7) we have that $Q(B) = a\lambda(I \cap B)$ for all $B \in \mathcal{B}(\mathbf{R})$, and since $Q(I) = 1$, we see that $\lambda(I)$ must be finite and

(2.24.3) $$P(X \in B) = \frac{\lambda(I \cap B)}{\lambda(I)} \qquad \forall B \in \mathcal{B}(\mathbf{R})$$

This means that if I is an unbounded interval, then there **does not** exist a model within the Kolmogorov system for random selection from I. But if I is a bounded interval, then by (2.8.6) there does exist a model for random selection from I.

If I is any Borel subset of $\mathbf{R}$ with $0 < \lambda(I) < \infty$ and X is a random variable satisfying (2.24.3), then we say that X is *uniformly distributed* over I. In particular, if X is uniformly distributed over the interval $I = [a, b]$ where $-\infty < a < b < \infty$, then we say that X has a *rectangular distribution* $R(a, b)$. More generally, if D is a Borel subset of $\mathbf{R}^k$ such that $0 < \lambda_k(D) < \infty$

where λ_k is the Lebesgue measure on $\mathbf{R}^k$, then we say that the random vector $X = (X_1, \ldots, X_k)$ is *uniformly distributed* over D if

$$(2.24.1) \qquad P(X \in B) = \frac{\lambda_k(B \cap D)}{\lambda_k(D)} \qquad \forall B \in \mathcal{B}(\mathbf{R}^k)$$

If $D = \{x \in \mathbf{R}^k \mid ||x|| < 1\}$ is *the unit ball* in $\mathbf{R}^k$, we speak about *the spherical distribution*, and if D is *the unit simplex*

$$D = \left\{ x \in \mathbf{R}^k \;\middle|\; x_1 > 0, \ldots, x_k > 0 \,,\; \sum_{j=1}^{k} x_j < 1 \right\}$$

then we speak about *the Dirichlet distribution* $D_k(1, 1, \ldots, 1)$ [see (8.16.2) in Volume II].

2.25 Random Intervals

Let $X_1, \ldots, X_n$ be independent random variables which are uniformly distributed over $[0, 1]$. Then $X_1, \ldots, X_n$ will divide the interval into $(n+1)$ subintervals. Let $L_0, L_1, \ldots, L_n$ denote the length of these intervals, that is,

$$L_j = X_{j+1,n} - X_{jn} \qquad \text{for } j = 0, 1, \ldots, n$$

where $X_{1n}, \ldots, X_{nn}$ are the order statistics of $X_1, \ldots, X_n$ and $X_{0n} := 0$ and $X_{n+1n} := 1$. Then I claim that $L_0, L_1, \ldots, L_n$ are identically distributed (but not independent – why not ?) and that L_j has a *beta distribution* $B(1, n)$, where the beta distribution $B(\alpha, \beta)$ ($\alpha, \beta > 0$) is the distribution with density given by

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 < x < 1 \\ 0 & \text{if } x \notin]0, 1[\end{cases}$$

and $\Gamma(\alpha)$ is the gamma function [see (3.37)]. This may seem a little strange since most people would expect that the two intervals at each end would behave differently from the others.

To see this let us roll the unit interval to a circle of length 1 and let us choose $(n+1)$ points $Y_0, Y_1, \ldots, Y_n$ independent and uniformly distributed over the circle. Then the circle is divided into $(n+1)$ arcs which, evidently, are identically distributed. Cut the circle at the point Y_0 and straighten out the circle to an interval of length 1. Then the points $Y_1, \ldots, Y_n$ clearly will be transformed into

n independent, uniformly distributed random variables $X_1, \ldots, X_n$, and the length of the $(n+1)$ arcs will be exactly $L_0, L_1, \ldots, L_n$. Thus, $L_0, \ldots, L_n$ all have the same distribution, and so

$$\begin{aligned} P(L_j \leq v) =& P(L_0 \leq v) = 1 - P(L_0 > v) \\ &= 1 - P(X_1 > v, \ldots, X_n > v) = 1 - (1-v)^n \end{aligned}$$

for all $0 \leq v \leq 1$. Hence, we have

(2.25.1) $\quad L_j \sim B(1, n) \qquad \forall j = 0, 1, \ldots, n$

Moreover, since the distribution function F of X_j equals $F(v) = v$ for $0 \leq v \leq 1$, we have by (2.23.3) and (2.23.7) that

(2.25.2) $\quad X_{kn} \sim B(k, nk+1)$

(2.25.3) $\quad nL_j$ is approximately $E(1)$-distributed as $n \to \infty$

(2.25.4) $\quad nX_{kn}$ is approximately $\Gamma(k, 1)$-distributed as $n \to \infty$

2.26 References to Chapter 2

In [3] and [5] you will find excellent introductions to elementary probability theory which don't use measure theory. For a long time Loeve's book [6] served as the standard textbook of advanced probability theory and it contains a wealth of information. If you want a more modern approach you should consult [1], [2] and [4]. In Feller's second volume (see [4]) you will find a rich sample of applications of probability to problems in other sciences, and in Dudley's book (see [2]) you will find a conjunction of probability and analysis.

[1] Chow, Y. S. and Teicher, H. (1978): *Probability Theory*, Springer Verlag, New York and Berlin

[2] Dudley, R. M. (1989): *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove

[3] Feller, W. (1957): *An Introduction to Probability and Its Applications*, Vol. I, Second ed., John Wiley & Sons, Inc., New York and London

[4] Feller, W. (1971): *An Introduction to Probability and Its Applications*, Vol. II second ed., John Wiley & Sons, Inc., New York and London

[5] Hoel, P. G., Port, S. C. and Stone, C. J. (1971): *Introduction to Probability*, Houghton Mifflin Co., Boston and London

[6] Loeve, M. (1963): *Probability Theory*, Third ed., D. Van Nostrand Co., Inc., New York and London

2.27 Exercises to Chapter 2

Exc. 2.1: Suppose that the probability p_n that a family has exactly n children is given by

$$p_n = \begin{cases} \alpha p^n & \text{if } n \geq 1 \\ \frac{1-(\alpha+1)p}{1-p} & \text{if } n = 0 \end{cases}$$

where α, p are two parameters satisfying $0 < p < 1$ and $0 \leq \alpha \leq 1 - p$. Let us also suppose that given that a family has exactly n children, then the events $G_j = \{$ the jth child is a girl $\}$ $(j = 1, \ldots, n)$ are independent and have the same probability, say $0 < q < 1$. What the probability that a randomly chosen family has exactly k girls, $k = 1, 2, \ldots$? What is the conditional probability that a randomly chosen family has two or more girls, given that it has at least one girl?

Remark: The parameter p is interpreted as the probability that a family gets a child, given that the parents are fertile, that is, capable of having children. The parameter α is a fertility parameter and $r = \frac{\alpha}{1-p}$ is interpreted as the probability that the parents are fertile. Observations from all over the world shows that $0.48 \leq q \leq 0.49$. In the year 1937 an American family satisfied the model above with $\alpha = 0.7358$.

Exc. 2.2: A cereal company puts coupons numbered $1, 2, \ldots, R$ in every cereal packet. When a costumer has collected a complete set of coupons, she or he wins a holiday in Hawaii. All the numbers $1, 2, \ldots, R$ are equally probable. Let $p(N, R)$ denote the probability that a customer with N coupons has a complete set of coupons. Show that

(i) $$p(N, R) = \sum_{j=0}^{R-1} \binom{R}{j} (-1)^j \left(1 - \frac{j}{R}\right)^N \qquad \forall N \geq R$$

Exc. 2.3: Ann has bought k lots in a lottery with a total of N lots. The lottery has m prices, and the winning lots is a random subset of $\{1, 2, \ldots N\}$ of size m such that all numbers have the same chance of being drawn. Let $p(k, m, N)$ denote the probability that Ann wins at least one of the m prices. Show that

$$\text{(i)} \qquad p(k, m, N) = 1 - \frac{\binom{N-k}{m}}{\binom{N}{m}} = 1 - \frac{\binom{N-m}{k}}{\binom{N}{k}}$$

Exc. 2.4: Let a, b, c be chosen uniformly in the interval $[-1, 1]$. What is the probability that the quadratic equation $ax^2 + bx + c = 0$ has a real root?

Exc. 2.5: A point Q is chosen uniformly in an equilateral triangle. Let X denote the smallest distance from Q to one of the three sides of the triangle. Find the distribution function of X.

Exc. 2.6: A point Q is chosen uniformly in a regular polygon with six sides. Let X denote the distance from Q to the center of the polygon. Find the distribution function of X.

Exc. 2.7: *(Banach's problem)* Stefan Banach (1892–1945) was an eminent Polish mathematician and a heavy smoker. He posed the following problem: He is carrying two boxes of matches each originally carrying 100 matches. Every time he lights a cigarette he chooses a box at random. What is the probability that when he chooses an empty box for the first time, the other box contains exactly r matches for $r = 0, 1, 2, \ldots, 100$?

Exc. 2.8: Mary draws cards from a card deck with 52 cards. After each draw she puts the card back and shuffle the deck. Let X denote the number of times until a card has been drawn twice. Show that

$$\text{(i)} \qquad P(X = k) = \frac{(k-1) \cdot (k-1)!}{52^k} \binom{52}{k-1} \qquad \forall 2 \leq k \leq 52$$

Exc. 2.9: When biologists count the population size of a species of wild animals, they often use the following method: A certain number, say M, of the animals are caught and marked; after the capture the animals are set loose. At some later time, another sample of size say n is caught. Let m be the number of marked animals in the new sample and let N be the unknown total population size. Show that $\widehat{N} = \frac{nM}{m}$ is the most likely value of N and give a probabilistic or statistical argument supporting your reasoning.

Exc. 2.10: Michael claims that he can taste the difference between a Tuborg beer and a Carlsberg beer. His good friend Molly designs the following test: Ten pair of glasses are presented for Michael, and he is told that each pair contains a glass with Tuborg beer and a glass with Carlsberg beer. He tastes each pair and tells which is which. How many correct answers out of the 10 possible should he have in order that you believe his claim? Give a probabilistic or statistical argument supporting your reasoning.

Exc. 2.11: In a paper "Testing the claim of a graphologist" (Journ. Pers., vol. 16, 1947, p. 192–197) G. R. Pascal and B. Suttel describes the following experiment: A graphologist claimed that she could tell from the hand writing of a person if that person was psychotic or not. Ten people, who were diagnosed to be psychotic by at least two independent psychiatrists, were matched with a healthy person of the same sex, age and educational and social background. The graphologist received a sample of hand writing from each of the 10 pair and was asked to decide for each of the ten pairs who was normal and who was psychotic. The graphologist had six correct answers and four wrong answers. Do you believe the claim of the graphologist? Give a probabilistic or statistical argument supporting your reasoning.

Exc. 2.12: When my nephew Hans-Henrik was 9 years old he showed me a gambling machine. A circular disk was mounted on a box, the disk was divided into nine arcs (of unequal lengths) numbered $1, 2, \ldots, 9$, and on the box there was a mark. The player choose a number between 1 and 9, say k , and the disk is rotated if the mark points at the chosen number k , the player receives k-times the stake; otherwise the player loses the stake (the stake was usually a colored marking ring). How shall Hans-Henrik divide the wheel if he wants to maximize his average profit under the assumption of an intelligent player? Find the average profit with the optimal division.

Exc. 2.13: In a group of k persons, each one throws a die. The player with the highest number of points wins the game. If two or more players receives the maximal number of points, the game undecided. Find the probability of the following three events:

A : the game is decided

B : player 1 wins the game

C_r : the player who wins has r points ($r = 1, 2, \ldots, 6$)

Exc. 2.14: A *formal power series* is an expression of the form

$$p(x) = \sum_{n=0}^{\infty} a_n x^n$$

where $a_n \in \mathbf{R}$ for all $n = 0, 1, 2, \ldots$ and x^n is a formal symbol, which only acts as a memo-technical device to remember the addition and multiplication rules below. If

$$p(x) = \sum_{n=0}^{\infty} a_n x^n \quad , \quad q(x) = \sum_{n=0}^{\infty} b_n x^n$$

are two formal power series. Then we define their sum $p(x) + q(x)$ and product $p(x) \cdot q(x)$ as follows:

$$p(x) + q(x) = \sum_{n=0}^{\infty} (a_n + b_n) x^n \quad , \quad p(x) \cdot q(x) = \sum_{n=0}^{\infty} \left\{ \sum_{j=0}^{n} a_j b_{n-j} \right\} x^n$$

Let $p_j(x) = \sum_0^\infty a_{jn} x^n$ be a formal power series for $j = 1, 2, \ldots k$. Show that

(i) $$\prod_{j=1}^{k} p_j(x) = \sum_{n=0}^{\infty} a_n x^n \quad \text{where} \quad a_n = \sum_{j_1 + \cdots + j_k = n} a_{1j_1} \cdot a_{2j_2} \cdots a_{kj_k}$$

Let there be given k subsets $A_1, \ldots, A_k$ of the non-negative integers $\mathbf{N}_0$, and let there be given k boxes numbered $1, 2, \ldots, k$. Let c_{nk} denote the number of ways that n balls can be placed in the k boxes obeying the rule that the number of balls in the jth box belongs to the set A_j for all $j = 1, \ldots, k$. Set

$$p_j(x) = \sum_{n=0}^{\infty} 1_{A_j}(n) x^n \qquad \forall j = 1, \ldots, k$$

and show that [you may either apply (i) or an induction in k]

(ii) $$\sum_{n=0}^{\infty} c_{nk} x^n = \prod_{j=1}^{k} p_j(x)$$

Exc. 2.15: In how many ways can you put n balls in k boxes in such a way that each box contains at most one ball? [Apply Exc.2.14 with $A_j = \{0, 1\}$.] In how many ways can you put n balls in k boxes in such a way that each box contains at least 1 ball?

Exc. 2.16: Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\{A_n\} \subseteq \mathcal{F}$. Show that

(i) $$P\left(\bigcap_{n=1}^{\infty} A_n\right) \geq 1 - \sum_{n=1}^{\infty} (1 - P(A_n))$$

Exc. 2.17: The three events A, B, C are pairwise independent. Moreover, $A^c \cup B^c$ and C^c are independent. Does it follow that A, B, C are independent?

Exc. 2.18: Show that an event A is independent of itself if and only if either $P(A) = 0$ or $P(A) = 1$, and if so, then A and F are independent for all $F \in \mathcal{F}$.

Exc. 2.19: Let $\{A_n\}$ be a sequence of events on the probability space $(\Omega, \mathcal{F}, P)$. Show that

(i) $$P(A_n) = 0 \quad \forall n \geq 1 \quad \Rightarrow \quad P\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$$

(ii) $$P(A_n) = 1 \quad \forall n \geq 1 \quad \Rightarrow \quad P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$$

Exc. 2.20: Let $A_1, \ldots, A_n$ be independent events and let $p_k = P(A_k)$ for $k = 1, \ldots, n$. Show that [use the inequality: $1 - x \leq e^{-x}$]

(i) $$P\left(\bigcup_{k=1}^{n} A_k\right) \geq 1 - \exp\left(-\sum_{k=1}^{n} p_k\right)$$

Exc. 2.21: Let $\{A_n\}$ be a sequence of independent events such that

$$\sum_{n=1}^{\infty} (1 - P(A_n)) = \infty$$

Show that $P(\liminf A_n) = 0$.

Exc. 2.22: Let $\{A_n\}$ be a sequence of independent events, and set $A_* = \liminf A_n$ and $A^* = \limsup A_n$. From the 0–1 law we know that $P(A_*)$ and $P(A^*)$ are either 0 or 1 . Apply the two Borel-Cantelli lemmas to give precise conditions in terms of $p_n = P(A_n)$ for $P(A_*) = 0$, $P(A_*) = 1$, $P(A^*) = 0$ and $P(A^*) = 1$.

Exc. 2.23: Let $\{A_n\}$ be a sequence of independent events such that $P(A_n) < 1$ for all $n \geq 1$ and $P(\cup_1^\infty A_n) = 1$. Show that $P(\limsup A_n) = 1$.

Hint: Let $p_n = 1 - P(A_n)$, and show that $\prod_1^\infty p_n = 0$. Since $p_n \neq 0$ for all $n \geq 1$, this implies

$$P\left(\bigcup_{n=j}^{\infty} A_n\right) = 1 - \prod_{n=j}^{\infty} p_n = 1 \qquad \forall j \geq 1$$

Finally, apply Exc.2.19.

Exc. 2.24: Let $\{\mathcal{F}_n\}$ be a sequence of independent σ-algebras on the probability space $(\Omega, \mathcal{F}, P)$. Set

$$\mathcal{G} = \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{j=n}^{\infty} \mathcal{F}_j\right)$$

Show that $P(G) = 0$ or 1 for every $G \in \mathcal{G}$.

Exc. 2.25: *(The inclusion-exclusion principle)* Let $A_1, \ldots, A_n$ be events on the probability space $(\Omega, \mathcal{F}, P)$. Set $L(\omega) = \#\{1 \leq j \leq n \,|\, \omega \in A_j\}$ and

$$s_0 = 1 \quad , \quad s_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$$

for all $k = 1, \ldots, k$. Show that

(i) $\quad P(L = k) = \sum_{j=k}^{n} (-1)^{j-k} \binom{j}{k} s_j \qquad \forall 0 \leq k \leq n$

(ii) $\quad P\left(\bigcup_{j=1}^{n} A_n\right) = P(L \geq 1) = \sum_{j=1}^{n} (-1)^{j-1} s_j$

[Take mean values on both sides in formulas (ii) and (iii) of Exc.1.8.] Suppose that $A_1, \ldots, A_n$ are independent and have the same probability $p = P(A_j)$ for all $j = 1, \ldots, n$. Show that

(iii) $\quad s_k = \binom{n}{k} p^k \qquad \forall k = 0, 1, \ldots, n$

(iv) $\quad P(L = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \forall k = 0, 1, \ldots, n$

Thus, (i) may be considered as a generalization of the binomial distribution. Formula (ii) is called the *inclusion-exclusion principle*, and it gives a formula for the probability of a non-disjoint union of events.

Exc. 2.26: *(Exchangeable events)* Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $p_n \in [0.1]$, and let $\{A_n\}$ be sequence of events satisfying

$$P\left(\bigcap_{j=1}^{k} A_{n_j}\right) = p_k \qquad \forall k \geq 1 \, \forall 1 \leq n_1 < \cdots < n_k$$

Show that

(i) $$P(\limsup A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} p_j$$

(ii) $$P(\liminf A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} p_n$$

Hint: Use the inclusion-exclusion principle [see Exc.2.25] to show that

$$P\left(\bigcup_{j=k}^{k+n-1} A_j\right) = \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} p_j$$

for all $k, n \geq 1$. Apply this to show (i).

Exc. 2.27: *(Euler's formula)* Let $\zeta(\alpha)$ denote *Riemann's zeta function*, that is,

$$\zeta(\alpha) = \sum_{n=1}^{\infty} n^{-\alpha} \qquad \forall 1 < \alpha < \infty$$

[Note that the series is finite if and only if $\alpha > 1$.] Let $\alpha > 1$ be given, and define the probability measure P on $(\mathbf{N}, 2^{\mathbf{N}})$ by

$$P(A) = \zeta(\alpha)^{-1} \sum_{n \in A} n^{-\alpha} \qquad \forall A \subseteq \mathbf{N}$$

Let $A(k) = \{n \in \mathbf{N} \mid k \text{ divides } n\} = \{k, 2k, \ldots, rk, \ldots\}$ and let $k_1, \ldots, k_m$ be given positive integers. Show that

(i) $P(A(k)) = k^{-\alpha} \qquad \forall k \geq 1$

(ii) $\bigcap_{j=1}^{m} A(k_j) = A(k_1 \cdot k_2 \cdots k_m)$ if $\gcd(k_i, k_j) = 1 \quad \forall i \neq j$

(iii) $A(k_1), \ldots, A(k_m)$ are independent if $\gcd(k_i, k_j) = 1 \;\; \forall i \neq j$

where $\gcd(p,q)$ denotes *the greatest common divisor* of the numbers p and q, that is, the largest integer which divides both p and q. Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots, p_{9487} = 98849, \ldots$ be the sequence of prime numbers. Show that

(iv) $\displaystyle\bigcap_{n=1}^{\infty} A(p_n)^c = \{1\}$

(v) $\displaystyle\zeta(\alpha)^{-1} = \prod_{n=1}^{\infty}\left(1 - \frac{1}{p_n^{\alpha}}\right)$

Formula (v) is called *Euler's formula.*

Hint: (v): Apply (iii) to show that $A(p_1), A(p_2), \ldots$ are independent. Conclude that $A(p_1)^c, A(p_2)^c, \ldots$ are independent, and apply (iv) and (i).

Exc. 2.28: Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots, p_{9487} = 98849, \ldots$ be the sequence of prime numbers. Set $\pi(x) = \#(\{j \in \mathbf{N} \mid p_j \le x\})$ for all $x \in \mathbf{N}$. Show that

(i) $\displaystyle\pi(x) \ge \frac{\log x}{2\log 2} \quad \forall x \in \mathbf{N}$ and $\displaystyle\sum_{n=1}^{\infty} \frac{1}{p_n} = \infty$

The estimate in (i) is very poor. Actually, it can be shown that $\pi(x)$ is approximately equal to $\frac{x}{\log x}$, but this is very difficult to prove.

Hint; Let $A_j(x)$ be the set of integers $1 \le y \le x$ whose prime factors all belong to the set $\{p_1, \ldots, p_j\}$ and let $\nu_j(x) = \#A_j(x)$. Let $y \in A_j(x)$, and show that y can be written on the form $y = z^2 p_1^{\varepsilon(1)} \cdots p_j^{\varepsilon(j)}$, where $1 \le z \le \sqrt{y}$ is an integer and $\varepsilon(i) = 0$ or 1 for all $1 \le i \le j$. Argue that this gives the following estimate of $\nu_j(x)$:

$$\nu_j(x) \le 2^j \sqrt{x} \quad \forall x \ge 1\, \forall j \ge 1$$

Show that the first part of (i) follows from this inequality by setting $j = \pi(x)$. Note that if $y \in \{1, \ldots, x\} \setminus A_j(x)$, then there exist $k > j$ and an integer $1 \le m \le x/p_k$ such that $y = m p_k$. Argue that this shows that following inequalities:

$$x - 2^j\sqrt{x} \le x - \nu_j(x) \le x \sum_{n=j+1}^{\infty} \frac{1}{p_n} \quad \forall x \ge 1\, \forall j \ge 1$$

Finally, show that these inequalities imply the second part of (i).

Exc. 2.29: *(Buffon's needle problem)* In the plane we draw parallel lines with distance $r > 0$. A needle of length $0 < t \leq r$ is thrown at random on the plane. Show that the probability that the needle intersects one of the lines equals $\frac{2t}{r\pi}$, where $\pi = 3.1415926535\ldots$ as usual denotes the length of a circle of diameter 1 .

This problem goes under the name *Buffon's needle problem.* The result has been used to find π experimentally. With $r = t = 5$ and 11 throws, I found 7 intersections giving the value $\pi \sim 3.142$, which is correct to 3 decimals places. Around 1850, a more persistent mathematician made the experiment 10,000 times with $r = t$ and found $\pi \sim 3.139$, which is only correct to 2 decimal places. A third and more clever mathematician made the experiment 355 times with $t = 5$ and $r = 10$ and found 113 intersections giving $\pi \sim 3.1415929$, which correct up 6 decimal places. There are very good reasons to choose $r = t$ and 11 throws, and to choose $r = 2t$ and 355 throws. Can you think of any reasons for these choices? [The answer to this question is tricky, and some knowledge of continued fractions is helpful.]

Exc. 2.30: This exercise offers you a cheap way of prolonging the lifetime of an electric bulb. Fix a certain time in the future, say tomorrow 11 A.M. At the fixed time, light the bulb and observe if it works. If the bulb works, then I claim that its average lifetime has been increased – why? Let us make a probabilistic model. Suppose that we are using a clock which only runs when the bulb is lit, and suppose that whenever the bulb breaks down it is replaced by a new bulb. Let $T_0 = 0 < T_1 < T_2 < \ldots$ be the random times when a new bulb is installed. Then the lifetime of the nth bulb equals $L_n = T_n - T_{n-1}$ $(n = 1, 2, \ldots)$. Suppose that the bulbs have no memory. Then by (2.22) it is reasonable to assume that $L_1, L_2, \ldots$ are independent and have an exponential distribution $E(\lambda)$ for some $\lambda > 0$. Then the average lifetime equals $E\,L_n = \frac{1}{\lambda}$. Let $a > 0$ be the fixed observation time, and let τ be the unique (random!) integer satisfying $T_{\tau-1} \leq a < T_\tau$. Then τ is the number of the bulb in function at the observation time, and L_τ is the lifetime for the observed bulb. Show that L_τ has density function $f(x)$ and mean value given by

$$\text{(i)} \qquad f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda^2\, x\, e^{-\lambda x} & \text{if } 0 \leq x < a \\ \lambda\,(1 + a\lambda)\, e^{-\lambda x} & \text{if } a \leq x \end{cases}$$

$$\text{(ii)} \qquad E\,L_\tau = \frac{2 - e^{-\lambda a}}{\lambda} > \frac{1}{\lambda} = E\,L_n \qquad \forall n = 1, 2, \ldots$$

Try to explain this paradox.

Hint: Note that $\{\tau = n\} = \{T_{n-1} \le a < T_n\}$ and show that

$$P(L_\tau \le x) = P(a \le L_1 \le x) + \sum_{n=2}^{\infty} P(T_{n-1} \le a < T_n \ , \ L_n \le x)$$

Let $n \ge 2$ be given. Show that T_{n-1} and L_n are independent and that (T_{n-1}, L_n) has density $f_n(u,v)$ given by

$$f_n(u,v) = \begin{cases} \dfrac{\lambda^n u^{n-2} e^{-\lambda(u+v)}}{(n-2)!} & \text{if } u > 0 \text{ and } v > 0 \\ 0 & \text{otherwise} \end{cases}$$

for all $n \ge 2$. Deduce that

$$\sum_{n=2}^{\infty} f_n(u,v) = \begin{cases} \lambda^2 e^{-\lambda v} & \text{if } u > 0 \text{ and } v > 0 \\ 0 & \text{otherwise} \end{cases}$$

and use the equality

$$\begin{aligned} P(T_{n-1} \le a < T_n \ , \ L_n \le x) &= P(T_{n-1} \le a < T_{n-1} + L_n \ , \ L_n \le x) \\ &= \int_{a-x}^{a} du \int_{a-u}^{x} f_n(u,v)\, dv \end{aligned}$$

to compute the distribution function of L_τ .

Exc. 2.31: Buses are arriving to a bus stop according to a given schedule with 10 minute intervals. But due to traffic the buses does not arrive exactly on schedule. Make a sensible stochastic model which can answer the following question: What is your average waiting time for next bus if you arrive to the bus stop without knowing the schedule? [There are many possible answers to this problem, but you will find that for all sensible models the mean waiting time is minimized if the busses arrive exactly on schedule. Can you find a good reason for this?]

Exc. 2.32: A secretary sets N letters in N envelopes with addresses. However, the letters have been mixed up, so they are put into the envelopes at random. Let $p(N,k)$ denote the probability that exactly k of the letters are placed in the correct envelope. Show that

(i) $$p(N,k) = \frac{1}{k!}\left(1 - \sum_{j=0}^{N-k} \frac{(-1)^j}{j!}\right) \qquad \forall k = 0,1,2,\ldots,N$$

Exc. 2.33: Let $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, and let P be the Lebesgue measure on $[0,1]$. Then $(\Omega, \mathcal{F}, P)$ is a probability space. Let

$$A_n = \bigcup_{j=1}^{2^n} \left[\frac{2j-1}{2^{n+1}}, \frac{2j}{2^{n+1}} \right] \quad \forall n = 0,1,2,\ldots$$

Show that $A_0, A_1, \ldots$ are independent and that $P(A_n) = \frac{1}{2}$ for all $n \geq 0$.

Exc. 2.34: Let $F : \mathbf{R} \to \mathbf{R}$ be an increasing right continuous function and let λ_F be the associated Lebesgue-Stieltjes measure [see (1.32)]. Let $F(x-) = \lim_{u \uparrow x} F(u)$ for all $-\infty < x \leq \infty$ and $F(\pm\infty) = \lim_{x \to \pm\infty} F(x)$. Show that

(i) $\lambda_F(]a,b]) = F(b) - F(a) \quad \forall -\infty \leq a \leq b < \infty$

(ii) $\lambda_F(\{a\}) = F(a) - F(a-) \quad \forall a \in \mathbf{R}$

(iii) $\lambda_F([a,b[) = F(b-) - F(a-) \quad \forall -\infty < a \leq b \leq \infty$

(iv) $\lambda_F(]a,b[) = F(b-) - F(a) \quad \forall -\infty \leq a \leq b \leq \infty$

(v) $\lambda_F([a,b]) = F(b) - F(a-) \quad \forall -\infty < a \leq b < \infty$

(vi) F is continuous at a if and only if $a \notin sp_+(\lambda_F)$ [see Exc.1.60]

(vii) F is continuous on $\mathbf{R}$ if and only if λ_F is diffuse [see Exc.1.60]

Exc. 2.35: Let $X = (X_1, \ldots, X_n)$ be an n-dimensional, random vector with distribution function $F(x_1, \ldots, x_n)$, and let $F_j(x)$ be the marginal distribution function of X_j for $j = 1, \ldots, n$. Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ and $a = (a_1, \ldots, a_n)$ be given vectors in $\mathbf{R}^n$, and let $\varepsilon > 0$. Show the following:

(i) $|F(x_1, \ldots, x_n) - F(y_1, \ldots, y_n)| \leq \sum_{j=1}^{n} |F_j(x_j) - F_j(y_j)|$

(ii)
$$P(X_j = a_j) \leq \sum_{i=1}^{n} (1 - F_i(x_i)) + F(x_1, \ldots, a_j, \ldots, x_n) - F(x_1, \ldots, a_j - \varepsilon, \ldots, x_n)$$

(iii) If F_j is continuous at a_j for all $j = 1, \ldots, n$, then F is continuous at a

(iv) X is simple (respectively discrete) if and only if $X_1, \ldots, X_n$ all are simple (respectively discrete)

(v) If one of the random variables $X_1, \ldots, X_n$ are diffuse, then so is X

(vi) F is continuous on $\mathbf{R}^n$ if and only if $X_1, \ldots, X_n$ all are diffuse

Hint: (i): Let $A = A_1 \cap \cdots \cap A_n$ and $B = B_1 \cap \cdots \cap B_n$ where $A_j = \{X_j \leq x_j\}$ and $B_j = \{X_j \leq y_j\}$. Then $F(x_1, \ldots, x_n) = P(A)$ and $F(y_1, \ldots, y_n) = P(B)$. Apply (1.3.1) and (v) in Exc.1.3 to show that

$$|P(A) - P(B)| \leq P(A \triangle B) \leq \sum_{j=1}^{n} P(A_j \triangle B_j)$$

$$P(A_j \triangle B_j) = |F_j(x_j) - F(y_j)| \quad \forall j = 1, \ldots, n$$

(ii): Show that

$$\{X_j = a_j\} \subseteq \{X_1 \leq x_1 , \ldots , a_j - \varepsilon < X_j \leq a_j , \ldots , X_n \leq x_n\} \cup \bigcup_{i=1}^{n} \{X_i > x_i\}$$

and argue that (ii) follows from this inclusion.

Exc. 2.36: Let $X = (X_1, \ldots, X_n)$ be an n-dimensional random vector with distribution function $F(x_1, \ldots, x_n)$ and let $M = \max\{X_1, \ldots, X_n\}$. Show that $P(M \leq s) = F(s, \ldots, s)$ for all $s \in \mathbf{R}$.

Exc. 2.37: Let $X_1, \ldots, X_n$ be independent, real, random variables with distribution functions $F_1, \ldots, F_n$. Let $M = \max\{X_1, \ldots, X_n\}$ and $L = \min\{X_1, \ldots, X_n\}$, and let $G(s)$, $H(t)$ and $\Phi(s,t)$ be the distribution functions of M , L and (M, L) , respectively. Show that

(i) $$G(s) = \prod_{j=1}^{n} F_j(s) \qquad \forall s \in \mathbf{R}$$

(ii) $$H(t) = 1 - \prod_{j=1}^{n} (1 - F_j(t)) \qquad \forall t \in \mathbf{R}$$

(iii) $$P(M \leq s \ , \ L > t) = \begin{cases} \prod_{j=1}^{n} (F_j(s) - F_j(t)) & \text{if } s > t \\ 0 & \text{if } s \leq t \end{cases}$$

(iv) $$\Phi(s,t) = \begin{cases} \prod_{j=1}^{n} F_j(s) - \prod_{j=1}^{n} (F_j(s) - F_j(t)) & \text{if } s > t \\ \prod_{j=1}^{n} F_j(s) & \text{if } s \leq t \end{cases}$$

Exc. 2.38: Let $X = (X_1, \ldots, X_n)$ be an n-dimensional random vector with distribution function $F(x_1, \ldots, x_n)$. Let $M = \max\{X_1, \ldots, X_n\}$ and $L = \min\{X_1, \ldots, X_n\}$, and let $\Phi(s,t)$ be the distribution functions of (M, L) . Let $\Gamma = \{0,1\}^n$ and let $s, t \in \mathbf{R}$ be given. If $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \in \Gamma$, then we set

$$|\varepsilon| = \varepsilon_1 + \cdots + \varepsilon_n \quad , \quad u_j(\varepsilon) = \varepsilon_j\, t + (1 - \varepsilon_j)\, s$$

for all $j = 1, \ldots, n$. Show that

$$\text{(i)} \qquad \Phi(s,t) = \begin{cases} F(s, \ldots, s) - \sum\limits_{\varepsilon \in \Gamma} (-1)^{|\varepsilon|} F(u_1(\varepsilon), \ldots, u_n(\varepsilon)) & \text{if } s > t \\ F(s, \ldots, s) & \text{if } s \le t \end{cases}$$

Hint: Suppose that $s > t$. Then show that

$$\Phi(s,t) = P(X_1 \le s\,, \ldots,\, X_n \le s) - P(t < X_1 \le s\,, \ldots,\, t < X_n \le s)$$

and apply (2.8.11).

Exc. 2.39: A random variable X is called *symmetric* if X and $-X$ have the same distribution. Let X be random variable with distribution function $F(t) = P(X \le t)$ and let $F(t-) = \lim\limits_{s \uparrow t} F(s)$ be the left limit of F at t . Show that

(i) X is symmetric if and only if $F(-t) = 1 - F(t-)$ for all $t \in \mathbf{R}$

(ii) If X and Y are independent and have the same distribution, then $X - Y$ is symmetric

Hint: (ii): Note that (X, Y) and (Y, X) both have distribution function $F(s)F(t)$. Then apply (2.8.15) with $f(s,t) = s - t$.

Exc. 2.40: Let $A_1, \ldots, A_n$ be events on the probability space $(\Omega, \mathcal{F}, P)$. Show that

$$\text{(i)} \qquad P(A_1 \cap \cdots \cap A_n) \le \prod_{j=1}^{n} P(A_j)^{1/n}$$

[Note that $P(A_1 \cap \cdots \cap A_n) \le P(A_j)$ for all $j = 1, \ldots, n$.] Let $F(x_1, \ldots, x_n)$ be an n-dimensional distribution function and let $F_1(x), \ldots, F_n(x)$ be the 1-dimensional marginals distribution functions of F . Show that

$$\text{(ii)} \qquad F(x_1, \ldots, x_n) \le \prod_{j=1}^{n} F_j(x_j)^{1/n} \qquad \forall\, (x_1, \ldots, x_n) \in \mathbf{R}^n$$

Exc. 2.41: Let $F : \mathbf{R}^2 \to [0,1]$ be the function

$$F(x,y) = \begin{cases} 1 & \text{if } x+y \geq 0 \\ 0 & \text{if } x+y < 0 \end{cases}$$

Show that $F(\cdot, y)$ and $F(x, \cdot)$ are 1-dimensional distribution functions for all $x, y \in \mathbf{R}$. Show that F is **not** a 2–dimensional distribution function [consider the increment over $]a,b]$, where $a = (-1,-1)$ and $b = (1,1)$.]

Exc. 2.42: Let F and G be 1–dimensional distribution functions. Show that for every $\alpha \in [-1,1]$, we have that

$$H_\alpha(x,y) = F(x)\,G(y)\,\{1 + \alpha\,(1 - F(x))\,(1 - G(y))\}$$

is a 2-dimensional distribution function with F and G as its two 1-dimensional marginals. Suppose that $F = G = N(0,1)$ where $N(0,1)$, as usual, denotes the normal distribution with mean 0 and variance 1 . Show that the distribution function H_α as defined above is a 2-dimensional normal distribution if and only if $\alpha = 0$. Thus, a 2-dimensional random vector (X,Y) may have normal marginal distributions without having a 2-dimensional, normal distribution.

Exc. 2.43: Let (X,Y) be a 2-dimensional random vector with distribution function $F(s,t) = P(X \leq s\ ,\ Y \leq t)$, and let F_1 and F_2 be the distribution functions of X and Y respectively. Let us set $H(s,t) = F_1(s) \wedge F_2(t)$. Show that

(i) H is a 2-dimensional distribution function with 1-dimensional marginals F_1 and F_2 such that $F(s,t) \leq H(s,t)$ for all $s,t \in \mathbf{R}$

Thus, H is the largest 2-dimensional distribution function with 1-dimensional marginals F_1 and F_2 . Suppose that F_1 and F_2 are strictly increasing and continuous, and let Φ_j be the inverse function of F_j for $j = 1,2$. Set $U = \Phi_1(F_2(Y))$, and show that (U,Y) has 2-dimensional distribution function H .

Exc. 2.44: Let X be a random variable and let $Y = g(X)$ where $g : \mathbf{R} \to \mathbf{R}$ is a Borel function. Show that X and Y are independent if and only if Y is degenerated. [Note that $\sigma(Y) \subseteq \sigma(X)$, and apply the 0–1 law (2.13) with $X_j = Y$ for all $j \geq 1$.]

Exc. 2.45: Let X be a random variable with distribution function F . Let $Y = \frac{1}{X}$ with the convention: $\frac{1}{0} = 0$. Find the distribution function of Y .

Exc. 2.46: Let $X_1, \ldots, X_n$ be independent, identically distributed random variables. If A is a Borel subset of $\mathbf{R}$, then we define

$$N_n(A, \omega) = \#\{1 \le j \le n \mid X_j(\omega) \in A\} = \sum_{j=1}^{n} 1_A(X_j(\omega))$$

Show that $N_n(A)$ has a binomial distribution $Bi(n,p)$ where $p = P(X_1 \in A)$. Let $A_1, \ldots, A_k$ be disjoint Borel subsets of $\mathbf{R}$. Show that the random vector $(N(A_1), \ldots, N(A_k))$ has a multinomial distribution $Mu_k(n, p_1, \ldots, p_k)$, where $p_j = P(X_1 \in A_j)$ for $j = 1, \ldots, k$.

Recall that the *k-dimensional multinomial distribution* $Mu_k(n, p_1, \ldots, p_k)$ where $p_1 \ge 0, \ldots p_k \ge 0$ and $p_1 + \cdots + p_k \le 1$, is the discrete k-dimensional distribution having probability mass function

$$p(x_1, \ldots, x_k) = \begin{cases} \dfrac{n!\, p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k} (1 - p_\bullet)^{n - x_\bullet}}{x_1! \cdot x_2! \cdots x_k! (n - x_\bullet)!} & \text{if } (x_1, \ldots, x_k) \in \Delta_n^k \\ 0 & \text{if } (x_1, \ldots, x_k) \notin \Delta_n^k \end{cases}$$

where $y_\bullet := y_1 + \cdots + y_k$ for all $y = (y_1, \ldots, y_k) \in \mathbf{R}^k$ and Δ_n^k is the set of all $(x_1, \ldots, x_k) \in \mathbf{N}_0^k$ with $x_\bullet \le n$.

Exc. 2.47: Let X and Y be random variables such that

$$P(X = 1\,,\, Y = 1) = P(X = 1\,,\, Y = 2) = \tfrac{1}{4} \quad , \quad P(X = 2\,,\, Y = 2) = \tfrac{1}{2}$$

Find the simultaneous distribution $F(s,t)$ of (X,Y) and the marginal distributions $F_1(s)$ and $F_2(t)$ of X and Y. At which points is F continuous? [The most convenient is to give the answer in the form of a drawing.]

Exc. 2.48: *(Positive types)* Let $Dis(\mathbf{R})$ denote the set of all 1-dimensional, distribution functions, and let Δ_a denote the distribution function associated with the Dirac measure δ_a [see (1.15)]. If $F \in Dis(\mathbf{R})$. Then we say that F is *non-degenerated* if $F \ne \Delta_a$ for all $a \in \mathbf{R}$ [cf. (2.14)]. Show that

(i) $\Delta_a(x) = \begin{cases} 1 & \text{if } x \ge a \\ 0 & \text{if } x < a \end{cases}$

(ii) F is non-degenerated if and only if there exists $x_0 \in \mathbf{R}$ such that $0 < F(x_0) < 1$

Let $F \in Dis(\mathbf{R})$ be a given distribution function, and let $\Phi_F(a,b)$ denote the distribution function: $F(\frac{x-a}{b})$ whenever $(a,b) \in \mathbf{R}\times]0,\infty[$. Then Φ_F is a function from $\mathbf{R}\times]0,\infty[$ into $Dis(\mathbf{R})$. We say that F and G are of *the same positive type*, and write $F \bowtie G$ if $G = \Phi_F(a,b)$ for some $(a,b) \in \mathbf{R}\times]0,\infty[$. Show the following:

(iii) $\bowtie$ is an equivalence relation on $Dis(\mathbf{R})$

(iv) The set of all degenerated distributions: $\{\Delta_a \mid a \in \mathbf{R}\}$ is a single equivalence class with respect to the equivalence relation $\bowtie$

(v) The set of all normal distributions: $\{N(\mu,\sigma^2) \mid (\mu,\sigma) \in \mathbf{R}\times]0,\infty[\}$ is a single equivalence class with respect to the equivalence relation $\bowtie$

(vi) If F is non-degenerated, then Φ_F is an injective function from $\mathbf{R}\times]0,\infty[$ into $Dis(\mathbf{R})$

Hint: (vi): Suppose that $\Phi_F(a,b) = \Phi_F(\alpha,\beta)$, and let $\gamma = \frac{b}{\beta}$ and $\delta = \frac{a-\alpha}{\beta}$. Show that $F(y) = F(\gamma y + \delta)$ for all $y \in \mathbf{R}$. Iterating this n-times gives

$$F(y) = \begin{cases} F(\gamma^n (y-\rho)+\rho) & \text{if } \gamma \neq 1 \\ F(y + n\,\delta) & \text{if } \gamma = 1 \end{cases}$$

for all $y \in \mathbf{R}$, where $\rho = \frac{\delta}{1-\gamma}$ whenever $\gamma \neq 1$. Let $n \to \infty$ and apply (ii).

Exc. 2.49: Let $Dis(\mathbf{R})$ be the set of all 1-dimensional distribution functions. Show that $Dis(\mathbf{R})$ is *convex*, that is, show that

(i) $\alpha F + (1-\alpha) G \in Dis(\mathbf{R}) \quad \forall 0 \le \alpha \le 1 \; \forall F, G \in Dis(\mathbf{R})$

Moreover, show that $Dis(\mathbf{R})$ is *σ-convex*, that is, show that

(ii) $\sum\limits_{n=1}^{\infty} \alpha_n F_n \in Dis(\mathbf{R})$ if $\alpha_n \ge 0$, $\sum\limits_{n=1}^{\infty} \alpha_n = 1$ and $F_1, F_2, \ldots \in Dis(\mathbf{R})$

Exc. 2.50: Let the random vector (X,Y) have a 2-dimensional normal distribution $N_2(0,\sigma)$, where σ is the 2×2-matrix

$$\sigma = \begin{Bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{Bmatrix}$$

Show that

(i) $P(XY < 0) = \dfrac{1}{\pi} \arccos\left(\dfrac{\sigma_{12}}{\sqrt{\sigma_{11}\,\sigma_{22}}}\right)$

Exc. 2.51: Let $\Phi(x)$ denote *the standard normal distribution function* $N(0,1)$ and let $\varphi(x)$ denote the corresponding density function; that is,

$$\varphi(x) = \Phi'(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

Set $\alpha := \sqrt{\frac{2}{\pi}}$, and show that for all $x > 0$, we have

(i) $$\left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x) \le 1 - \Phi(x) \le \frac{1}{x}\varphi(x)$$

(ii) $$\frac{\varphi(x)}{\alpha + x} \le 1 - \Phi(x) \le \frac{2\,\varphi(x)}{2\,\alpha + x}$$

Hint: To prove the four inequalities, subtract and differentiate with respect x using the formulas $\varphi'(x) = -x\,\varphi(x)$ and $\Phi'(x) = \varphi(x)$.

Exc. 2.52: Let $X_0, X_1, X_2 \ldots$ be a sequence of independent, identically distributed random variables, and let A be a Borel subset of $\mathbf{R}$ such that $0 < P(X_0 \in A) < 1$. Let $\tau(\omega) = \min\{n \in \mathbf{N}_0 \mid X_n(\omega) \in A\}$. Show that τ has a geometric distribution $Ge(p)$, where $p = P(X_0 \notin A)$.

Recall that *the geometric distribution* $Ge(p)$, where $0 < p < 1$, is given by the probability mass function

$$q(x) = \begin{cases} (1-p)p^x & \forall x \in \mathbf{N}_0 \\ 0 & \forall x \notin \mathbf{N}_0 \end{cases}$$

Exc. 2.53: *(The problem of points)* Two men A and B are playing "balla" (probably a medieval Italian ball game). Both have set 10 gold coins at stake, and the one who first wins N games gets the 20 gold coins. However, at a certain time, the play interrupted. At the time of interruption A has won a games, and B has won b games ($0 \le a, b < N$). It is supposed that the two men are equally skillful. How should they divide the stake? This problem with $N = 6$, $a = 5$, $b = 3$ is the first recorded probabilistic problem in the history. It was posed by Luca Pacioli (1445–1509) in his book *Summa de arithmetica, geometria, propotioni et proportionalita* published in 1494. Luca gave the answer $5 : 3$; in 1556, Niccolo Tartaglia (1499–1557) objected to the answer of Luca and claimed, that the stake should be divided in the proportions $2 : 1$. Two years later, G. F. Peverone claimed the right answer to be $6 : 1$; and in 1654, Blaise Pascal gave the answer $7 : 1$. Which of these answers is the correct one?

Solution: Consider the general problem of points, and let p denote the probability that A would have won the play if they had continued. Then

$$\text{(i)} \qquad p = \sum_{j=0}^{k-1} \binom{j+m-1}{m-1} 2^{-j-m}$$

where $m = N - a$ and $k = N - b$.

Exc. 2.54: *(Chevalier de Méré 's problem)* On July 29, 1654, Blaise Pascal (1623–1662) wrote to his friend Pierre de Fermat (1601–1665) :

Je n'ai pas le temps de envoyer la démonstration d'une difficulté qui étonnoit fort M. de Méré, car il a très bon esprit, mais il n'est pas géomètre (c'est, comme vous savez, un grand défaut) . . .

or in English:

I have no time to send you the proof of a difficulty which greatly puzzled M. de Méré, for he is very able, but he is not a mathematician (this, you know, is a great defect) . . . .

The problem that puzzled M. de Méré goes as follows: The probability of getting at least one "6" in 4 throws with a die equals $\frac{671}{1296} > \frac{1}{2}$, and since 4 is to 6 as 24 is to 36, this should (according to Chevalier de Méré) also be the probability of getting a "double 6" in 24 throws with 2 dice. However, by experience Chevalier de Méré found that the latter probability was less than $\frac{1}{2}$. This upset de Méré so much that he claimed the arithmetic to be wrong. But unlike you, de Méré was not a mathematician. Show that de Méré was wrong – and not the arithmetic, and find the correct probability for getting at least on "double 6" in 24 throws with 2 dice.

Solution: The probability equals $1-(\frac{35}{36})^{24} = 0.49140\ldots$. The good de Méré must have played the game a huge number times in order to note a difference of even less than 1%.

Exc. 2.55: Let $\{X_n\}$ be a sequence of random variables, and set $L^* = \limsup X_n$ and $L_* = \liminf X_n$. Show that

$$\text{(i)} \qquad P(L_* > y) \le \liminf_{n\to\infty} P(X_n > y) \le \limsup_{n\to\infty} P(X_n > y) \le P(L^* \ge y)$$

for all $y \in \mathbf{R}$.

Hint: Let $F_n = \{X_n > y\}$, and show that $\{L_* > y\} \subseteq \liminf F_n$ and $\{L^* \ge y\} \supseteq \limsup F_n$. Then apply (1.4.4) and (1.4.5).

Exc. 2.56: Let $\{U_j\}$ and $\{L_j\}$ be two sequences of random variables and let $\{V_{jn} \mid 1 \le j \le n\}$ be a triangular array of random variables satisfying

(*) V_{jn} and $(U_1, \dots U_j)$ are independent for all $1 \le j \le n < \infty$

(**) $U_j \le V_{jn} + L_n \qquad \forall 1 \le j \le n$

Let us define the following random variables and functions:

$$M_n = \max_{1 \le j \le n} U_j \ , \ M = \sup_{j \ge 1} U_j \ , \ L^* = \limsup_{n \to \infty} L_n \ , \ L_* = \liminf_{n \to \infty} L_n$$

$$R_n(y) = \max_{1 \le j \le n} P(V_{jn} > y) \ , \ R_*(y) = \liminf_{n \to \infty} R_n(y) \ , \ R^*(y) = \limsup_{n \to \infty} R_n(y)$$

$$G_n(y) = P(L_n > y) \ , \ G_*(y) = \liminf_{n \to \infty} G_n(y) \ , \ G^*(y) = \limsup_{n \to \infty} G_n(y)$$

Show the following:

(i) $P(L_* > y) \le G_*(y) \le G^*(y) \le P(L^* \ge y) \qquad \forall y \in \mathbf{R}$

(ii) $P(M_n > x) \le P(L_n > x - y) + R_n(y)\, P(M_n > x)$

(iii) $P(M > x) \le G^*(x - y) + R_*(y)\, P(M > x)$

(iv) $P(M > x) \le G_*(x - y) + R^*(y)\, P(M > x)$

for all $x, y \in \mathbf{R}$ and all $n \ge 1$.

Hint: (i): Follows from Exc.2.55. (ii): Follows easily from (2.18.4). (iii)+(iv): Follows from (ii) and (P.4.19).

Exc. 2.57: Let $\{U_j\}$ and $\{L_j\}$ be two sequences of random variables and let $\{V_{jn} \mid 1 \le j \le\}$ be a triangular array of random variables satisfying

(*) V_{jn} and $(U_1, \dots U_j)$ are independent for all $1 \le j \le n < \infty$

(**) $U_j \le U_{i-1} + V_{in} + V_{jn} + L_n \qquad \forall 1 \le i \le j \le n$

where we set $U_0 = 0$. Let $M_n, M, R_n, R_*, R^*, G_n, G_*$ and G^* be defined as in Exc.2.56, and let $0 \le y \le \frac{x}{2}$ and $n \ge 1$ be given. Show that

(i) $P(M_n > 2x) \le G_n(x - 2y) + 2\, R_n(y)\, P(M_n > x)$

(ii) $P(M > 2x) \le G^*(x - 2y) + 2\, R_*(y)\, P(M > x)$

(iii) $P(M > 2x) \le G_*(x - 2y) + 2\, R^*(y)\, P(M > x)$

Hint: (i): Set

$$D_j = \{U_j > x\ ,\ U_k \le x \quad \forall 0 \le k < j\} \quad , \quad Z_n = \sum_{i=1}^{n} 1_{D_i} V_{in}^+$$

Show that $D_1, D_2, \ldots$ are disjoint sets with $D_1 \cup \cdots \cup D_n = \{M_n > x\}$. Then show that $U_j \le x + Z_n + L_n^+ + V_{jn}^+$ and that D_j and V_{jn} are independent for all $1 \le j \le n$. Conclude that

$$P(Z_n > y) = \sum_{i=1}^{n} P(D_i) \cdot P(V_{in} > y) \le R_n(y)\, P(M_n > x)$$

Hence, by Exc.2.56.(ii) we have

$$\begin{aligned} P(M_n > 2x) &\le P\big(Z_n + L_n^+ > x - y\big) + R_n(y)\, P(M > 2x) \\ &\le P(Z_n > y) + P(L_n > x - 2y) + R_n(y)\, P(M > 2x) \end{aligned}$$

Show that this implies (i).

Exc. 2.58: Let $\{X_n\}$ be a sequence of independent random variables and set $S_k = X_1 + \cdots + X_k$ and

$$M_n = \max_{1 \le j \le n} |S_j| \ , \quad M = \sup_{j \ge 1} |S_j| \ , \quad L_n = \max_{1 \le j \le n} |X_j| \ , \quad L = \sup_{j \ge 1} |X_j|$$

$$R_n(y) = \max_{1 \le j \le n} P(|S_j| > y) \ , \quad R(y) = \sup_{j \ge 1} P(|S_j| > y) \ , \quad p_\infty = \lim_{y \to \infty} R(y)$$

for all $y \in \mathbf{R}$ and all $n \le 1$. Let $x \ge 0$, $y \in \mathbf{R}$ and $n \in \mathbf{N}$ be given. Show that

(i) $\quad R_n(y) \le P(M_n > y) \quad$ and $\quad R(y) \le P(M > y)$

(ii) $\quad P(M_n > x) \le P(|S_n| > x - 2y) + 2\, R_n(y)\, P(M_n > x)$

(iii) $\quad P(M > x) \le \liminf_{n \to \infty} P(|S_n| > x - 2y) + 2\, R(y)\, P(M > x)$

(iv) $\quad P(M_n > 2x) \le P(L_n > x - 4y) + 4\, R_n(y)\, P(M_n > x)$

(v) $\quad P(M > 2x) \le P(L > x - 4y) + 4\, R(y)\, P(M > x)$

(vi) $\quad P(M = \infty) = 0 \ \text{ or } \ 1$

(vii) $\quad P(M = \infty) = 0 \quad \Leftrightarrow \quad p_\infty = 0 \quad \Leftrightarrow \quad p_\infty < \frac{1}{3}$

Hint: Note that $P(|S_n - S_j| > 2y) \le 2\, R_n(y)$ for all $y \in \mathbf{R}$ and all $1 \le j \le n$. Show that (ii) and (iii) follow from (2.18.7) with $a_n \equiv 1$. Setting $U_j = |S_j|$ and $V_{jn} = |S_n - S_j|$, show that (iv) and (v) follows from Exc.2.57. Finally, show that (vi) and (vii) follow from (iii) and the 0-1 law (2.13).

Chapter 3

INTEGRATION

3.1 Integrals of Positive Functions

The integral was first introduced by Isaac Newton (1643-1727) and Gottfried Wilhelm Leibniz (1646-1716) for functions defined on an interval. In its first 150 years of existence, the definition of the integral was loose and imprecise (measured with the standard of today's mathematics), and the precise definition that you (probably) have learned in your elementary calculus course is due to Bernard Riemann (1826-1866). It was extended to arbitrary measures in the beginning of this century by mathematicians such as E. Borel, H. Lebesgue, C. Carathéodory and many others.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f be a function from Ω into $\mathbf{R}$. Then, loosely speaking, the μ-integral is a weighted average of f where we use μ as the weight, that is, something like the following:

$$\int f \, d\mu \approx \sum f(\omega) \, \mu(d\omega)$$

where $d\omega$ is an "infinitesimal small neighborhood" of ω. Actually, the integral sign $\int$ is a degenerated S and stands for "sum." However, the sum above has no precise mathematical sense, and we must, therefore, search for another and more precise definition of the integral. There are many ways (at least 10) to do this, and here I have chosen the fastest and most convenient way for our purposes.

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f be a measurable function from Ω into $\overline{\mathbf{R}}_+ = [0, \infty]$. Then we define *the μ-integral* of f as

$$s_n(f) := 2^{-n} \sum_{j=1}^{\infty} \mu(f > j2^{-n}) \qquad \forall n = 0, 1, 2, \ldots \tag{3.1.1}$$

$$\int f \, d\mu := \int_\Omega f \, d\mu = \int_\Omega f(\omega) \, \mu(d\omega) := \lim_{n \to \infty} s_n(f) \tag{3.1.2}$$

Note that we allow three different notations for the μ-integral of f, viz. $\int f \, d\mu$ and $\int_\Omega f \, d\mu$ and $\int_\Omega f(\omega) \, \mu(d\omega)$. We shall in a moment see that the limit of $s_n(f)$ as $n \to \infty$ always exists (but possibly equal to $+\infty$), and so the integral is well defined by (3.1.2) for every non-negative measurable function. In Section (3.2) we shall see how to extend the integral to arbitrary measurable functions taking values in the extended real line $[-\infty, \infty]$, and in (3.8) we shall see that (3.1.2) really is a formalization of the weighted sum above.

Having defined the integral, we have to show that the integral, as defined above, has the properties that an integral should have, such as linearity, monotonicity, good limit results, etc. This work is lengthy but of an elementary and routine character. It will now be done in eight steps.

The eight steps: *Let $f, g, f_1, f_2, \cdots$ be measurable functions from Ω into the extended positive line $\overline{\mathbf{R}}_+ = [0, \infty]$. Then we have the following:*

(3.1.3) $\quad s_n(f) \le s_{n+1}(f)$ *and so the limit* $\lim_{n \to \infty} s_n(f)$ *exists in* $\overline{\mathbf{R}}_+$

(3.1.4) $\quad f \le g$ *μ-a.s.* $\Rightarrow s_n(f) \le s_n(g)$ *and* $\int f \, d\mu \le \int g \, d\mu$

(3.1.5) $\quad$ *If* $s_0(f) < \infty$, *then* $f < \infty$ *μ-a.s.*

(3.1.6) $\quad s_n(f) + s_n(g) \le s_n(f+g) \le s_n(f) + s_n(g) + 2^{-n} \mu(f+g > 2^{-n})$

(3.1.7) $\quad$ *If* $\int f \, d\mu < \infty$, *then* $\lim_{n \to \infty} 2^{-n} \mu(f > 2^{-n}) = 0$

(3.1.8) $\quad \int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu$

(3.1.9) $\quad$ *If* $f_n \uparrow f$, *then* $\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu$

(3.1.10) $\quad \int (af) \, d\mu = a \int f \, d\mu \qquad$ *if* $a \in \overline{\mathbf{R}}_+$

Proof: (3.1.3): Let $A_j^n = \{f > j2^{-n}\}$, then $A_{2j}^{n+1} = A_j^n \subseteq A_{2j-1}^{n+1}$,

and so we find

$$s_n(f) = 2^{-n} \sum_{j=1}^{\infty} \mu(A_j^n) \le 2^{-n} \sum_{j=1}^{\infty} \tfrac{1}{2}(\mu(A_{2j}^{n+1}) + \mu(A_{2j-1}^{n+1}))$$
$$= 2^{-n-1} \sum_{j=1}^{\infty} \mu(A_j^{n+1}) = s_{n+1}(f)$$

(3.1.4): If $f \le g$ μ-a.s., then we have $\{f > j2^{-n}\} \subseteq \{g > j\,2^{-n}\}$ μ-a.s., and so (3.1.4) follows from (1.19.4).

(3.1.5): Since $\sum \mu(f > j) < \infty$, we have that $\mu(f > j) \to 0$ as $j \to \infty$, and since $\{f > j\} \downarrow \{f = \infty\}$, then by (1.4.3) we have that $\mu(f = \infty) = 0$, and so $f < \infty$ μ-a.s.

(3.1.6): If $s_n(f) = \infty$, then $s_n(f+g) = \infty$ because $f+g \ge f$; thus, all three expressions in (3.1.6) become $+\infty$. Hence, we may assume that $s_n(f) < \infty$, then by (3.1.5) we have that $f < \infty$ μ-a.s., and by (3.1.4) we see that we may assume that $f(\omega) < \infty$ for all $\omega \in \Omega$. Let $j, n \ge 0$ be integers and let us define

$$F_j^n = \{f > j2^{-n}\} \quad , \quad G_j^n = \{g > j2^{-n}\} \quad , \quad H_j^n = \{f + g > j2^{-n}\}$$
$$D_0^n = \{f \le 2^{-n}\} \quad , \quad D_j^n = \{j2^{-n} < f \le (j+1)2^{-n}\} \quad \text{if } j \ge 1$$

Then $D_0^n, \dots, D_{j-1}^n, F_j^n$ are disjoint and since $f(\omega) < \infty$ for all ω, we get

$$\Omega = \bigcup_{k=0}^{\infty} D_k^n = F_j^n \cup \bigcup_{k=0}^{j-1} D_k^n \qquad \forall j = 1, 2, \cdots$$

Note that $G_{j-k}^n \cap D_k^n \subseteq H_j^n \cap D_k^n \subseteq G_{j-k-1}^n \cap D_k^n$ and $F_j^n \subseteq H_j^n$; hence, we have

$$(*) \qquad \mu(H_j^n) = \mu(F_j^n) + \sum_{k=0}^{j-1} \mu(D_k^n \cap H_j^n) \ge \mu(F_j^n) + \sum_{i=1}^{j} \mu(D_{j-i}^n \cap G_i^n)$$

$$(**) \qquad \mu(H_j^n) = \mu(F_{j-1}^n) + \sum_{k=0}^{j-2} \mu(D_k^n \cap H_j^n) \le \mu(F_{j-1}^n) + \sum_{i=1}^{j-1} \mu(D_{j-i-1}^n \cap G_i^n)$$

Summing (*) over $j = 1, 2, \cdots$ gives

$$s_n(f+g) \ge s_n(f) + 2^{-n} \sum_{j=1}^{\infty} \sum_{i=1}^{j} \mu(D_{j-i}^n \cap G_i^n)$$
$$= s_n(f) + 2^{-n} \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \mu(D_{j-i}^n \cap G_i^n)$$
$$= s_n(f) + s_n(g)$$

Summing (**) over $j = 2, 3, \cdots$ gives

$$\begin{aligned} s_n(f+g) &\le 2^{-n}\,\mu(H_1^n) + s_n(f) + 2^{-n}\sum_{j=2}^{\infty}\sum_{i=1}^{j-1}\mu(D_{j-i-1}^n \cap G_i^n) \\ &= 2^{-n}\,\mu(H_1^n) + s_n(f) + 2^{-n}\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty}\mu(D_{j-i-1}^n \cap G_i^n) \\ &= 2^{-n}\,\mu(H_1^n) + s_n(f) + s_n(g) \end{aligned}$$

and so (3.1.6) is proved.

(3.1.7): Set $a_n = 2^{-n}\,\mu(f > 2^{-n})$ and $a = \int f\,d\mu$. Then $a_n \le s_n(f) \le a$, and if $F_j^n = \{f > j2^{-n}\}$, then $F_{2j-1}^{n+1} \supseteq F_{2j}^{n+1} = F_j^n$, and so we get

$$\begin{aligned} \tfrac{1}{2}\,a_n + s_{n+1}(f) &= 2^{-n-1}\,\mu(F_1^n) + a_{n+1} + 2^{-n-1}\sum_{j=2}^{\infty}\mu(F_j^{n+1}) \\ &= a_{n+1} + 2^{-n-1}\Big(\mu(F_1^n) + \sum_{j=2}^{\infty}\mu(F_{2j-1}^{n+1}) + \sum_{j=1}^{\infty}\mu(F_{2j}^{n+1})\Big) \\ &\ge a_{n+1} + 2^{-n-1}\Big(\sum_{j=1}^{\infty}\mu(F_j^n) + \sum_{j=1}^{\infty}\mu(F_j^n)\Big) \\ &= a_{n+1} + s_n(f) \end{aligned}$$

Let $c = \limsup_{n\to\infty} a_n$, ; then using (P.4.19) we find that $\frac{1}{2}c + a \ge c + a$, and since $0 \le c \le a < \infty$, we conclude that $c = 0$, and so (3.1.7) is proved.

(3.1.8): Letting $n \to \infty$ in (3.1.6), we see that

$$\int f\,d\mu + \int g\,d\mu \le \int (f+g)\,d\mu$$

Hence, if $\int f\,d\mu = \infty$ or $\int g\,d\mu = \infty$, then (3.1.8) holds. So suppose that $\int f\,d\mu < \infty$ and $\int g\,d\mu < \infty$. Since

$$2^{-n}\mu(f+g > 2^{-n}) \le 2^{-n}\mu(f > 2^{-n-1}) + 2^{-n}\,\mu(g > 2^{-n-1})$$

then letting $n \to \infty$ in (3.1.6) and using (3.1.7), we get

$$\int (f+g)\,d\mu \le \int f\,d\mu + \int g\,d\mu$$

Thus, (3.1.8) is proved.

(3.1.9): Let $a < \int f\,d\mu$; then by definition (3.1.2) there exists $p \geq 1$ such that $s_p(f) > a$. Hence, by definition (3.1.1) there exists $q \geq 1$ such that

$$2^{-p} \sum_{j=1}^{q} \mu(f > j \cdot 2^{-p}) > a$$

and since $\{f_n > j \cdot 2^{-p}\} \uparrow \{f > j \cdot 2^{-p}\}$ when $n \uparrow \infty$ and j and p are fixed, then by (1.4.2) there exists $k \geq 1$ such that

$$\begin{aligned} a < 2^{-p} \sum_{j=1}^{q} \mu(f_k > j2^{-p}) &\leq s_p(f_k) \leq \int f_k\,d\mu \\ &\leq \lim_{n\to\infty} \int f_n\,d\mu \leq \int f\,d\mu \end{aligned}$$

Letting $a \uparrow \int f\,d\mu$, we see that (3.1.9) holds.

(3.1.10): Using (3.1.8) and a simple induction argument, we see that (3.1.10) holds if $a \in \mathbf{N}_0$. Hence if $k \in \mathbf{N}$ and $m \in \mathbf{N}_0$, we have

$$m \cdot \int f\,d\mu = \int (m \cdot f)\,d\mu = \int k\left(\tfrac{m}{k}\right) \cdot f\,d\mu = k \cdot \int \left(\tfrac{m}{k}\right) \cdot f\,d\mu$$

and so (3.1.10) holds if a is a non-negative rational number. Let $a \in \overline{\mathbf{R}}_+$ be arbitrary; then there exists a sequence $\{q_n\}$ of non-negative rational numbers such that $q_n \uparrow a$. Then $q_n f \uparrow af$ and so by (3.1.9) we find

$$\int (af)\,d\mu = \lim_{n\to\infty} \int (q_n f)\,d\mu = \lim_{n\to\infty} q_n \int f\,d\mu = a \int f\,d\mu$$

and so (3.1.10) holds. □

3.2 Integrals of Real Functions

In (3.1) we have defined the integral of a non-negative function f . Suppose that $f : \Omega \to \overline{\mathbf{R}}$ is a measurable function with values in the extended real line $\overline{\mathbf{R}} = [-\infty, \infty]$. To define the integral of f , we split f into its *positive part* f^+ and its *negative part* f^- [see (P.4)]. Then by (P.4.1), we have

(3.2.1) $\quad f = f^+ - f^- \quad , \quad |f| = f^+ + f^-$

and f^+, f^- and $|f|$ are measurable functions from Ω into $\overline{\mathbf{R}}_+$. Hence if $f : \Omega \to \overline{\mathbf{R}}$ is measurable, we may define *the μ-integral of* f as

$$\int f\,d\mu = \int_\Omega f\,d\mu = \int_\Omega f(\omega)\,\mu(d\omega) := \int f^+\,d\mu - \int f^-\,d\mu \tag{3.2.2}$$

provided that the last expression is defined; that is, not of the form: $\infty - \infty$. Hence, the μ-integral of f is defined by (3.2.2) for all measurable functions f satisfying

$$\text{either} \quad \int_\Omega f^+\,d\mu < \infty \quad \text{or} \quad \int_\Omega f^-\,d\mu < \infty \tag{3.2.3}$$

and if so, we say that *the μ-integral of f exists* or that $f \in L(\mu)$, that is, $L(\mu)$ is the set of all functions $f : \Omega \to \overline{\mathbf{R}}$ for which the μ-integral exists:

$$L(\mu) := \left\{ f \in \overline{\mathbf{M}}(\Omega, \mathcal{F}) \;\middle|\; \text{the } \mu\text{-integral of } f \text{ exists} \right\} \tag{3.2.4}$$

where $\overline{\mathbf{M}}(\Omega, \mathcal{F})$ denotes the of all measurable functions from Ω into $\overline{\mathbf{R}}$.

Note that the integral $\int f\,d\mu$ may take the values $+\infty$ or $-\infty$. If the μ-integral of f exists and is finite, that is, not $+\infty$ nor $-\infty$, we say that f is *μ-integrable*. The set of all μ-integrable functions f is denoted $L^1(\mu)$; that is,

$$L^1(\mu) := \left\{ f \in L(\mu) \;\middle|\; \int_\Omega f\,d\mu \neq \pm\infty \right\} \tag{3.2.5}$$

By the definition of the arithmetic on $\overline{\mathbf{R}}$ [see (P.4)] we have

$$L^1(\mu) = \left\{ f \in \overline{M}(\Omega, \mathcal{F}) \;\middle|\; \int_\Omega f^+\,d\mu < \infty \text{ and } \int_\Omega f^-\,d\mu < \infty \right\} \tag{3.2.6}$$

and since $f^+ + f^- = |f|$, by (3.1.8) we have

$$L^1(\mu) = \left\{ f \in \overline{\mathbf{M}}(\Omega, \mathcal{F}) \;\middle|\; \int_\Omega |f|\,d\mu < \infty \right\} \tag{3.2.7}$$

If $F \in \mathcal{F}$ and $f : \Omega \to \overline{\mathbf{R}}$ is measurable, then we define *the μ-integral of f over F* as

$$\int_F f\,d\mu = \int_F f(\omega)\,\mu(d\omega) := \int_\Omega 1_F(\omega) f(\omega)\,\mu(d\omega) \tag{3.2.8}$$

provided, of course, that $1_F f \in L(\mu)$.

3.3 The Calculus of Integrals

Let $(\Omega, \mathcal{F}, \mu)$ *be a measure space, and suppose that* $f, g \in L(\mu)$ *and that* $h : \Omega \to \overline{\mathbf{R}}$ *is measurable. Then we have the following:*

(3.3.1) *If* $\int f\,d\mu > -\infty$ *and* $f \le h$ μ*-a.s., then* $h \in L(\mu)$

(3.3.2) *If* $\int f\,d\mu < \infty$ *and* $f \ge h$ μ*-a.s., then* $h \in L(\mu)$

(3.3.3) *If* $f \le g$ μ*-a.s., then* $\int f\,d\mu \le \int g\,d\mu$

(3.3.4) *If* $f = g$ μ*-a.s., then* $\int f\,d\mu = \int g\,d\mu$

(3.3.5) *If* $f \in L^1(\mu)$, $a \in \mathbf{R}_+$ *and* $|h| \le a|f|$ μ*-a.s. then* $h \in L^1(\mu)$

(3.3.6) *If* $\int f\,d\mu < \infty$, *then* $f < \infty$ μ*-a.s.*

(3.3.7) *If* $\int f\,d\mu > -\infty$, *then* $f > -\infty$ μ*-a.s.*

(3.3.8) $\int |f|\,d\mu = 0 \iff f = 0$ μ*-a.s.*

(3.3.9) $|\int f\,d\mu| \le \int |f|\,d\mu$

(3.3.10) $L^1(\mu)$ *is a vector space and* $f \curvearrowright \int f\,d\mu$ *is linear on* $L^1(\mu)$

(3.3.11) $\int (af)\,d\mu = a\int f d\mu \qquad \forall a \in \mathbf{R}$

(3.3.12) $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$

(3.3.13) *If* $0 \le h \le a$ *for some* $a \in \mathbf{R}_+$, *then* $h \cdot f \in L(\mu)$

where in (3.13.12), we must assume that $f(\omega) + g(\omega)$ *and* $\int f\,d\mu + \int g\,d\mu$ *are well-defined (that is, not of the form* $\infty - \infty$ *).*

Proof: All of the propositions (3.3.1)-(3.3.13) are easy consequences of (3.1.3)-(3.1.10) and definition (3.2.2). I will illustrate this by proving (3.3.1), (3.3.8) and (3.3.12), and leaving the proofs of the 10 other propositions to the reader.

(3.3.1): If $\int f\,d\mu > -\infty$, then $\int f^-\,d\mu < \infty$, and since $x \curvearrowright x^-$ is decreasing, we have that $h^- \le f^-$ μ-a.s. Hence, by (3.1.4) we have $\int h^-\,d\mu < \infty$ and so $h \in L(\mu)$.

(3.3.8): If $f = 0$ μ-a.s., then $f^+ = 0$ μ-a.s. and $f^- = 0$ μ-a.s. Hence, by (3.1.10) and (3.1.4) we have that $\int f^+\,d\mu = \int f^-\,d\mu = 0$. So by (3.1.8) we have that $\int |f|\,d\mu = 0$ because $|f| = f^+ + f^-$. Conversely, suppose that $\int |f|\,d\mu = 0$. Then by (3.1.3) we have that $s_n(|f|) = 0$ for all n . In particular, we see that $\mu(|f| > 2^{-n}) = 0$, and since $\{|f| > 2^{-n}\} \uparrow \{f \ne 0\}$, by (1.4.2) we have that $\mu(f \ne 0) = 0$. Hence $f = 0$ μ-a.s.

(3.3.12): By assumption, $h(\omega) = f(\omega) + g(\omega)$ exists μ-a.s. and since:

$$h = h^+ - h^- = (f^+ - f^-) + (g^+ - g^-)$$

we find $h^+ + f^- + g^- = f^+ + g^+ + h^-$. Hence, by (3.1.8) we find that

$$\int h^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu = \int f^+ \, d\mu + \int g^+ \, d\mu + \int h^- \, d\mu$$

Moreover, by assumption we have that the sum $\int f \, d\mu + \int g \, d\mu$ is well-defined, that is, not of the form $\infty - \infty$. But then we conclude that

$$\begin{aligned} \int h \, d\mu &= \int h^+ \, d\mu - \int h^- \, d\mu \\ &= \int f^+ \, d\mu - \int f^- \, d\mu + \int g^+ \, d\mu - \int g^- \, d\mu \\ &= \int f \, d\mu + \int g \, d\mu \end{aligned}$$

and so (3.3.12) holds. □

3.4 Monotone Convergence Theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f, f_1, f_2, \cdots$ be measurable functions from Ω into $\overline{\mathbf{R}}$. If $f_n \uparrow f$ μ-a.s. and $\int f_1 \, d\mu > -\infty$, then we have

$$f \in L(\mu) \quad \text{and} \quad \int_\Omega f \, d\mu = \lim_{n \to \infty} \int_\Omega f_n \, d\mu \tag{3.4.1}$$

Similarly, if $f_n \downarrow f$ μ-a.s. and $\int f_1 \, d\mu < \infty$, then (3.4.1) holds.

Proof: Note that we have already proved the monotone convergence theorem, when f_n is non-negative and $f_n(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega$ [see (3.1.9)]. The point of the proof below is to replace $f_1, f_2, \ldots$ by appropiately defined a non-negative functions $h_1, h_2, \ldots$ such that $h_n \uparrow h$, and then apply (3.1.9).

Suppose that $f_n \uparrow f$ μ-a.s. and $\int f_1 \, d\mu > -\infty$. Since $f_n \geq f_1$ μ-a.s. and $f \geq f_1$ μ-a.s., we see that $f_n \in L(\mu)$ and $f \in L(\mu)$. Let $N \in \mathcal{F}$ be a μ-null set such that $f_n(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega \setminus N$, and set

$$g_n = 1_{\Omega \setminus N} \cdot f_n, \qquad g = 1_{\Omega \setminus N} \cdot f$$

Then $g_n = f_n$ μ-a.s., $g = f$ μ-a.s. and $g_n \uparrow g$. Moreover, since

$$\int g_1 \, d\mu = \int f_1 \, d\mu > -\infty$$

we have that $g_1^- \in L^1(\mu)$ and $h_n = g_n + g_1^- \geq g_1 + g_1^- \geq 0$. Since $h_n \uparrow g + g_1^-$, then by (3.1.9) we get

$$\int (g + g_1^-) \, d\mu = \lim_{n\to\infty} \int (g_n + g_1^-) \, d\mu$$

So by (3.3.4) and (3.3.12) we have

$$\begin{aligned} \int f \, d\mu = \int g \, d\mu &= \int (g + g_1^-) \, d\mu - \int g_1^- \, d\mu \\ &= \lim_{n\to\infty} \left\{ \int (g_n + g_1^-) \, d\mu - \int g_1^- \, d\mu \right\} \\ &= \lim_{n\to\infty} \int g_n \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu \end{aligned}$$

Suppose that $f_n \downarrow f$ μ-a.s. and $\int f_1 \, d\mu < \infty$. If $\hat{f}_n = -f_n$ and $\hat{f} = -f$, then $\hat{f}_n \uparrow \hat{f}$ and $\int \hat{f}_1 \, d\mu > -\infty$. Thus, the second part of the theorem follows from the first part. □

3.5 Fatou's Lemma

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f_1, f_2, \cdots$ be measurable functions from Ω into $\overline{\mathbf{R}}$. If there exists a function $h \in L^1(\mu)$ such that $f_n \geq h$ μ-a.s. for all $n \geq 1$, then

$$\int_\Omega \liminf_{n\to\infty} f_n(\omega) \, \mu(d\omega) \leq \liminf_{n\to\infty} \int_\Omega f_n(\omega) \, \mu(d\omega) \tag{3.5.1}$$

In particular, if $f_n \geq 0$ μ-a.s. for all $n \geq 1$, then (3.5.1) holds. Similarly, if there exists a function $h \in L^1(\mu)$ such that $f_n \leq h$ μ-a.s. for all $n \geq 1$, then we have

$$\int_\Omega \limsup_{n\to\infty} f_n(\omega) \, \mu(d\omega) \geq \limsup_{n\to\infty} \int_\Omega f_n(\omega) \, \mu(d\omega) \tag{3.5.2}$$

In particular, if $f_n \leq 0$ μ-a.s. for all $n \geq 1$, then (3.5.2) holds.

Proof: Let $g_n(\omega) = \inf_{j \geq n} f_j(\omega)$, then $g_1 \geq h$ μ-a.s. and, thus, we have

$$\int g_1 \, d\mu \geq \int h \, d\mu > -\infty \quad \text{and} \quad g_n \uparrow \liminf_{n\to\infty} f_n$$

Hence, by the monotone convergence theorem (3.4) we find

$$\int_\Omega \liminf_{n\to\infty} f_n(\omega) \, \mu(d\omega) = \lim_{n\to\infty} \int g_n \, d\mu$$

But obviously, $g_n \leq f_n$, and so $\int g_n \, d\mu \leq \int f_n \, d\mu$. Thus,

$$\int (\liminf_n f_n) \, d\mu = \lim_{n\to\infty} \int g_n \, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu$$

and so (3.5.1) is proved. (3.5.2) is proved similarly. □

3.6 Lebesgue's Dominated Convergence Theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f, f_1, f_2, \cdots$ be measurable functions such that $f_n \to f$ μ-a.s. If there exists a function $h \in L^1(\mu)$ such that $|f_n| \leq h$ μ-a.s. for all $n = 1, 2, \cdots$, then f_n and f are μ-integrable and

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu \tag{3.6.1}$$

$$\lim_{n\to\infty} \int |f - f_n| \, d\mu = 0 \tag{3.6.2}$$

Proof: Since $|f_n| \leq h$ μ-a.s. and $|f_n| \to |f|$ μ-a.s., we see that $|f| \leq h$ μ-a.s. Hence, f_n and f are μ-integrable by (3.3.5). Let

$$f^* = \limsup_{n\to\infty} f_n \quad , \quad f_* = \liminf_{n\to\infty} f_n$$

Then $f_* \leq f^*$ and $f_* = f = f^*$ μ-a.s. because $f_n \to f$ μ-a.s.. Since $f_n \leq h$ μ-a.s. and $f_n \geq -h$ μ-a.s., then by Fatou's lemma (3.5) and (3.3.4) we get

$$\begin{aligned} \limsup_{n\to\infty} \int f_n \, d\mu \leq \int f^* \, d\mu = \int f \, d\mu = \int f_* \, d\mu \\ \leq \liminf_{n\to\infty} \int f_n \, d\mu \end{aligned}$$

because $f_* = f = f^*$ μ-a.s. Hence (3.6.1) holds. Finally, since $|f_n - f| \to 0$ μ-a.s. and $|f_n - f| \leq 2h$ μ-a.s., we see that (3.6.2) follows from (3.6.1). □

3.7 Beppo Levi's Theorem

Let $(\Omega, \mathcal{F}, \mu)$ *be a measure space and let* $f_1, f_2, \cdots$ *be measurable functions from* Ω *into* $\mathbf{R}$. *If* $\sum_{n=1}^{\infty} \int f_n^+ \, d\mu < \infty$, *then*

(3.7.1) $\sum_{n=1}^{\infty} f_n(\omega)$ *exists and is* $< \infty$ μ*-a.s.*

(3.7.2) $$\int_\Omega \left\{ \sum_{n=1}^{\infty} f_n(\omega) \right\} \mu(d\omega) = \sum_{n=1}^{\infty} \left\{ \int_\Omega f_n(\omega) \, \mu(d\omega) \right\}$$

Similarly, if $\sum_{n=1}^{\infty} \int f_n^- \, d\mu < \infty$, *then*

(3.7.3) $\sum_{n=1}^{\infty} f_n(\omega)$ *exists and is* $> -\infty$ μ*-a.s.*

(3.7.4) $$\int_\Omega \left\{ \sum_{n=1}^{\infty} f_n(\omega) \right\} \mu(d\omega) = \sum_{n=1}^{\infty} \left\{ \int_\Omega f_n(\omega) \, \mu(d\omega) \right\}$$

In particular, if $f_n \geq 0$ μ*-a.s. for all* $n \geq 1$, *or if* $f_n \leq 0$ μ*-a.s. for all* $n \geq 1$, *then (3.7.2) holds.*

Remark: Recall [see (P.4) and (1.19)] that (3.7.1) means that there exists a μ-null set $N \in \mathcal{F}$ such that

(3.7.5) $f(\omega) = \lim_{k \to \infty} (f_1(\omega) + \cdots + f_k(\omega))$ exists for all $\forall \omega \in \Omega \setminus N$

(3.7.6) $-\infty \leq f(\omega) < \infty \qquad \forall \omega \in \Omega \setminus N$

and (3.7.3) has a similar meaning.

Proof: Let $g = \sum_1^\infty f_n^+$ and $h = \sum_1^\infty f_n^-$. Since the terms are non-negative, the two sums exist everywhere. Moreover, since

$$f_1^+ + \cdots + f_k^+ \uparrow g \quad \text{and} \quad f_1^- + \cdots + f_k^- \uparrow h$$

then by the monotone convergence theorem (3.4) and linearity of the integral [see (3.3.12)], we have

(i) $$\int g \, d\mu = \sum_{n=1}^{\infty} \int f_n^+ \, d\mu$$

(ii) $$\int h \, d\mu = \sum_{n=1}^{\infty} \int f_n^- \, d\mu$$

Hence, if $\sum \int f_n^+ \, d\mu < \infty$, then by (3.3.6) there exists a μ-null set $N \in \mathcal{F}$ such that $g(\omega) < \infty$ for $\omega \in \Omega \setminus N$. Thus, if $\omega \in \Omega \setminus N$, then we have the limit $f(\omega) := \lim_{n\to\infty} (f_1(\omega) + \cdots + f_n(\omega))$ exists and

$$f(\omega) = \lim_{n\to\infty} \left\{ \sum_{j=1}^{n} f_j^+(\omega) - \sum_{j=1}^{n} f_j^-(\omega) \right\} = g(\omega) - h(\omega) < \infty$$

Thus, (3.7.1) holds, and since $f \le g$ μ-a.s. and $\int g \, d\mu < \infty$, we have that $f \in L(\mu)$. From (3.3.11), (3.3.12), (i) and (ii) we find

$$\begin{aligned} \int f \, d\mu &= \int g \, d\mu - \int h \, d\mu \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} \int f_n^+ \, d\mu - \lim_{k\to\infty} \sum_{n=1}^{k} \int f_n^- \, d\mu \\ &= \lim_{k\to\infty} \sum_{n=1}^{k} \int f_n \, d\mu \end{aligned}$$

because $\sum_1^\infty \int f_n^+ \, d\mu < \infty$. Hence, (3.7.2) holds. The second part of the theorem is proved similarly. □

3.8 Integrals of Simple and Discrete Functions

The four convergence theorems (3.4)–(3.7) represent the culmination of the theory of integration. They tell us exactly how much we can juggle around with integrals, limits and infinite sums. They are important both conceptually and in practical computation of integrals. It is the convergence theorems (3.4)–(3.7), which show that the integral defined by (3.1.1) and (3.1.2) is superior to the classical Riemann integral. Propositions (3.3.1)–(3.3.13) give the elementary rules for computing integrals, but we have not yet seen how to compute integrals of concrete functions.

So let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f = 1_F$ where $F \in \mathcal{F}$. Then evidently we have that $\{f > j2^{-n}\} = F$ if $1 \le j \le 2^n - 1$ and $\{f > j2^{-n}\} = \emptyset$ if $j \ge 2^n$. Hence, we find

$$s_n(f) = 2^{-n} \sum_{j=1}^{\infty} \mu(f > j2^{-n}) = (1 - 2^{-n})\mu(F)$$

and since $2^{-n} \to 0$ as $n \to \infty$, we see that

$$\int_F 1 \, d\mu = \int_\Omega 1_F \, d\mu = \mu(F) \qquad \forall F \in \mathcal{F} \tag{3.8.1}$$

Thus, by Beppo Levi's theorem (3.7) we have

$$\int_\Omega \left(\sum_{n=1}^{\infty} a_n \, 1_{F_n} \right) d\mu = \sum_{n=1}^{\infty} a_n \, \mu(F_n) \tag{3.8.2}$$

whenever $a_n \in \mathbf{R}$ and $F_n \in \mathcal{F}$ satisfies

$$\text{Either} \quad \sum_{n=1}^{\infty} a_n^+ \, \mu(F_n) < \infty \quad \text{or} \quad \sum_{n=1}^{\infty} a_n^- \, \mu(F_n) < \infty \tag{3.8.3}$$

In particular, if f is a discrete function from Ω into $\mathbf{R}$ [see (1.45)] and if $a_1, a_2, \cdots$ are the possible values of f, then we have

$$\int f \, d\mu = \sum_n a_n \, \mu(f = a_n) \tag{3.8.4}$$

provided that we have

$$\text{Either} \quad \sum_{n=1}^{\infty} a_n^+ \, \mu(f = a_n) < \infty \quad \text{or} \quad \sum_{n=1}^{\infty} a_n^- \, \mu(f = a_n) < \infty \tag{3.8.5}$$

Thus, (3.8.4) allows us to compute the integral of every discrete function f. Having established the four convergence theorems for integrals we shall now turn to four important inequalities for integrals.

3.9 Markov's Inequality

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\varphi : \overline{\mathbf{R}}_+ \to \overline{\mathbf{R}}_+$ be an increasing function. If $a > 0$ and $0 < \varphi(a) < \infty$, then we have

$$\mu(|f| \geq a) \leq \frac{1}{\varphi(a)} \int_\Omega \varphi(|f(\omega)|) \, \mu(d\omega) \tag{3.9.1}$$

Remark: Andrei Andreiwich Markov (1856–1922) was a student of Lvovich Chebyshev (1821–1884). Together with his teacher Chebyshev and one of his fellow students, Alexander Mikhailovich Lyapounov (1857–1918), Markov was the forerunner of modern probability.

Proof: Let $F = \{|f| \geq a\}$ and $g(\omega) = \varphi(|f(\omega)|)$. If $\omega \in F$, then $g(\omega) \geq \varphi(a)$ because φ is increasing, and $g \geq 0$ because $\varphi(x) \geq 0$ for all $x \geq 0$. Hence, $g \geq \varphi(a) \, 1_F$, so by (3.3.11) and (3.8.1.) we have $\int g \, d\mu \geq \varphi(a) \, \mu(F)$. Dividing both sides by $\varphi(a)$, we see that (3.9.1) follows. □

3.10 Jensen's Inequality

Let $(\Omega, \mathcal{F}, \mu)$ *be a measure space, let* $F \in \mathcal{F}$ *such that* $0 < \mu(F) < \infty$ *, and let* $f \in L^1(\mu)$ *. Suppose that* I *is an interval on* $\mathbf{R}$ *such that* $f(\omega) \in I$ *for all* $\omega \in F$ *and that* φ *is a convex function from* I *into* $\mathbf{R}$ *. Then*

$$\frac{1}{\mu(F)} \int_F f \, d\mu \in I \tag{3.10.1}$$

$$\varphi\left(\frac{1}{\mu(F)} \int_F f \, d\mu\right) \leq \frac{1}{\mu(F)} \int_F \varphi(f(\omega)) \; \mu(d\omega) \tag{3.10.2}$$

Remark: Johan Ludvig William Valdemar Jensen (1859–1925) was a Danish mathematician. Jensen and his fellow mathematician A.K. Erlang (1878–1929) were among the pioneers, who introduced the telephone in Denmark. He was director of the Copenhagen Telephone Company from 1890 to 1925.

Proof: Let $m = \mu(F)^{-1} \int_F f \, d\mu$ and let a and b denote the left and right end points of I . Then $a 1_F \leq 1_F f \leq b 1_F$ because $f(\omega) \in I$ for all $\omega \in F$. Hence,

$$a \, \mu(F) \leq \int_F f \, d\mu \leq b \, \mu(F)$$

Thus, we see that $a \leq m \leq b$. If $m = a$, then

$$\int 1_F (f - a) \, d\mu = 0 \qquad \text{and} \qquad 1_F (f - a) \geq 0$$

Thus, by (3.3.8) we have that $1_F (f - a) = 0$ μ-a.s., and so $f = a$ μ-a.s. on F . Hence, $m = a \in I$ and

$$\varphi(m) = \frac{1}{\mu(F)} \int_F \varphi(m) \, d\mu = \frac{1}{\mu(F)} \int_F \varphi(f(\omega)) \; \mu(d\omega)$$

Thus, (3.10.1) and (3.10.2) hold. In a similar manner, we see that (3.10.1) and (3.10.2) hold if $m = b$.

So suppose that $a < m < b$. Then m belongs to the interior of the interval I and if λ denotes the right-hand derivative of φ of the point m [which exists by convexity of φ ; see Exc.3.33], we have

$$\varphi(x) \geq \lambda (x - m) + \varphi(m)$$

[The graph of a convex function lies above its tangents.] Hence, we have

$$\varphi(f(\omega)) \geq \lambda (f(\omega) - m) + \varphi(m) \qquad \forall \omega \in F$$

and so we find

$$\int_F (\varphi \circ f)\, d\mu \geq \lambda \left(\int_F f\, d\mu - m\mu(F) \right) + \varphi(m)\mu(F)$$
$$= \varphi(m)\, \mu(F)$$

because $m\,\mu(F) = \int_F f\, d\mu$. Dividing through by $\mu(F)$ we see that (3.10.1) and (3.10.2) hold. □

3.11 Hölder's Inequality

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f and g be measurable functions from Ω into $\overline{\mathbf{R}}$. If p, q and r are positive numbers such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, then we have

$$\left\{ \int_\Omega |fg|^r\, d\mu \right\}^{1/r} \leq \left\{ \int_\Omega |f|^p\, d\mu \right\}^{1/p} \left\{ \int_\Omega |g|^q\, d\mu \right\}^{1/q} \tag{3.11.1}$$

Remark: If $p = q = 2$, then $r = 1$ and in this case, inequality (3.11.1) is called *the Cauchy-Schwarz inequality.*

Proof: Since $\log x$ is concave on $]0, \infty[$, we have

$$\log(tx + (1-t)y) \geq t \log x + (1-t) \log y$$

for all $x, y > 0$ and all $0 \leq t \leq 1$. Exponentiating this inequality we find

$$tx + (1-t)y \geq x^t y^{1-t} \qquad \forall\, x, y \geq 0 \qquad \forall\, 0 < t < 1$$

Set $t = \frac{r}{p}$, then $0 < t < 1$ and $1 - t = \frac{r}{q}$ because $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Let a and b be non-negative numbers and set $x = |a\, f(\omega)|^p$ and $y = |b\, g(\omega)|^q$. Then we have

$$a^r b^r |f(\omega) g(\omega)|^r = x^t y^{1-t} \leq t a^p |f(\omega)|^p + (1-t) b^q |g(\omega)|^q$$

and so

$$a^r b^r \int |fg|^r\, d\mu \leq t a^p \int |f|^p\, d\mu + (1-t) b^q \int |g|^q\, d\mu$$
$$= t a^p \alpha^p + (1-t) b^q \beta^q$$

where α and β are defined by

$$\alpha = \left\{ \int |f|^p\, d\mu \right\}^{1/p} \quad , \quad \beta = \left\{ \int |g|^q\, d\mu \right\}^{1/q}$$

If $\alpha = 0$, then $f = fg = 0$ μ-a.s., hence $\int |fg|^r \, d\mu = 0$ and so (3.11.1) holds. Similarly, we see that (3.11.1) holds if $\beta = 0$. If $\alpha > 0$ and $\beta > 0$, and either $\alpha = \infty$ or $\beta = \infty$, then (3.11.1) holds trivially because the right-hand side becomes infinite. So suppose that $0 < \alpha < \infty$ and $0 < \beta < \infty$, and set $a = \frac{1}{\alpha}$ and $b = \frac{1}{\beta}$. Then we find

$$\alpha^{-r}\beta^{-r} \int |fg|^r \, d\mu \leq t + (1-t) = 1$$

and so multiplying through with $\alpha^r \, \beta^r$ and taking the rth root we see that (3.11.1) follows. □

3.12 Minkowski's Inequality

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f and g be measurable functions from Ω into $\overline{\mathbf{R}}$. If $1 \leq p < \infty$, we have

$$\left\{\int_\Omega |f+g|^p \, d\mu\right\}^{1/p} \leq \left\{\int_\Omega |f|^p \, d\mu\right\}^{1/p} + \left\{\int_\Omega |g|^p \, d\mu\right\}^{1/p} \tag{3.12.1}$$

Moreover, if $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is increasing and concave, then

$$\int_\Omega \varphi(|f+g|) \, d\mu \leq \int_\Omega \varphi(|f|) \, d\mu + \int_\Omega \varphi(|g|) \, d\mu \tag{3.12.2}$$

In particular, if $0 < p \leq 1$, then

$$\int_\Omega |f+g|^p \, d\mu \leq \int_\Omega |f|^p \, d\mu + \int_\Omega |g|^p \, d\mu \tag{3.12.3}$$

Proof: (3.12.2): Let $x, y \geq 0$ and set $\lambda = \frac{x}{(x+y)}$ with the convention: $\frac{0}{0} = 0$. Then $x = \lambda(x+y) + (1-\lambda)0$ and $y = (1-\lambda)(x+y) + \lambda \cdot 0$. Thus, by concavity of φ we find

$$\begin{aligned} \varphi(x) &\geq \lambda\varphi(x+y) + (1-\lambda)\varphi(0) \geq \lambda\varphi(x+y) \\ \varphi(y) &\geq (1-\lambda)\varphi(x+y) + \lambda\varphi(0) \geq (1-\lambda)\varphi(x+y) \end{aligned}$$

because $\varphi(0) \geq 0$. Hence, $\varphi(x+y) \leq \varphi(x) + \varphi(y)$, and since φ is increasing, we have

$$\varphi(|f(\omega) + g(\omega)|) \leq \varphi(|f(\omega)| + |g(\omega)|) \leq \varphi(|f(\omega)|) + \varphi(|g(\omega)|)$$

Integrating this inequality, we obtain (3.12.2).

(3.12.3) follows from (3.12.2) since $x \curvearrowright x^p$ is increasing and concave on $\mathbf{R}_+$ for all $0 < p \leq 1$.

(3.12.1): If $p = 1$, then (3.12.1) follows from (3.12.3). So suppose that $1 < p < \infty$, then by Hölder's inequality with $r = 1$, $p = p$ and $q = \frac{p}{p-1}$ we have

$$\int |f|\,|f+g|^{p-1}\,d\mu \leq \left\{\int |f|^p\,d\mu\right\}^{1/p}\left\{\int |f+g|^{q(p-1)}\,d\mu\right\}^{1/q}$$
$$\int |g|\,|f+g|^{p-1}\,d\mu \leq \left\{\int |g|^p\,d\mu\right\}^{1/p}\left\{\int |f+g|^{q(p-1)}\,d\mu\right\}^{1/q}$$

Note that $p = q(p-1)$ and $|f+g|^p \leq |f|\,|f+g|^{p-1} + |g|\,|f+g|^{p-1}$; hence, summing the two inequalities we find:

$$\int |f+g|^p\,d\mu \leq \left\{\int |f+g|^p\,d\mu\right\}^{1/q}\left(\left\{\int |f|^p\,d\mu\right\}^{1/p} + \left\{\int |g|^p\,d\mu\right\}^{1/p}\right)$$

and since $\frac{1}{p} = 1 - \frac{1}{q}$, we obtain (3.12.1) by moving the first factor on the right-hand side of the inequality to the other side of the inequality. □

3.13 Continuity of the Integral

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f \in L(\mu)$. Suppose that $F_1, F_2, \cdots$ all belong to $\mathcal{F}$. Then we have

(3.13.1) $$\int_{F_n} f\,d\mu \to \int_F f\,d\mu \quad \text{if} \begin{cases} \text{either } F_n \uparrow F \\ \text{or } F_n \downarrow F \text{ and } f \in L^1(\mu) \end{cases}$$

(3.13.2) $$\int_F f\,d\mu = \sum_{n=1}^{\infty} \int_{F_n} f\,d\mu \quad \text{if } \{F_n\} \text{ are disjoint and } F = \bigcup_{n=1}^{\infty} F_n$$

(3.13.3) $$\lim_{\mu(A)\to 0} \int_A f\,d\mu = 0 \qquad \text{if } f \in L^1(\mu)$$

where (3.13.3) is to be understood in the following sense:

(3.13.4) $$\forall \varepsilon > 0\ \exists \delta > 0 \text{ so that } \left|\int_A f\,d\mu\right| \leq \varepsilon \quad \text{if } \mu(A) \leq \delta$$

Remark: Let ν be a function from $\mathcal{F}$ into $[-\infty, \infty]$ satisfying the following continuity condition:

(3.13.5) $\quad \forall \varepsilon > 0 \; \exists \delta > 0$ so that $|\nu(A)| \leq \varepsilon$ if $\mu(A) \leq \delta$

Then we say that ν is *μ-continuous*, and we write $\lim_{\mu(A)\to 0} \nu(A) = 0$. Thus, (3.13.3) states that the function $A \curvearrowright \int_A f\, d\mu$ is μ-continuous whenever f is μ-integrable. Suppose that $\nu : \mathcal{F} \to [-\infty, \infty]$ is μ-continuous; then evidently we have

(3.13.6) $\quad \mu(A) = 0 \quad \Rightarrow \quad \nu(A) = 0$

A function $\nu : \mathcal{F} \to [-\infty, \infty]$ satisfying (3.13.6) is said to be *μ-saturated.* Thus, every μ-continuous function $\nu : \mathcal{F} \to [-\infty, \infty]$ is μ-saturated.

Proof: (3.13.1): Suppose that $F_n \uparrow F$, then (3.13.1) follows from the monotone convergence theorem (3.4) applied to $1_{F_n} f^+$ and $1_{F_n} f^-$. Suppose that $F_n \downarrow F$ and $f \in L^1(\mu)$; then (3.13.1) follows from Lebesgue's dominated convergence theorem (3.6) applied to $1_{F_n} f$.

(3.13.2): Let $f_n = 1_{F_n} f$. Since $\{F_n\}$ are disjoint and $F = \cup_1^\infty F_n$ we have

$$1_F f = \sum_{n=1}^{\infty} f_n \quad , \quad 1_F f^+ = \sum_{n=1}^{\infty} f_n^+ \quad , \quad 1_F f^- = \sum_{n=1}^{\infty} f_n^-$$

and since $f \in L(\mu)$, we have that either f^+ or f^- is μ-integrable. Hence, (3.13.2) follows from Beppo Levi's theorem (3.7).

(3.13.3): Let $\varepsilon > 0$ be given and set

$$\nu(A) = \int_A |f|\, d\mu \qquad \forall A \in \mathcal{F}$$

Then ν is a measure on $\mathcal{F}$, and since $f \in L^1(\mu)$, we have that $\nu(\Omega) < \infty$ and, thus, $\mu(|f| = \infty) = \nu(|f| = \infty) = 0$. Note that $\{\, |f| > n\} \downarrow \{|f| = \infty\}$, so by (1.4.3) we have

$$\lim_{n\to\infty} \nu(|f| > n) = 0$$

Hence, there exists an integer $p \geq 1$ such that $\nu(|f| > p) \leq \frac{\varepsilon}{2}$. Set $\delta = \frac{\varepsilon}{2p}$, then by (3.3.9) we have

$$\begin{aligned} \left| \int_A f\, d\mu \right| &\leq \nu(A) \leq \nu(|f| > p) + \nu(A \cap \{|f| \leq p\}) \\ &\leq \tfrac{\varepsilon}{2} + p\,\mu(A) \leq \varepsilon \end{aligned}$$

whenever $A \in \mathcal{F}$ and $\mu(A) \leq \delta$. Thus, (3.13.3) is proved. $\square$

3.14 The Standard Proof

Let $(\Omega, \mathcal{F})$ *be a measurable space and let* $M^+(\Omega, \mathcal{F})$ *denote the set of all* $\mathcal{F}$*-measurable non-negative functions* $f : \Omega \to \overline{\mathbf{R}}_+$.

(1): Let L_1 *and* L_2 *be two functions from* $M^+(\Omega, \mathcal{F})$ *into* $\overline{\mathbf{R}}_+$ *satisfying*

(3.14.1) $\quad L_j(f+g) = L_j(f) + L_j(g), \qquad L_j(af) = aL_j(f) \qquad \forall a \geq 0$

(3.14.2) $\quad$ *If* $f_n \uparrow f$, *then* $L_j(f) = \lim_{n\to\infty} L_j(f_n)$

(3.14.3) $\quad L_1(1_F) = L_2(1_F) \qquad \forall F \in \mathcal{F}$

whenever $j = 1, 2$ *and* $f, g, f_1, f_2, \ldots$ *are non-negative measurable functions. Then we have*

(3.14.4) $\quad L_1(f) = L_2(f) \qquad \forall f \in M^+(\Omega, \mathcal{F})$

(2): Let $\mathcal{W} \subseteq \mathcal{V}$ *be two sets of functions from* Ω *into* $\overline{\mathbf{R}}_+$ *satisfying (1.49.1)–(1.49.5), and such that* $\mathcal{F} = \sigma(\mathcal{W})$. *Moreover, let* L_1 *and* L_2 *be two functions from* $\mathcal{V}$ *into* $\overline{\mathbf{R}}_+$ *satisfying (3.14.1+2) for all* $f, g, f_1, f_2, \ldots \in \mathcal{V}$, *and such that*

(3.14.5) $\quad$ *If* $a + 1_F \in \mathcal{V}$ *and* $L_1(a + 1_F) = L_2(a + 1_F)$ $\forall a > 0$, *then* 1_F , $1_{F^c} \in \mathcal{V}$ *and* $L_1(1_F) = L_2(1_F)$ *and* $L_1(1_{F^c}) = L_2(1_{F^c})$

(3.14.6) $\quad L_1(g) = L_2(g) \qquad \forall g \in \mathcal{W}$

Then $M^+(\Omega, \mathcal{F}) \subseteq \mathcal{V}$ *and we have*

(3.14.7) $\quad L_1(f) = L_2(f) \qquad \forall f \in M^+(\Omega, \mathcal{F})$

Remark: In spite of its simplicity (1) is extremely useful in showing equality between expressions containing integrals. Below I shall give numerous applications of this result, which is known as *the standard proof.* The idea is the following: We go from indicator functions to simple functions by linearity [see (3.14.1)] and we pass from simple functions to non-negative measurable functions by monotone convergence [see (3.14.2)]. The second part is more involved since it uses the functional Sierpińsky lemma (1.49), but for most purposes the simple proposition (1) will suffice.

Proof: (1): By (3.14.1)–(3.14.3) we have that $L_1(f) = L_2(f)$ for every non-negative simple function, and so (3.14.4) follows from (3.14.2) and (1.46).

(2): Let $\mathcal{V}_0 = \{f \in \mathcal{V} \mid L_1(f) = L_2(f)\}$, then $\mathcal{V}_0$ satisfies (1.49.1)–(1.49.3) by (3.14.1+2+5), and $\mathcal{V}_0 \supseteq \mathcal{W}$ by (3.14.6). Hence $M^+(\Omega, \mathcal{F}) \subseteq \mathcal{V}_0$ by the functional Sierpińsky lemma (1.49), and so (3.14.7) holds. □

3.15 The First Transformation Theorem

Let $(\Omega, \mathcal{F}, \mu)$ *be a measure space and let* T *be a measurable function from* $(\Omega, \mathcal{F})$ *into a measurable space* $(M, \mathcal{B})$ *. Let* μ_T *denote the image measure [see (1.44)] and let* $g : M \to \overline{\mathbf{R}}$ *be measurable. Then we have*

$$\int_M g(x)\mu_T(dx) = \int_\Omega g(T(\omega))\,\mu(d\omega) \tag{3.15.1}$$

in the sense that if one of the integrals exists, then so does the other and the two integrals are equal. Moreover, if μ *is given in the form:*

$$\mu(F) = \sum_{n=1}^{\infty} a_n\,\mu_n(F) \qquad \forall F \in \mathcal{F} \tag{3.15.2}$$

where $a_1, a_2, \ldots$ *are non-negative numbers and* $\mu_1, \mu_2, \ldots$ *are measures on* $(\Omega, \mathcal{F})$ *, then we have*

$$\int_\Omega f\,d\mu = \sum_{n=1}^{\infty} a_n \int_\Omega f\,d\mu_n \qquad \forall f \in L(\mu) \tag{3.15.3}$$

Remark: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. If μ can be written on the form (3.15.2), where $a_n = 1$ and μ_n is a finite measure for all $n \geq 1$, then we say that μ is *sum-finite*. Recall that a measure μ is *σ-finite* if there exist $\{F_n\} \in \mathcal{F}$ such that $\mu(F_n) < \infty$ for all $n \geq 1$ and $\Omega = \cup_{n=1}^{\infty} F_n$. A measure μ is called *finitely founded* if for every set $F \in \mathcal{F}$ with $\mu(F) > 0$, there exists a set $F_0 \in \mathcal{F}$ such that $F_0 \subseteq F$ and $0 < \mu(F_0) < \infty$. The Lebesgue measure on $\mathbf{R}^k$ is infinite, but σ-finite, sum-finite and finitely founded. The counting measure on an uncountable set is finitely founded, but neither σ-finite nor sum-finite.

Proof: (3.15.1): Let $L_1(g) = \int g\,d\mu_T$ and $L_2(g) = \int (g \circ T)\,d\mu$. Then L_1 and L_2 satisfies (3.14.1) and (3.14.2) by linearity of the integral and the monotone convergence theorem (3.4). Moreover, if $B \in \mathcal{B}$, then by (3.8.1) and (1.44.1) we have

$$L_1(1_B) = \mu_T(B) = \mu(T^{-1}(B)) = L_2(1_B)$$

because $1_B \circ T = 1_{T^{-1}(B)}$. Hence, (3.15.1) holds for all $g \in M^+(M, \mathcal{B})$ by (3.14.4). Thus, if $g : M \to \overline{\mathbf{R}}$ is measurable, then

$$\int_M g^+(x)\mu_T(dx) = \int_\Omega g^+(T(\omega))\,\mu(d\omega)$$
$$\int_M g^-(x)\mu_T(dx) = \int_\Omega g^-(T(\omega))\,\mu(d\omega)$$

and so (3.15.1) follows from definition (3.2.2) of the integral.

(3.15.3): Follows from the standard proof in exactly the same manner as above. □

3.16 Semi-finite Measures

Suppose that μ and ν are measures on $(\Omega, \mathcal{F})$ and that λ is a set function from $\mathcal{F}$ into $[0, \infty]$. Then we have the following:

(3.16.1) *Every finite measure is σ-finite*

(3.16.2) *Every σ-finite measure is sum-finite and finitely founded*

(3.16.3) *If μ is sum-finite, then there exists a finite measure ξ such that $\mathcal{N}(\mu) = \mathcal{N}(\xi)$*

(3.16.4) *If μ is sum-finite and ν is μ-saturated, then ν is sum–finite*

(3.16.5) *μ is sum-finite if and only if μ is ξ-saturated for some finite (or σ-finite or sum-finite) measure ξ*

(3.16.6) *If λ is μ-saturated and subadditive on $\mathcal{F}$ and if $\lambda(F_n) \to 0$ whenever $F_n \in \mathcal{F}$ and $F_n \downarrow \emptyset$, then λ is μ-continuous*

Proof: (3.16.1) and (3.16.2) are easy and I shall leave the proof to the reader.

(3.16.3): Let μ be a sum–finite measure, then $\mu = \sum \mu_n$, where μ_n is a finite measure. Setting $a_n = 2^{-n}/(1 + \mu_n(\Omega))$, we see that

$$\xi(F) = \sum_{n=1}^{\infty} a_n \mu_n(F) \qquad \forall F \in \mathcal{F}$$

is a finite measure [actually, $\xi(\Omega) \leq 1$] such that $\xi(F) = 0$ if and only if $\mu(F) = 0$. Thus, we have proved (3.16.3).

(3.16.4): Suppose that ν is μ-saturated and μ is sum–finite; then by (3.16.3) there exists a finite measure ξ such that ν is ξ-saturated. Let us choose $A_1, A_2, \cdots$ in $\mathcal{F}$ such that $\nu(A_n) < \infty$ and $\xi(A_n) \to \alpha$, where

$$\alpha = \sup\{\xi(F) \mid F \in \mathcal{F}\ ,\ \nu(F) < \infty\}$$

and consider the disjungation of $\{A_n\}$ [see (1.2.11–13)]:

$$B_1 = A_1, \qquad B_n = A_n \setminus \bigcup_{j=1}^{n-1} A_j \qquad B = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$$

Let $F \in \mathcal{F}$ be given. If $\nu(F \setminus B) < \infty$, then $\nu(A_n \cup (F \setminus B)) < \infty$ and $\xi(A_n \cup (F \setminus B)) = \xi(A_n) + \xi(F \setminus B) \leq \alpha$. Hence, we see that $\xi(F \setminus B) = 0$ and so $\nu(F \setminus B) = 0$; that is,

$$\nu(F \setminus B) = \begin{cases} 0 & \text{if } \xi(F \setminus B) = 0 \\ \infty & \text{if } \xi(F \setminus B) > 0 \end{cases}$$

Thus, if we set

$$\nu_{2n}(F) = \nu(F \cap B_n) \quad \text{and} \quad \nu_{2n-1}(F) = \xi(F \setminus B)$$

for $n = 1, 2, \cdots$, then $\nu_1, \nu_2, \cdots$ are finite measures such that

$$\nu(F) = \nu(F \cap B) + \nu(F \setminus B) = \sum_{n=1}^{\infty} \nu_n(F)$$

and so ν is sum-finite.

(3.16.5) follows easily from (3.16.3) and (3.16.4).

(3.16.6): Suppose that λ is not μ-continuous. Then by (3.13.5) there exists $\varepsilon_0 > 0$ and $F_n \in \mathcal{F}$ such that $\lambda(F_n) \geq \varepsilon_0$ and $\mu(F_n) \leq 2^{-n}$ for all $n \geq 1$. Let

$$F^k = \bigcup_{n=k}^{\infty} F_n \qquad F = \bigcap_{n=1}^{\infty} F^k = \limsup F_n$$

By (1.4.6) we have that $\mu(F) = 0$, hence $\lambda(F) = 0$ because λ is μ-saturated. By subadditivity of λ we have

$$\varepsilon_0 \leq \lambda(F_k) \leq \lambda(F^k \setminus F) + \lambda(F) = \lambda(F^k \setminus F)$$

because $F_k \subseteq (F^k \setminus F) \cup F$. But $F^k \setminus F \downarrow \emptyset$ and $\varepsilon_0 > 0$ which contradicts the last assumption on λ. Thus, λ must be μ-continuous. □

3.17 Indefinite Integrals

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\varphi : \Omega \to \overline{\mathbf{R}}_+$ be a non-negative $\mathcal{F}$-measurable function. Let $\mathcal{G} \subseteq \mathcal{F}$ be a given σ-algebra and set

(3.17.1) $$\nu(G) = \int_G \varphi \, d\mu \qquad \forall G \in \mathcal{G}$$

Then ν *is a measure on* $(\Omega, \mathcal{G})$ *and if* $f : \Omega \to \overline{\mathbf{R}}$ *is* $\mathcal{G}$*-measurable, then we have*

$$\int_\Omega f\, d\nu = \int_\Omega (f \cdot \varphi)\, d\mu \tag{3.17.2}$$

in the sense that if one of the integrals exists, then so does the other and they are equal. Moreover, let f *and* g *be* $\mathcal{G}$*-measurable functions such that* $f, g \in L(\mu)$ *and*

$$\int_G f\, d\mu \le \int_G g\, d\mu \qquad \forall G \in \mathcal{G} \tag{3.17.3}$$

If either $f \in L^1(\mu)$ *, or if* $g \in L^1(\mu)$ *, or if the restriction of* μ *to* $\mathcal{G}$ *is finitely founded, then* $f \le g$ μ*-a.s.*

Proof: By (3.13.2) we have that ν is a measure. Set $L_1(f) = \int f d\nu$ and $L_2(f) = \int f\varphi\, d\mu$. Then

$$L_1(1_G) = \nu(G) = \int_G \varphi\, d\mu = L_2(1_G) \qquad \forall G \in \mathcal{G}$$

Thus, by the standard proof we have that (3.17.2) holds for all $f \in M^+(\Omega, \mathcal{G})$. But then (3.17.2) follows for arbitrary $\mathcal{G}$-measurable functions by the definition of the integral [see (3.2.2)].

Suppose that $f \in L^1(\mu)$ or $g \in L^1(\mu)$ and let $D = \{f > g\}$. Then $D \in \mathcal{G}$ and $1_D(f - g) > 0$. But $\int_D (f - g)\, d\mu \le 0$ by (3.17.3) and (3.3.10). Thus, by (3.3.8) we have $\mu(D) = 0$, and so $f \le g$ μ-a.s.

Suppose that the restriction of μ to $\mathcal{G}$ is finitely founded. As above we set $D = \{f > g\}$, since the integrals $\int_D f\, d\mu$ and $\int_D g\, d\mu$ both might be infinite we can no longer subtract the integrals. But suppose that $\mu(D) > 0$. Then by finite foundedness there exists $G \subseteq D$ such that $G \in \mathcal{G}$ and $0 < \mu(G) < \infty$. Let $G_n = G \cap \{g \le n, f \ge -n\}$. Then $G_n \in \mathcal{G}$ and by (3.17.3) we have

$$-\infty < -n\,\mu(G_n) \le \int_{G_n} f\, d\mu \le \int_{G_n} g\, d\mu \le n\,\mu(G_n) < \infty$$

because $\mu(G_n) \le \mu(G) < \infty$. Hence, as above, we have $\mu(G_n) = 0$ and since $g(\omega) < \infty$ and $f(\omega) > -\infty$ for all $\omega \in D$, we have that $G = \cup_1^\infty G_n$. Hence, $\mu(G) = 0$ contradictory to our assumption. Thus, $\mu(D) = 0$ and so $f \le g$ μ-a.s. □

Remark: Setting $\mathcal{G} = \mathcal{F}$ in (3.17.1), we obtain a measure ν on $(\Omega, \mathcal{F})$, which is called *the indefinite* μ*-integral of* φ , and we write

$$d\nu = \varphi\, d\mu \quad \text{or} \quad \nu(d\omega) = \varphi(\omega)\, \mu(d\omega) \quad \Leftrightarrow \quad \nu(F) = \int_F \varphi d\mu \ \ \forall F \in \mathcal{F} \tag{3.17.4}$$

Note that (3.17.2) justifies this notation. Let μ and ν be two given measures on $\mathcal{F}$. If there exists a measurable function $\varphi : \Omega \to \overline{\mathbf{R}}_+$ satisfying (3.17.4), then we say that ν is *absolutely μ-continuous*, and we write $\nu \ll \mu$. Suppose that $\nu \ll \mu$; then the function φ is called a *μ-density* of ν, and is denoted $\frac{d\nu}{d\mu} := \varphi$. Note that if $\nu \ll \mu$, then by (3.17.2) and (3.17.3) we have the following:

(3.17.5) If μ is finitely founded, then any two μ-densities of ν agree μ-a.s.

(3.17.6) $$d\nu = \frac{d\nu}{d\mu}\, d\mu$$

(3.17.7) $$\int_\Omega f(\omega)\, \nu(d\omega) = \int_\Omega f(\omega) \cdot \frac{d\nu}{d\mu}(\omega)\, \mu(d\omega)$$

where (3.17.7) is to be understood in the usual sense: If one of the integrals exists then so does the other and they are equal. If φ is a μ-density of ν and if $\psi = \varphi$ μ-a.s, then by (3.3.4) we have that ψ is a μ-density of ν. This slight lack of uniqueness of densities usually causes no problem since, by (3.17.5), we have that any two densities agree μ-a.s., at least if μ is finitely founded. There has, however, been made several attempts to select a unique "canonical" density (for instance. using "approximate limits" or "approximate continuity" or "liftings," etc.), but none of these attempts has been really satisfactory. Actually, it has been shown that it is not possible to select a unique density in a satisfactory way. Thus, we have to live with the slight lack of uniqueness of densities.

Let μ and ν be two arbitrary measures on $(\Omega, \mathcal{F})$, then by (3.13.6) and (3.16.6) we have the following:

(3.17.8) If ν is μ-continuous, then ν is μ-saturated

(3.17.9) If $\nu \ll \mu$, then ν is μ-saturated

(3.17.10) If $\nu \ll \mu$ and ν is finite, then ν is μ-continuous

Moreover, we have the following important converse to (3.17.9):

3.18 Radon-Nikodym-Lebesgue's Theorem

Let μ be a σ-finite measure on $(\Omega, \mathcal{F})$ and let ν be a sum–finite measure on $(\Omega, \mathcal{F})$. Then there exists a μ-null set $N \in \mathcal{F}$ and a non-negative measurable function $\varphi : \Omega \to \overline{\mathbf{R}}_+$ satisfying

(3.18.1) $$\nu(F) = \nu(F \cap N) + \int_F \varphi\, d\mu \qquad \forall F \in \mathcal{F}$$

In particular, if μ is a σ-finite measure and ν is a μ-saturated measure, then ν is absolutely μ-continuous [see (3.16.4)].

Remark: Let μ and ν be two measures on the measurable space $(\Omega, \mathcal{F})$. If $\mu \ll \nu$ and $\nu \ll \mu$, then we say that μ and ν are *equivalent* and write $\mu \approx \nu$. If there exists a set $H \in \mathcal{F}$ such that $\mu(H) = 0$ and $\nu(H^c) = 0$, then we say that μ and ν are *singular* and we write $\mu \perp \nu$. Moreover, inspired by the theorem above we say that ν is *Radon-Nikodym μ-differentiable* with *Radon-Nikodym μ-derivative* φ and *μ-singular set* N if N is a μ-null set and φ is a non-negative measurable functionn satisfying (5.18.1) and if so, we write $\varphi = \frac{d\nu}{d\mu}$. Then by the theorem above we have that every sum-finite measure ν is Radon-Nikodym μ-differentiable for any σ-finite measure μ. Moreover, if $\nu \ll \mu$ or $\nu \perp \mu$, then, evidently, we have that ν is Radon-Nikodym μ-differentiable.

Proof: First we prove the theorem under the additional assumption that μ and ν are finite measures, and then we extend the result to the case where μ is σ-finite and ν is sum-finite. So suppose that μ and ν are finite measures and set

$$\mathcal{V} = \left\{ f \in M^+(\Omega, \mathcal{F}) \;\middle|\; \int_F f\, d\mu \le \nu(F) \quad \forall F \in \mathcal{F} \right\}$$

$$\alpha = \sup\left\{ \int_\Omega f\, d\mu \;\middle|\; f \in \mathcal{V} \right\} \;,\quad \beta = \sup\{\, \nu(B) \mid B \in \mathcal{F} \cap \mathcal{N}(\mu) \}$$

where $M^+(\Omega, \mathcal{F})$ denote the set of all measurable functions from Ω into $[0, \infty]$. By definition of α and β we can find functions $f_n \in \mathcal{V}$ and sets $N_n \in \mathcal{F} \cap \mathcal{N}(\mu)$ such that $\int f_n\, d\mu \to \alpha$ and $\nu(N_n) \to \beta$, and we set

$$\varphi_n = f_1 \vee \cdots \vee f_n \qquad \varphi = \sup_{n \ge 1} f_n \qquad N = \bigcup_{n=1}^{\infty} N_n$$

Then $N \in \mathcal{F} \cap \mathcal{N}(\mu)$ and $\nu(N) = \beta$ because $N_n \subseteq N$ for all $n \ge 1$. Let $f, g \in \mathcal{V}$and $F \in \mathcal{F}$ be given. Set $G = F \cap \{f \le g\}$ and $H = F \cap \{f > g\}$, then H and G are disjoint with union F, and since $f \vee g = g$ on G and $f \vee g = f$ on H, we get

$$\int_F (f \vee g)\, d\mu = \int_G g\, d\mu + \int_H f\, d\mu \le \nu(G) + \nu(H) = \nu(F)$$

Thus, $f \vee g \in \mathcal{V}$ for all $f, g \in \mathcal{V}$. In particular, we see that φ_n belongs to $\mathcal{V}$ and $\varphi_n \uparrow \varphi$. Hence, by the monotone convergence theorem (3.4) we have that $\varphi \in \mathcal{V}$ and $\int \varphi\, d\mu = \alpha$. Summing up, we have

$$(*) \qquad \int_\Omega \varphi\, d\mu = \alpha \quad \text{and} \quad \int_F \varphi\, d\mu \le \nu(F) \qquad \forall F \in \mathcal{F}$$

$$(**) \qquad N \in \mathcal{F} \;,\quad \nu(N) = \beta \quad \text{and} \quad \mu(N) = 0$$

and we shall now see that (3.18.1) follows from (*) and (**).

So let $F \in \mathcal{F}$ and $\varepsilon > 0$ be given and let us define the paving $\mathcal{H}$ and the number m_1 as follows:

$$\mathcal{H} = \left\{ H \in \mathcal{F} \,\middle|\, \nu(H) \leq \int_H \varphi \, d\mu + \varepsilon\mu(H) \right\}$$
$$m_1 = \sup\{\mu(H) \mid H \in \mathcal{H} \,,\ H \subseteq F\}$$

Then we may choose $H_1 \in \mathcal{H}$ such that $\mu(H_1) \geq \frac{1}{2} m_1$ and $H_1 \subseteq F$. Let $m_2 = \sup\{\mu(H) \mid H \in \mathcal{H} \,,\ H \subseteq F \setminus H_1\}$, and choose $H_2 \in \mathcal{H}$ such that $\mu(H_2) \geq \frac{1}{2} m_2$ and $H_2 \subseteq F \setminus H_1$. Continuing like this, we can construct a sequence of disjoint sets $\{H_n\} \subseteq \mathcal{H}$ such that $H_n \subseteq F$ and $\mu(H_n) \geq \frac{1}{2} m_n$, where

$$m_n = \sup\left\{ \mu(H) \,\middle|\, H \in \mathcal{H}, \quad H \subseteq F \setminus \bigcup_{j=1}^{n-1} H_j \right\}$$

Set $H = \cup_1^\infty H_n$. Then $H \subseteq F$, and since $\mathcal{H}$ is $(\sum c)$-stable and $m_n \leq 2\,\mu(H_n)$, we see that $H \in \mathcal{H}$ and $\sum_1^\infty m_n \leq 2\,\mu(H) < \infty$. In particular, we see that $m_n \to 0$. Let $M = F \setminus H$ and $G \in \mathcal{F}$ be given. If $G \cap M \in \mathcal{H}$, then $\mu(G \cap M) \leq m_n$ because $G \cap M \subseteq F \setminus (H_1 \cup \cdots \cup H_{n-1})$, and so $\mu(G \cap M) = 0$, but then

$$\int_G (\varphi + \varepsilon 1_M)\, d\mu = \int_G \varphi \, d\mu \leq \nu(G)$$

On the other hand, if $G \cap M \notin \mathcal{H}$, then

$$\begin{aligned}\int_G (\varphi + \varepsilon 1_M)\, d\mu &= \int_{G \setminus M} \varphi \, d\mu + \int_{G \cap M} \varphi \, d\mu + \varepsilon\mu(G \cap M) \\ &\leq \nu(G \setminus M) + \nu(G \cap M) \leq \nu(G)\end{aligned}$$

Hence, we see that $(\varphi + \varepsilon 1_M) \in \mathcal{V}$, and so we have

$$\alpha \geq \int_\Omega (\varphi + \varepsilon 1_M)\, d\mu = \int_\Omega \varphi \, d\mu + \varepsilon\mu(M) = \alpha + \varepsilon\mu(M)$$

Thus, $\mu(M) = 0$ and, hence,

$$\beta \geq \nu(N \cup M) = \nu(N) + \nu(M \setminus N) = \beta + \nu(M \setminus N)$$

showing that $\nu(M \setminus N) = 0$. Since $F \setminus N \subseteq H \cup (M \setminus N)$, we have

$$\nu(F \setminus N) \leq \nu(H) + \nu(M \setminus N) = \nu(H) \leq \int_F \varphi \, d\mu + \varepsilon\mu(\Omega)$$

because $H \in \mathcal{H}$ and $H \subseteq F \subseteq \Omega$. Moreover, because N is a μ-null set, then by (*) and (**), we find

$$\nu(F \cap N) + \int_F \varphi \, d\mu = \nu(F \cap N) + \int_{F \setminus N} \varphi \, d\mu \leq \nu(F \cap N) + \nu(F \setminus N)$$
$$= \nu(F) \leq \nu(F \cap N) + \int_F \varphi \, d\mu + \varepsilon \mu(\Omega)$$

for all $F \in \mathcal{F}$. Thus, (3.18.1) follows by letting $\varepsilon \to 0$.

Let us now consider the general case. Then $\nu = \sum \nu_n$ where ν_n is finite for all n , and there exist disjoint sets $\{A_k\} \subseteq \mathcal{F}$ such that $\mu(A_k) < \infty$ and $\Omega = \cup_1^\infty A_k$. Set

$$\nu_{nk}(F) = \nu_n(F \cap A_k), \qquad \mu_k(F) = \mu(F \cap A_k)$$

then ν_{nk} and μ_k are finite measures. Hence, there exists $N_{nk} \in \mathcal{F}$ and measurable functions $\varphi_{nk} : \Omega \to \overline{\mathbf{R}}_+$ such that $\mu(N_{nk} \cap A_k) = 0$ and

$$\nu_n(F \cap A_k) = \nu_n(F \cap A_k \cap N_{nk}) + \int_F 1_{A_k} \varphi_{nk} \, d\mu \qquad \forall F \in \mathcal{F}$$

Summing over k and setting

$$N_n = \bigcup_{k=1}^{\infty} (A_k \cap N_{nk}) \ , \ N = \bigcup_{n=1}^{\infty} N_n \ , \ \varphi_n = \sum_{k=1}^{\infty} 1_{A_k} \varphi_{nk} \ , \ \varphi = \sum_{k=1}^{\infty} \varphi_n$$

shows that

$$\nu_n(F) = \nu_n(F \cap N_n) + \int_F \varphi_n \, d\mu$$

$$\nu_n(F \cap N) = \nu_n(F \cap N \cap N_n) + \int_{F \cap N} \varphi_n \, d\mu = \nu_n(F \cap N_n)$$

because N is a μ-null set containing N_n . Thus, (3.18.1) follows by summing over n . □

3.19 Radon-Nikodym Derivatives

Let μ , ν and λ be measures on $(\Omega, \mathcal{F})$ such that ν is Radon-Nikodym μ-differentiable with μ-singular set N , and that μ is Radon-Nikodym λ-differentiable with λ-singular set L . Let ν_a and ν_s be the measures given by

$$\text{(3.19.1)} \qquad \nu_a(F) = \int_F \tfrac{d\nu}{d\mu}(\omega) \, \mu(d\omega) \ , \ \nu_s(F) = \nu(F \cap N) \quad \forall F \in \mathcal{F}$$

and let D_0 *and* D_∞ *be the sets given by*

$$D_0 = \left\{ \omega \in \Omega \;\middle|\; \frac{d\nu}{d\mu}(\omega) = 0 \right\} \quad , \quad D_\infty = \left\{ \omega \in \Omega \;\middle|\; \frac{d\nu}{d\mu}(\omega) = \infty \right\}$$

If $f : \Omega \to \overline{\mathbf{R}}$ *is a measurable function, then we have*

$$\int_\Omega f(\omega)\,\nu(d\omega) = \int_N f(\omega)\,\nu(d\omega) + \int_\Omega f(\omega)\,\frac{d\nu}{d\mu}(\omega)\,\mu(d\omega) \tag{3.19.2}$$

in the usual sense: If one of the two sides exists, then so does the other and they are equal. Moreover, we have the following:

(3.19.3) $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ *and* $\nu_s \perp \mu$. *Conversely, if* ν_0 *and* ν_1 *are two measures such that* $\nu = \nu_0 + \nu_1$, $\nu_0 \ll \mu$ *and* $\nu_1 \perp \mu$, *then* $\nu_0 = \nu_a$ *and* $\nu_1 = \nu_s$

(3.19.4) $\nu \perp \mu \iff \dfrac{d\nu}{d\mu} = 0 \;\; \mu\text{-a.s.} \iff \nu(N^c) = 0$

(3.19.5) $\nu \ll \mu \iff d\nu = \dfrac{d\nu}{d\mu}\,d\mu \iff \nu(N) = 0$

(3.19.6) *If* $\nu(\Omega) < \infty$, *then* $\nu \ll \mu$ *if and only if*

$$\nu(\Omega) = \int_\Omega \frac{d\nu}{d\mu}(\omega)\,\mu(d\omega)$$

(3.19.7) *If* $\lambda(N \setminus M) = 0$ *for some* ν*-null set* M *(for instance, if* $\nu \ll \mu$ *or if* $\lambda \ll \mu$ *), then* ν *is Radon-Nikodym* λ*-differentiable with* λ*-singular set* $(N \cup L) \setminus M$ *and*

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

(3.19.8) *If* $\nu(N \setminus D_\infty) = \nu(D_0 \cap N \setminus M) = \nu(D_\infty \setminus M) = 0$ *for some* μ*-null set* M *(for instance, if* $\nu \ll \mu$ *and* $\frac{d\nu}{d\mu} < \infty$ ν*-a.s.), then* μ *is Radon-Nikodym* ν*-differentiable with* ν*-singular set* $(D_0 \cup D_\infty) \setminus M$ *and*

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

(3.19.9) *If* $\nu \ll \mu$ *and* $0 < \frac{d\nu}{d\mu} < \infty$ μ*-a.s., then* $\mu \approx \nu$

(3.19.10) *If* ν *is finitely founded, then* $\frac{d\nu}{d\mu} < \infty$ μ*-a.s.*

Remark: The unique pair (ν_a, ν_s) satisfying (3.19.3) is called *the Lebesgue decomposition* of ν with respect to μ, and ν_a is called *the absolutely μ-continuous part* of ν, and ν_s is called the *μ-singular part* of ν.

Proof: (3.19.2): By definition we have that (3.19.2) holds whenever $f = 1_F$ for some $F \in \mathcal{F}$. But then (3.19.2) follows from the standard proof (3.14).

(3.19.3): The first part of (3.19.3) is evident. So suppose that $\nu = \nu_0 + \nu_1$ where $\nu_0 \ll \mu$ and $\nu_1 \perp \mu$. Then there exists a μ-null set $\hat{N} \in \mathcal{F}$ such that $\nu_1(\Omega \setminus \hat{N}) = 0$. Let $M = N \cup \hat{N}$. Then M is a μ-null set, and $\nu_s(M^c) = \nu_1(M^c) = \nu_0(M) = \nu_a(M) = 0$. Hence,

$$\nu_s(F) = \nu_s(F \cap M) = \nu(F \cap M) = \nu_0(F \cap M) + \nu_1(F \cap M) = \nu_1(F)$$

$$\nu_a(F) = \nu_a(F \setminus M) = \nu(F \setminus M) = \nu_0(F \setminus M) + \nu_1(F \setminus M) = \nu_0(F)$$

for all $F \in \mathcal{F}$. Thus, $\nu_s = \nu_1$ and $\nu_a = \nu_0$.

(3.19.4)–(3.19.6): Evident!

(3.19.7): Let $\hat{N} = (N \cup L) \setminus M$. Since $\mu(N) = \nu(M) = 0$, then by (3.18.1) and (3.19.2) we get

$$\begin{aligned}
\nu(F) &= \nu(F \cap N) + \int_{F \cap L} \frac{d\nu}{d\mu}\, d\mu + \int_F \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda}\, d\lambda \\
&= \nu(F \cap N) + \int_{F \cap (L \setminus N)} \frac{d\nu}{d\mu}\, d\mu + \int_F \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda}\, d\lambda \\
&= \nu(F \cap N) + \nu(F \cap (L \setminus N)) + \int_F \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda}\, d\lambda \\
&= \nu(F \cap (N \cup L)) + \int_F \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda}\, d\lambda \\
&= \nu(F \cap \hat{N}) + \int_F \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda}\, d\lambda
\end{aligned}$$

for all $F \in \mathcal{F}$; and since $\lambda(N \setminus M) = \lambda(L) = 0$, we see that (3.19.7) holds.

(3.19.8): Let $D = D_0 \cup D_\infty$ and $\hat{N} = D \setminus M$. By our convention $0 \cdot \infty = \frac{1}{\infty} = 0$, we see that $(\frac{d\nu}{d\mu}) \cdot (\frac{d\nu}{d\mu})^{-1} = 1_{\Omega \setminus D}$ and $\frac{d\nu}{d\mu}(\omega)^{-1} = 0$ for all $\omega \in D_\infty$. Since $\nu(N \setminus D_\infty) = 0$, by (3.19.2) we get:

$$\int_F \left(\frac{d\nu}{d\mu}\right)^{-1} d\nu = \int_{F \cap N} \left(\frac{d\nu}{d\mu}\right)^{-1} d\nu + \int_F \left(\frac{d\nu}{d\mu}\right)^{-1} \left(\frac{d\nu}{d\mu}\right) d\mu = \mu(F \setminus D)$$

for all $F \in \mathcal{F}$. Moreover, since $\mu(M) = 0$ and

$$\begin{aligned} 0 \le \nu(\hat{N}) &= \nu(D_\infty \setminus M) + \nu(D_0 \setminus M) \\ &= \nu(D_\infty \setminus M) + \nu(N \cap D_0 \setminus M) + \int_{D_0 \setminus M} \frac{d\nu}{d\mu}\, d\mu \\ &= \nu(D_\infty \setminus M) + \nu(N \cap D_0 \setminus M) = 0 \end{aligned}$$

we see that $\hat{N}$ is a ν-null set satisfying

$$\mu(F) = \mu(F \cap D) + \mu(F \setminus D) = \mu(F \cap \hat{N}) + \int_F \left(\frac{d\nu}{d\mu}\right)^{-1} d\nu$$

for all $F \in F$. Thus, (3.19.8) is proved.

(3.19.9): Easy consequence of (3.19.8).

(3.19.10). By (3.18.1) we have

$$\nu(F \cap D_\infty \setminus N) = \int_{F \cap D_\infty} \frac{d\nu}{d\mu}(\omega)\, \mu(d\omega) = \begin{cases} \infty & \text{if } \mu(F \cap D_\infty) > 0 \\ 0 & \text{if } \mu(F \cap D_\infty) = 0 \end{cases}$$

for all $F \in \mathcal{F}$. Thus, by finite foundedness of ν we see that $\mu(D_\infty) = 0$. □

3.20 Product Measures

Let $(S, \mathcal{A}, \mu)$ and $(T, \mathcal{B}, \nu)$ be two sum-finite measure spaces, and let $\{\mu_n\}$ and $\{\nu_k\}$ be two sequences of finite measures such that

$$\mu(A) = \sum_{n=1}^{\infty} \mu_n(A) \quad , \quad \nu(B) = \sum_{n=1}^{\infty} \nu_n(B)$$

for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$. Consider the product space $S \times T$ with its product σ-algebra $\mathcal{A} \otimes \mathcal{B}$. Let $n, k \in \mathbf{N}$ and let $\mathcal{V}_{nk}$ be the set of all non-negative measurable functions $f : S \times T \to \overline{\mathbf{R}}_+$ satisfying

(3.20.1) $\quad s \curvearrowright \int_T f(s,t)\nu_k(dt) \quad$ is $\mathcal{A}$-measurable: $S \to \overline{\mathbf{R}}_+$

(3.20.2) $\quad t \curvearrowright \int_S f(s,t)\mu_n(ds) \quad$ is $\mathcal{B}$-measurable: $T \to \overline{\mathbf{R}}_+$

Then evidently $\mathcal{V}_{nk}$ is a $(\uparrow)$-stable convex cone, and since μ_n and ν_k are finite measures, it follows easily that $\mathcal{V}_{nk}$ satisfies (1.49.3). Let

$$L_{nk}(f) = \int_S \left\{ \int_T f(s,t)\nu_k(dt) \right\} \mu_n(ds)$$
$$M_{nk}(f) = \int_T \left\{ \int_S f(s,t)\mu_n(ds) \right\} \nu_k(dt)$$

for all $f \in \mathcal{V}_{nk}$. Then L_{nk} and M_{nk} satisfy (3.14.1), (3.14.2) and (3.14.5). Let $\mathcal{G} = \{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}$ and let $\mathcal{W} = \{1_G \mid G \in \mathcal{G}\}$. Then $\mathcal{G}$ is $(\cap f)$-stable, and so $\mathcal{W}$ satisfies (1.49.4). Moreover, since

$$1 - 1_{A\times B} = 1_{A^c \times B} + 1_{S \times B^c}$$

we see that $\mathcal{W}$ satisfies (1.49.5). Let $A \in \mathcal{A}$ and $B \in \mathcal{B}$ be given; then

$$\int_T 1_{A\times B}(s,t)\,\nu_k(dt) = 1_A(s)\,\nu_k(B)$$
$$\int_S 1_{A\times B}(s,t)\,\mu_n(ds) = 1_B(t)\,\mu_n(A)$$
$$L_{nk}(1_{A\times B}) = \mu_n(A)\,\nu_k(B) = M_{nk}(1_{A\times B})$$

Hence, $\mathcal{W} \subseteq \mathcal{V}_{nk}$ and $L_{nk}(f) = M_{nk}(f)$ for all $f \in \mathcal{W}$. Thus, by the standard proof (3.14), we have that $M^+(S \times T, \mathcal{A} \otimes \mathcal{B}) \subseteq \mathcal{V}_{nk}$ and

$$L_{nk}(f) = M_{nk}(f) \qquad \forall f \in M^+(S \times T, \mathcal{A} \otimes \mathcal{B})$$

Hence, if $f : S \times T \to \overline{\mathbf{R}}_+$ is $\mathcal{A} \otimes \mathcal{B}$- measurable, then by (3.15.3) we have

(3.20.3) $$F(s) = \int_T f(s,t)\,\nu(dt) \quad \text{is } \mathcal{A}\text{-measurable}$$

(3.20.4) $$G(t) = \int_S f(s,t)\,\mu(ds) \quad \text{is } \mathcal{B}\text{-measurable}$$

(3.20.5) $$\int_S d\mu \int_T f\,d\nu = \int_T d\nu \int_S f\,d\mu$$

where we use the shorthand notation

$$\int_S d\mu \int_T f\,d\nu := \int_S \left\{ \int_T f(s,t)\,\nu(dt)) \right\} \mu(ds)$$
$$\int_T d\nu \int_S f\,d\mu := \int_T \left\{ \int_S f(s,t)\,\mu(ds) \right\} \nu(dt)$$

In particular, we may define *the product measure of* μ *and* ν, denoted $\mu \otimes \nu$, by the formula

(3.20.6) $$\mu \otimes \nu(F) := \int_S d\mu \int_T 1_F\,d\nu = \int_T d\nu \int_S 1_F\,d\mu \qquad \forall F \in \mathcal{A} \otimes \mathcal{B}$$

By Beppo Levi's theorem, we see that $\mu \otimes \nu$ is a measure on $\mathcal{A} \otimes \mathcal{B}$. Setting $F = A \times B$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$, we find

$$\begin{aligned}\mu \otimes \nu(A \times B) &= \int_S \mu(ds) \int_T 1_A(s)\, 1_B(t)\, \nu(dt) \\ &= \int_S 1_A(s)\, \nu(B) \mu(ds) = \mu(A)\, \nu(B)\end{aligned}$$

because $1_{A \times B}(s,t) = 1_A(s) 1_B(t)$. Hence, the product measure $\mu \otimes \nu$ satisfies

(3.20.7) $\quad \mu \otimes \nu(A \times B) = \mu(A)\, \nu(B) \qquad \forall A \in \mathcal{A} \quad \forall B \in \mathcal{B}$

That is, the "area" of a "rectangle" equals the product of the "lengths" of the "sides." Note that by the uniqueness theorem (1.7), we have that

(3.20.8) If μ and ν are σ-finite, and λ is a measure on $\mathcal{A} \otimes \mathcal{B}$ such that $\lambda(A \times B) = \mu(A)\, \nu(B)$ for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$, then we have that $\lambda = \mu \otimes \nu$

Observe that

$$\mu \otimes \nu(F) = \int_{S \times T} 1_F \, d(\mu \otimes \nu) = \int_S d\mu \int_T 1_F \, d\nu = \int_T d\nu \int_S 1_F \, d\mu$$

Hence, by the standard proof we obtain the following very important result about interchanging the order of integration:

3.21 Tonelli-Fubini's Theorem

Let $(S, \mathcal{A}, \mu)$ and $(T, \mathcal{B}, \nu)$ be two sum-finite measure spaces and let $\mu \otimes \nu$ be the product measure on $(S \times T, \mathcal{A} \otimes \mathcal{B})$ of μ and ν. If $f : S \times T \to \overline{\mathbf{R}}_+$ is measurable, then

(3.21.1) $$\int_{S \times T} f \, d(\mu \otimes \nu) = \int_S d\mu \int_T f \, d\nu = \int_T d\nu \int_S f \, d\mu$$

and if $h : S \times T \to \overline{\mathbf{R}}$ is measurable such that just one of the following six integrals is finite:

$$\begin{array}{ccccc} \displaystyle\int_{S \times T} h^+ d(\mu \otimes \nu) & , & \displaystyle\int_S d\mu \int_T h^+ \, d\nu & , & \displaystyle\int_T d\nu \int_S h^+ \, d\mu \\ \displaystyle\int_{S \times T} h^- d(\mu \otimes \nu) & , & \displaystyle\int_S d\mu \int_T h^- \, d\nu & , & \displaystyle\int_T d\nu \int_S h^- \, d\mu \end{array}$$

then we have

$$\int_{S\times T} h\, d(\mu\otimes\nu) = \int_S d\mu \int_T h\, d\nu = \int_T d\nu \int_S h\, d\mu \tag{3.21.2}$$

Remark: The first part of the theorem, that is, (3.21.1) for non-negative functions f is called *Tonelli's theorem.* The second part, that is, (3.21.2) for $\overline{\mathbf{R}}$-valued functions h, is called *Fubini's theorem.* The theorem above is a very powerful tool as we shall see in the following sections and in the exercises.

3.22 Convergence in q-Mean

Up to now we have met three kinds of convergence notions for sequences of real-valued functions, namely, pointwise convergence [see (P.6)], uniform convergence [see (P.6)] and μ-a.s. convergence [see (1.19)]; they are denoted

$$f_n \to f \quad , \quad f_n \rightrightarrows f \quad \text{and} \quad f_n \to f \quad \mu\text{-a.s.}$$

respectively. I shall now introduce some convergence notions which connects convergence of functions with the theory of integration. So let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f, f_1, f_2, \ldots$ be measurable functions from Ω into the extended real line $\overline{\mathbf{R}}$. Then we say that $\{f_n\}$ *converges in μ-measure* to f if for all $\varepsilon > 0$ we have that $\mu(|f_n - f| > \varepsilon) \to 0$ as $n \to \infty$ and if so, we write

$$f_n \to f \text{ in } \mu\text{-measure} \quad \Leftrightarrow \quad \lim_{n\to\infty} \mu(|f_n - f| > \varepsilon) = 0 \quad \forall \varepsilon > 0 \tag{3.22.1}$$

where we use the convention: $\infty - \infty = 0$. We say that $\{f_n\}$ *converges μ-a.s uniformly* to f and we write $f_n \rightrightarrows f$ μ-a.s. if there exists a μ-null set N such that $\{f_n\}$ converges uniformly to f on $\Omega \setminus N$.

Let $\mathcal{L}(0,\infty)$ denote the set of all increasing, left continuous functions $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ such that $\varphi(x) > 0$ for all $x > 0$ and $\varphi(0) = \lim_{x\downarrow 0} \varphi(x) = 0$. If $\varphi \in \mathcal{L}(0,\infty)$, then we define $|f|_\varphi$ and convergence in *φ-mean* as follows:

$$|f|_\varphi = \int_\Omega \varphi(|f(\omega)|)\, \mu(d\omega) \tag{3.22.2}$$

$$f_n \to f \text{ in } \varphi\text{-mean} \quad \Longleftrightarrow \quad \lim_{n\to\infty} |f_n - f|_\varphi = 0 \tag{3.22.3}$$

Note that by Markov's inequality (3.9), we have

$$\mu(|f| > \varepsilon) \le \frac{|f|_\varphi}{\varphi(\varepsilon)} \qquad \forall \varepsilon > 0 \ \forall \varphi \in \mathcal{L}(0,\infty) \tag{3.22.4}$$

Hence, if $\varphi \in \mathcal{L}(0,\infty)$, then we have

(3.22.5) If $f_n \to f$ in φ-mean, then $f_n \to f$ in μ-measure

and by Minkowski's inequality (3.12), we get

(3.22.6) $|f+g|_\varphi \le |f|_\varphi + |g|_\varphi$ if $\varphi \in \mathcal{L}(0,\infty)$ is concave

Suppose that $\varphi \in \mathcal{L}(0,\infty)$ is bounded. Then there exists $0 \le a < \infty$ such that $\varphi(x) \le a$ for all $x \ge 0$. Since $\varphi(|f(\omega)|)$ is 0 on $\{f=0\}$, less than $\varphi(\varepsilon)$ on $\{0 < |f| \le \varepsilon\}$, and less than a on $\{|f| > \varepsilon\}$, we have

(3.22.7) $|f|_\varphi \le \varphi(\varepsilon)\,\mu(0 < |f| \le \varepsilon) + a\,\mu(|f| > \varepsilon) \qquad \forall \varepsilon > 0$

Hence, we have the following converse to (3.22.5):

(3.22.8) If $\varphi \in \mathcal{L}(0,\infty)$ is bounded and $\sup_n \mu(0 < |f_n - f| \le b) < \infty$ for some $b > 0$, then $f_n \to f$ in μ-measure if and only if $f_n \to f$ in φ-mean

(3.22.9) If $\varphi \in \mathcal{L}(0,\infty)$ is bounded and $\mu(\Omega) < \infty$, then $f_n \to f$ in φ-mean if and only if $f_n \to f$ in μ-measure.

We shall frequently use (3.22.9) with $\varphi(t) = \arctan(t)$ which, evidently, is a bounded function belonging to $\mathcal{L}(0,\infty)$. We shall also frequently consider convergence in φ-mean, in case $\varphi(t) = t^q$ for some $0 < q < \infty$, and if so we speak about convergence in *q-mean*, that is,

(3.22.10) $f_n \to f$ in q-mean $\iff \lim_{n\to\infty} \int_\Omega |f_n(\omega) - f(\omega)|^q \,\mu(d\omega) = 0$

We say that $\{f_n\}$ *converges in* 0*-mean* (respectively *in* ∞*-mean*) to f if $\{f_n\}$ converges to f in μ-measure (respectively μ-a.s. uniformly to f). We shall now see that convergence in q-mean ($0 \le q \le \infty$) is induced by certain pseudo-metrics on the set of all real-valued measurable functions. Let f be a measurable function from Ω into $\overline{\mathbf{R}}$; then we define

(3.22.11) $||f||_0 := \inf\{a \in \overline{\mathbf{R}}_+ \mid \mu(|f| > a) \le a\}$

(3.22.12) $||f||_q := \int_\Omega |f(\omega)|^q \,\mu(d\omega)$ if $0 < q \le 1$

(3.22.13) $||f||_q := \left\{ \int_\Omega |f(\omega)|^q \,\mu(d\omega) \right\}^{1/q}$ if $1 \le q < \infty$

(3.22.14) $\quad ||f||_\infty := \inf\{a \in \overline{\mathbf{R}}_+ \mid \mu(|f| > a) = 0\}$

Note that $||f||_\infty$ is the smallest number $c \in \overline{\mathbf{R}}_+$ satisfying $|f(\omega)| \le c$ μ-a.s. Using Minkowski's inequality (3.12) if $0 < q < \infty$, and the inequality

$$\mu(|f+g| > a+b) \le \mu(|f| > a) + \mu(|g| > b) \qquad \forall a, b \in \mathbf{R}_+$$

if $q = 0$ or ∞, we find

(3.22.15) $\quad ||f+g||_q \le ||f||_q + ||g||_q \qquad \forall 0 \le q \le \infty$

Thus, if we let

(3.22.16) $\quad M(\mu) = \{ f : \Omega \to \mathbf{R} \mid f \text{ is measurable } \}$

then by (3.22.15) we have

(3.22.17) $\quad ||f-g||_q$ is a pseudo-metric on $M(\mu) \quad \forall 0 \le q \le \infty$

Observe that

(3.22.18) $\quad \mu(|f| > \varepsilon) \le ||f||_0 \qquad$ if $\quad ||f||_0 \le \varepsilon$

(3.22.19) $\quad ||f||_0 \le \varepsilon + \mu(|f| > \varepsilon) \qquad \forall \varepsilon \ge 0$

Hence, we see that convergence in μ-measure is the same as convergence in the pseudo-metric $||f-g||_0$. Similarly if $0 < q \le \infty$, then convergence in q-mean is the same as convergence in the pseudo-metric $||f-g||_q$; that is, for any $0 \le q \le \infty$, we have

(3.22.20) $\quad f_n \to f_0$ in q-mean $\iff \lim_{n\to\infty} ||f_n - f_0||_q = 0$

Using (3.3.8) and (3.22.18), we find

(3.22.21) $\quad ||f||_q = 0$ if and only if $f = 0$ μ-a.s.

(3.22.22) $\quad ||af||_0 \le t\,|a|^{1/2} + \mu(|f| > t|a|^{-1/2}) \qquad \forall t \ge 0 \quad \forall a \in \mathbf{R}$

(3.22.23) $\quad ||af||_q = |a|^q\, ||f||_q \qquad \forall a \in \mathbf{R} \quad \forall 0 < q \le 1$

(3.22.24) $\quad ||af||_q = |a|\ ||f||_q \qquad \forall a \in \mathbf{R} \quad \forall 1 \le q \le \infty$

Let me introduce the so-called *L^q-spaces*:

(3.22.25) $\quad L^q(\mu) = \{f \in M(\mu) \mid \quad ||f||_q < \infty\} \qquad \forall 0 \le q \le \infty$

Since $\{|f| > a\} \downarrow \emptyset$ as $a \to \infty$ whenever $f \in M(\mu)$, then by (1.4.3) we have

(3.22.26) $L^0(\mu) = \{f \in M(\mu) \mid \lim_{t\to\infty} \mu(|f| > t) = 0\}$

(3.22.27) $L^0(\mu) = \{f \in M(\mu) \mid \exists\, t \in \mathbf{R}_+ \text{ so that } \mu(|f| > t) < \infty)\}$

(3.22.28) $L^0(\mu) = M(\mu)$ if μ is a finite measure.

(3.22.29) $L^q(\mu) = \left\{f \in M(\mu) \,\middle|\, \int_\Omega |f|^q \, d\mu < \infty\right\} \qquad \forall\, 0 < q < \infty$

(3.22.30) $L^\infty(\mu) = \{f \in M(\mu) \mid f \text{ is bounded a.s. }\}$

If $0 \le q \le \infty$, then by (3.22.15)–(3.22.24) we have

(3.22.31) $(L^q(\mu), \|\cdot\|_q)$ is a pseudo–metric vector space such that the addition: $(f, g) \curvearrowright f + g$ and scalar multiplication: $(a, f) \curvearrowright af$ are continuous.

If $\{f_n\}$ converges to g and also to h in q-mean, then we have:

$$0 \le \|g - h\|_q \le \lim_{n\to\infty} \|g - f_n\|_q + \lim_{n\to\infty} \|f_n - h\|_q = 0$$

and so by (3.22.21) we get

(3.22.32) If $\{f_n\}$ converges in q-mean to g and also to h, then $g = h$ μ-a.s.

The convergence theorems (3.4)–(3.7) represent almost all that can be said about interchanging limits and integrals. However, a slight, but important improvement is possible. Note that (3.6.2) says that $f_n \to f$ in 1-mean if $f_n \to f$ μ-a.s. and the $|f_n|$'s are dominated by an integrable function h. The condition of domination is sufficient but not necessary for convergence in 1-mean, as will be seen in (3.33).

Let $\mathcal{V}$ be a subset of $L^1(\mu)$; then we define the set function $\nu : \mathcal{F} \to \overline{\mathbf{R}}_+$ as follows:

(3.22.33) $\nu(F) = \sup\left\{\int_F |f| \, d\mu \,\middle|\, f \in \mathcal{V}\right\} \qquad \forall F \in \mathcal{F}$

and we say that $\mathcal{V}$ is *uniformly integrable* if

(3.22.34) $\nu(\Omega) < \infty$ and $\nu(F_n) \to 0 \quad \forall F_n \subseteq \mathcal{F}$ with $F_n \downarrow \emptyset$

Moreover, we say that $\mathcal{V}$ is *finitely μ-concentrated* if

(3.22.35) $\nu(\Omega) < \infty$ and $\forall\, \varepsilon > 0\ \exists\, A \in \mathcal{F} : \nu(A^c) < \varepsilon$ and $\mu(A) < \infty$

In (3.24) we shall see that any μ-uniformly integrable set of functions is finitely μ-concentrated. Note that $\nu(N) = 0$ for every μ-null set N, and by Beppo Levi's theorem (3.7), we see that ν is σ-subadditive. Moreover, if $\mu(\Omega) < \infty$, then setting $A = \Omega$ in (3.22.35) we get

(3.22.36) The set function ν defined by (3.22.33) is σ-subadditive and μ-saturated. And if μ is finite, then $\mathcal{V}$ is finitely μ-concentrated if and only if $\nu(\Omega) < \infty$

We say that $\mathcal{V}$ is L^1-*dominated* if there exists $h \in L^1(\mu)$ such that $|f| \leq h$ μ-a.s. for all $f \in \mathcal{V}$. Suppose that $\mathcal{V}$ is L^1-dominated and let $h \in L^1(\mu)$ be chosen such that $|f| \leq h$ μ-a.s. for all $f \in \mathcal{V}$. Then $\nu(F) \leq \int_F h\, d\mu$ and if $A_n = \{h \leq n\}$, then by (3.13.1) and Markov's inequality (3.9.1) we have $\mu(A_n) < \infty$ and $\nu(A_n^c) \to 0$. Thus, by (3.13.1) we have the following sufficient criteria for uniform integrability of $\mathcal{V}$:

(3.22.37) If $\mathcal{V}$ is L^1-dominated, then $\mathcal{V}$ is uniformly integrable and finitely μ-concentrated. In particular, every finite subset of $L^1(\mu)$ uniformly integrable and finitely μ-concentrated

We say that $\mathcal{V}$ is *uniformly bounded* if there exists $0 \leq a < \infty$ such that $|f| \leq a$ μ-a.s. for all $f \in \mathcal{V}$. Suppose that $\mathcal{V}$ is uniformly bounded; then we have

$$\int_F |f| \; d\mu \leq a\, \mu(F \cap A) + \nu(A^c)$$

for all $f \in \mathcal{V}$ and all $F, A \in \mathcal{F}$. Let $\varepsilon > 0$ be given; then choosing A according to (3.22.35) we find

(3.22.38) If $\mathcal{V}$ is uniformly bounded and finitely μ-concentrated, then $\mathcal{V}$ is uniformly integrable

Moreover, it is easily verified that we have

(3.22.39) Let $0 < c < \infty$ be a given positive number. If $\mathcal{V} \subseteq L^1(\mu)$ and $\mathcal{W} \subseteq L^1(\mu)$ are uniformly μ-integrable, then so is the set $\mathcal{U}_c$ of all measurable functions $f : \Omega \to \overline{\mathbf{R}}$ satisfying

$$\exists\, g \in \mathcal{V}\,,\ \exists\, h \in \mathcal{W} \ :\ |f(\omega)| \leq c\big(|g(\omega)| + |h(\omega)|\big) \ \ \mu\text{-a.s.}$$

With this new notion we can prove a slight, but very important improvement of Fatou's lemma (3.5) and Lebesgue's dominated convergence theorem (3.6).

3.23 Dunford-Pettis' Theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f_1, f_2, \ldots$ be sequence of measurable functions from Ω into $\overline{\mathbf{R}}$. Set

$$f^*(\omega) = \limsup_{n\to\infty} f_n(\omega), \qquad f_*(\omega) = \liminf_{n\to\infty} f_n(\omega) \qquad \forall \omega \in \Omega$$

and let g and h be measurable functions such that $g \geq f^*$ and $h \leq f_*$. Then we have the following:

(1): If $\{f_n^+ \mid n \geq 1\}$ is uniformly integrable, then $(f_*)^+ \in L^1(\mu)$ and we have

$$\limsup_{n\to\infty} \int_F f_n \, d\mu \leq \int_F g \, d\mu \quad \text{if } F \in \mathcal{F} \text{ and } g \cdot 1_F \in L(\mu) \tag{3.23.1}$$

(2): If $\{f_n^- \mid n \geq 1\}$ is uniformly integrable, then $(f^*)^- \in L^1(\mu)$ and we have

$$\liminf_{n\to\infty} \int_F f_n \, d\mu \geq \int_F h \, d\mu \quad \text{if } F \in \mathcal{F} \text{ and } h \cdot 1_F \in L(\mu) \tag{3.23.2}$$

(3): If $\{f_n \mid n \geq 1\}$ is uniformly integrable, then $(f_*)^+ \in L^1(\mu)$ and $(f^*)^- \in L^1(\mu)$; moreover, for every $F \in \mathcal{F}$ we have

$$\int_F f_* \, d\mu \leq \liminf_{n\to\infty} \int_F f_n \, d\mu \leq \limsup_{n\to\infty} \int_F f_n \, d\mu \leq \int_F f^* \, d\mu \tag{3.23.3}$$

(3.23.4) *If $f_n \to f$ μ-a.s., then $f_n \to f$ in 1-mean and* $\displaystyle\int_F f_n \, d\mu \to \int_F f \, d\mu$

Remark: If $f_n \leq \varphi$ μ-a.s. for all $n \geq 1$, for some $\varphi \in L^1(\mu)$, then $0 \leq f_n^+ \leq \varphi^+$ and $(f^*)^+ \leq \varphi^+$. Hence, $\{f_n^+\}$ is uniformly integrable by (3.22.37) and $f^* \in L(\mu)$. Similarly, if $f_n \geq \varphi$ μ-a.s. for all $n \geq 1$ for some $\varphi \in L^1(\mu)$, then $\{f_n^-\}$ is uniformly integrable and $f_* \in L(\mu)$. Hence, setting $g = f^*$ and $h = f_*$, we see that (1)+(2) is an extension of Fatou's lemma (3.5). If $|f_n| \leq \varphi$ μ-a.s. for all $n \geq 1$, for some $\varphi \in L^1(\mu)$, then as above we have that $\{f_n\}$ is uniformly integrable. Hence, (3) is an extension of Lebesgue's dominated convergence theorem (3.6).

Proof: (1): Since $(f_*)^+ = \liminf f_n^+$ and $\{f_n^+\}$ is uniformly integrable, then by Fatou's lemma we have

$$\int_\Omega (f_*)^+ \, d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n^+ \, d\mu \leq \nu(\Omega) < \infty$$

where ν is defined by

$$\nu(F) = \sup_n \left\{ \int_F f_n^+ \, d\mu \right\} \qquad \forall F \in \mathcal{F}$$

Hence, $(f_*)^+ \in L^1(\mu)$. Let $\varepsilon > 0$ be given, and set $g_n = (f_n^+ - g^+)^+$ with the convention $\infty - \infty = 0$, then $0 \le g_n \le f_n^+ \le g_n + g^+$. Since $g^+ \ge \limsup f_n^+$, we have $g_n(\omega) \to 0$ for all $\omega \in \Omega$. Let

$$A_j = \bigcup_{n=j}^{\infty} \{g_n \ge 1\} \qquad B_j = \bigcap_{n=1}^{j} \{g_n \le 2^{-j}\} \qquad B = \bigcup_{n=1}^{\infty} \{g_n > 0\}$$

Then $B_j \cap B \downarrow \emptyset$ and since $g_n \to 0$, we have that $A_j \downarrow \emptyset$. Hence, by (3.22.34) there exists an integer $p \ge 1$ such that $\nu(B \cap B_p) < \varepsilon$ and $\nu(A_p) \le \varepsilon$. Let $C = (A_p \cup B_p)^c = A_p^c \cap B_p^c$. Then

$$C \subseteq B_p^c = \bigcup_{k=1}^{p} \{g_k > 2^{-p}\}$$

Therefore, by Markov's inequality (3.9) we have

$$\mu(C) \le \sum_{k=1}^{p} \mu(g_k > 2^{-p}) \le \sum_{k=1}^{p} 2^p \int_\Omega g_k \, d\mu \le p\, 2^p \, \nu(\Omega) < \infty$$

because $0 \le g_k \le f_k^+$. Moreover, $C \subseteq A_p^c$ and so $g_n(\omega) \le 1$ for all $n \ge p$ and all $\omega \in C$; hence, $0 \le 1_C \, g_n \le 1_C$ for all $n \ge p$ and since $1_C \in L^1(\mu)$ and $1_C \, g_n \to 0$, we have by Lebesgue's dominated convergence theorem that there exists an integer $k \ge p$ such that $\int_C g_n \, d\mu \le \varepsilon$ for all $n \ge k$. Let $n \ge k$. Since $\Omega = C \cup A_p \cup B_p$, $0 \le g_n \le f_n^+$ and $g_n(\omega) = 0$ for all $\omega \in B_p \setminus B$, we have

$$\begin{aligned}
\int_\Omega g_n \, d\mu &\le \int_C g_n \, d\mu + \int_{A_p} g_n \, d\mu + \int_{B_p} g_n \, d\mu \\
&= \int_C g_n \, d\mu + \int_{A_p} g_n \, d\mu + \int_{B_p \cap B} g_n \, d\mu \\
&\le \varepsilon + \nu(A_p) + \nu(B_p \cap B) \\
&\le 3\varepsilon
\end{aligned}$$

Let $F \in \mathcal{F}$. Since $f_n^+ \le g_n + g^+$, we have

$$\limsup_{n\to\infty} \int_F f_n^+ \, d\mu \le \limsup_{n\to\infty} \int_F (g_n + g^+) \, d\mu \le 3\varepsilon + \int_F g^+ \, d\mu$$

and since $g^- \le \liminf f_n^-$, then by Fatou's lemma (3.5) we have

$$\liminf_{n\to\infty} \int_F f_n^- \, d\mu \ge \int_F g^- \, d\mu$$

Suppose that $1_F g \in L(\mu)$. If $\int_F g \, d\mu = \infty$, then (3.23.1) holds trivially. So suppose that $\int_F g \, d\mu < \infty$; then we have

$$0 \le \limsup_{n\to\infty} \int_F f_n^+ \, d\mu \le 3\varepsilon + \int_F g^+ \, d\mu < \infty$$

Therefore, by (P.4.19) we find

$$\begin{aligned}
\limsup_{n\to\infty} \int_F f_n \, d\mu &= \limsup_{n\to\infty} \Big\{ \int_F f_n^+ \, d\mu - \int_F f_n^- \, d\mu \Big\} \\
&\le \limsup_{n\to\infty} \int_F f_n^+ \, d\mu - \liminf_{n\to\infty} \int_F f_n^- \, d\mu \\
&\le 3\varepsilon + \int_F g^+ \, d\mu - \int_F g^- \, d\mu \\
&\le 3\varepsilon + \int_F g \, d\mu
\end{aligned}$$

Letting $\varepsilon \downarrow 0$, we see that (3.23.1) holds.

(2): Apply (1) to the sequence $(-f_n)$.

(3): In this case $\{f_n^+\}$ and $\{f_n^-\}$ are both uniformly integrable; thus, setting $g = f^*$ and $h = f_*$, we have that h^+ and g^- are μ-integrable by (1) and (2) and so $h, g \in L(\mu)$. But then $1_F h$ and $1_F g$ both belong to $L(\mu)$ for all $F \in \mathcal{F}$, and so (3.23.3) follows from (1) and (2). Suppose that $f_n \to f$ a.s. then $f = g = h$ a.s. and so $f \in L^1(\mu)$. Moreover, $|f_n - f| \to 0$ a.s. and since $|f_n - f| \le |f_n| + |f|$, then by (3.22.37) and (3.22.39) we see that $\{\, |f_n - f| \mid n \ge 1\}$ is uniformly integrable. Hence, (3.23.4) follows from (3.23.3) applied to the sequence $\{\, |f_n - f|\}$. □

3.24 Criteria for Uniform Integrability

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\mathcal{V} \subseteq L^1(\mu)$ be a given set of integrable functions. If $\nu(F)$ is defined by (3.22.33), then the following five statements are equivalent:

(3.24.1) *$\mathcal{V}$ is uniformly μ-integrable*

(3.24.2) *ν is μ-continuous and $\mathcal{V}$ is finitely μ-concentrated*

(3.24.3) *$\mathcal{V}$ is finitely μ-concentrated, and for all $\varepsilon > 0$ there exists a μ-continuous set function $\rho : \mathcal{F} \to \overline{\mathbf{R}}_+$ and a finite set $\mathcal{S} \subseteq \mathcal{V}$ such that*

$$\int_F |f(\omega)|\,\mu(d\omega) < \varepsilon + \rho(F) \qquad \forall f \in \mathcal{V} \setminus \mathcal{S}\ \forall F \in \sigma(f)$$

where $\sigma(f)$, as usual, denotes the smallest σ-algebra making f measurable

(3.24.4) *$\mathcal{V}$ is finitely μ-concentrated, and for all $\varepsilon > 0$ there exists $a \in \mathbf{R}_+$ such that*

$$\int_{\{|f|>a\}} |f(\omega)|\,\mu(d\omega) < \varepsilon \qquad \forall f \in \mathcal{V}$$

(3.24.5) *$\mathcal{V}$ is finitely μ-concentrated, and there exists a number $c \in \mathbf{R}_+$ and a Borel function $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ such that*

$$\lim_{t\to\infty} \frac{\varphi(t)}{t} = \infty \text{ and } \int_\Omega \varphi(|f(\omega)|)\,\mu(d\omega) \le c \quad \forall f \in \mathcal{V}$$

Remarks: (1): Condition (3.24.5) is particularly useful for verifying uniform integrability of $\mathcal{V}$. In applications, one often uses $\varphi(t) = t^2$ or $\varphi(t) = t^q$, where $q > 1$, or $\varphi(t) = t\log(1+t)$, which all satisfy: $\lim_{t\to\infty} \frac{\varphi(t)}{t} = \infty$.

(2): Suppose that $\mu(\Omega) < \infty$. Then by (3.22.36) we have that $\mathcal{V}$ is finitely μ-concentrated if and only if $\nu(\Omega) < \infty$, that is, if and only if $\sup_{f\in\mathcal{V}} \int_\Omega |f|\,d\mu < \infty$.

Proof: (3.24.1) $\Rightarrow$ (3.24.2): By (3.22.36), we have that ν is σ-subadditive and μ-saturated. Hence, by (3.16.6) we see that (3.24.1) implies that ν is μ-continuous. So let us show that uniform integrability implies finite concentration. So suppose that $\mathcal{V}$ is uniformly integrable, then $\nu(\Omega) < \infty$, and we can find $f_1 \in \mathcal{V}$ such that

$$\nu(\Omega) \le 2^{-1} + \int_\Omega |f_1|\,d\mu$$

and since $\{0 < |f_1| < 2^{-j}\} \downarrow \emptyset$ as $j \uparrow \infty$, there exists $a_1 > 0$ such that $\nu(0 < |f_1| < a_1) < 2^{-1}$. Let $F_1 = \{|f_1| \ge a_1\}$ and choose $f_2 \in \mathcal{V}$ such that

$$\nu(\Omega \setminus F_1) \le 2^{-2} + \int_{\Omega\setminus F_1} |f_2|\,d\mu$$

Continuing in this way we can find $f_1, f_2 \ldots \in \mathcal{V}$ and positive numbers $a_1, a_2 \ldots$ satisfying

$$\nu(\Omega \setminus F_n) \le 2^{-n-1} + \int_{\Omega\setminus F_n} |f_{n+1}|\,d\mu \quad \text{and} \quad \nu(H_n) \le 2^{-n-1}$$

for all $n \geq 0$ where $F_0 = \emptyset$ and

$$F_n = \bigcup_{j=1}^{n} \{ |f_j| \geq a_j \} \ , \ H_{n-1} = \{0 < f_n < a_n\} \ , \ G_{n-1} = \{f_n = 0\}$$

for all $n \geq 1$. If $F = \cup_1^\infty F_n$, then $F \setminus F_n \downarrow \emptyset$, and since $\Omega \setminus F_n \subseteq H_n \cup (F \setminus F_n) \cup G_n$, we have

$$\begin{aligned}
\nu(\Omega \setminus F_n) &\leq 2^{-n-1} + \int_{\Omega \setminus F_n} |f_{n+1}| \, d\mu \\
&\leq 2^{-n-1} + \int_{F \setminus F_n} |f_{n+1}| \, d\mu + \int_{H_n} |f_{n+1}| \, d\mu + \int_{G_n} |f_{n+1}| \, d\mu \\
&\leq 2^{-n-1} + \nu(F \setminus F_n) + \nu(H_n) \leq 2^{-n} + \nu(F \setminus F_n)
\end{aligned}$$

Hence, $\nu(\Omega \setminus F_n) \to 0$ by uniform integrability. Moreover, by Markov's inequality (3.9) we have

$$\mu(F_n) \leq \sum_{j=1}^{n} \mu(|f| \geq a_j) \leq \sum_{j=1}^{n} \frac{1}{a_j} \int_\Omega |f_j| \, d\mu < \infty$$

Thus, we see that $\mathcal{V}$ is finitely μ-concentrated.

(3.24.2) $\Rightarrow$ (3.24.3): Evident! (Set $\rho = \nu$).

(3.24.3) $\Rightarrow$ (3.24.4): Let $\varepsilon > 0$ be given and choose ρ and $\mathcal{S}$ according to (3.24.3). Then by μ-continuity of ρ there exists $\delta > 0$ such that $\rho(F) \leq \varepsilon$ whenever $F \in \mathcal{F}$ and $\mu(F) \leq \delta$. Let $c = \nu(\Omega)/\delta$; then by Markov's inequality (3.9) we have

$$\mu(|f| > c) \leq \tfrac{1}{c} \int_\Omega |f| \, d\mu \leq \tfrac{1}{c} \nu(\Omega) \leq \delta$$

for all $f \in \mathcal{V}$ and since $\{|f| > c\} \in \sigma(f)$, we get

$$\int_{\{|f|>c\}} |f| \, d\mu \leq \rho(\{|f| > c\}) + \varepsilon \leq 2\varepsilon \qquad \forall f \in \mathcal{V} \setminus \mathcal{S}$$

Moreover, since $\mathcal{S}$ is a finite subset of $L^1(\mu)$, there exists $a \geq c$ such that the inequality above holds for $f \in \mathcal{S}$ with c replaced by a . Hence, (3.24.4) follows.

(3.24.4) $\Rightarrow$ (3.24.5): Let $a = \nu(\Omega)$, then by (3.24.4) there exist integers $0 = a_0 < a_1 < a_2 < \ldots$ such that $a_k \to \infty$ and

$$\int_{\{|f| \geq a_k\}} |f| \, d\mu \leq 4^{-k} a \qquad \forall f \in \mathcal{V} \qquad \forall k = 0, 1, 2, \ldots$$

Let $\varphi(t) = 2^k t$ if $t \in [a_k, a_{k+1}[$ and $k \geq 0$. Then φ is an increasing function from $\mathbf{R}_+$ into $\mathbf{R}_+$ such that $\frac{\varphi(t)}{t} \to \infty$ as $t \to \infty$. Let $f \in \mathcal{V}$ and set $F_k = \{a_k \leq |f| < a_{k+1}\}$. Then we have

$$\int_\Omega \varphi(|f(\omega)|)\mu(d\omega) = \sum_{k=0}^{\infty} 2^k \int_{F_k} |f|\, d\mu$$
$$\leq \sum_{k=0}^{\infty} 2^k \int_{\{|f| \geq a_k\}} |f|\, d\mu \leq \sum_{k=0}^{\infty} 2^k 4^{-k} a = 2a$$

Thus, we see that (3.24.5) follows from (3.24.4).

(3.24.5) $\Rightarrow$ (3.24.1): Let $F_n \in \mathcal{F}$ such that $F_n \downarrow \emptyset$, and let $\varepsilon > 0$ be given. Since $\mathcal{V}$ is finitely μ-concentrated by assumption, we have $\nu(\Omega) < \infty$ and there exists $F \in \mathcal{F}$ such that $\mu(F) < \infty$ and $\nu(\Omega \setminus F) \leq \varepsilon$. Let φ and c be chosen according to assumption (3.24.5). Then we choose $t_0 > 0$ such that $\varphi(t) \geq \frac{c}{\varepsilon} t$ for all $t \geq t_0$. Let $f \in \mathcal{V}$ and $A \in \mathcal{F}$ be given; if $L = \{\, |f| < t_0\}$ and $M = \{\, |f| \geq t_0\}$, then $|f| \leq t_0$ on L and $|f| \leq \frac{\varepsilon}{c}\varphi(|f|)$ on M . Moreover, since $A \subseteq F^c \cup (A \cap F \cap L) \cup M$ we find

$$\int_A |f|\, d\mu \leq \int_{F^c} |f|\, d\mu + \int_{A \cap F \cap L} |f|\, d\mu + \int_M |f|\, d\mu$$
$$\leq \nu(F^c) + t_0\, \mu(A \cap F) + \frac{\varepsilon}{c} \int_\Omega \varphi(|f|)\, d\mu$$

Setting $A = F_n$ and taking supremum over $f \in \mathcal{V}$, we find

$$\nu(F_n) \leq \varepsilon + t_0\, \mu(F \cap F_n) + \varepsilon = 2\varepsilon + t_0\, \mu(F \cap F_n)$$

and since $\mu(F) < \infty$, we have $\mu(F \cap F_n) \to 0$ by (1.4.3). Hence there exists $k \in \mathbf{N}$ such that $t_0\, \mu(F \cap F_n) \leq \varepsilon$ for $n \geq k$. Thus, $\nu(F_n) \leq 3\varepsilon$ for all $n \geq k$ and so $\nu(F_n) \to 0$. That is, $\mathcal{V}$ is uniformly integrable and the equivalence of (3.24.1)–(3.24.5) is established. □

3.25 Convergence Relations

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $\psi, \varphi \in \mathcal{L}(0, \infty)$ be given functions such that $\psi(2x) \leq C\,\psi(x)$ for all $x \geq 0$ for some finite constant C . If f and $f_1, f_2, \ldots$ are measurable functions from Ω into $\mathbf{R}$, then we have the following:

(3.25.1) *If $f_n \to f$ μ-a.s. and $\{\varphi(|f_n - f|) \mid n \geq 1\}$ [respectively $\{\psi(|f_n|) \mid n \geq 1\}$] is uniformly integrable, then $f_n \to f$ in φ-mean [respectively in ψ-mean]*

(3.25.2) *If there exists* $\varepsilon_n > 0$ *such that* $\sum_{n=1}^{\infty} \mu(|f_n - f| > \varepsilon_n) < \infty$ *and* $\varepsilon_n \to 0$ *, then* $f_n \to f$ μ*-a.s.*

(3.25.3) *If* $\sum_{n=1}^{\infty} |f_n - f|_\varphi < \infty$ *, then* $f_n \to f$ μ*-a.s.*

(3.25.4) *If* $f_n \to f$ *in* μ*-measure, then there exists integers* $\sigma(1) < \sigma(2) < \sigma(3) < \cdots$ *such that* $f_{\sigma(n)} \to f$ μ*-a.s.*

(3.25.5) *If* $f_n \to f$ *in* φ*-mean, then* $f_n \to f$ *in* μ*-measure*

(3.25.6) *If* $f_n \to f$ *in* μ*-measure and* $\{\varphi(|f_n - f|) \mid n \geq 1\}$ *[respectively* $\{\psi(|f_n|) \mid n \geq 1\}$ *] is uniformly integrable, then* $f_n \to f$ *in* φ*-mean [respectively in* ψ*-mean]*

(3.25.7) *If* $f_n \to f$ μ*-a.s. and* $\mu(\Omega) < \infty$ *, then* $f_n \to f$ *in* μ*-measure*

(3.25.8) *If* φ *is bounded and* $\mu(\Omega) < \infty$ *, then* $f_n \to f$ *in* μ*-measure if and only if* $f_n \to f$ *in* φ*-mean*

(3.25.9) *If* $|f_n - f_m|_\varphi \to 0$ *as* $n, m \to \infty$ *, then there exists a measurable function* h *such that* $f_n \to h$ *in* φ*-mean*

(3.25.10) $(L^q(\mu), \|\cdot\|_q)$ *is a complete pseudo–metric space* $\forall\, 0 \leq q \leq \infty$

Remark: The ten propositions above represent virtually all relations between the convergence notions introduced in (3.22), and no other relation holds in general. Let $0 < q < \infty$ be a given number, let $f, f_1, f_2, \ldots$ be given measurable functions from Ω into $\mathbf{R}$, and let (U^q) denote the following condition:

(U^q) : $\{\, |f_n|^q \mid n \geq 1 \,\}$ is uniformly μ-integrable

Note that $|2x|^q = 2^q\, |x|^q$. Hence, we have the following scheme of implications for the convergence notions introduced in Section (3.22):

$$\begin{array}{ccccccc} \mu\text{-measure} + (U^q) & \Rightarrow & q\text{-mean} & \Rightarrow & \mu\text{-measure} & \Leftrightarrow & 0\text{-mean} \\ & & \Uparrow & & & & \\ & & \mu\text{-a.s.} + (U^q) & & & & \end{array} \tag{3.25.11}$$

Suppose that $\mu(\Omega) < \infty$ and that $0 < r < q < \infty$. Setting $p = \frac{rq}{q-r}$, then $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ and so by Hölder's inequality (3.11), we have

$$\left\{\int_\Omega |f|^r\, d\mu\right\}^{1/r} = \left\{\int_\Omega |f \cdot 1_\Omega|^r\, d\mu\right\}^{1/r} \leq \left\{\int_\Omega |f|^q\, d\mu\right\}^{1/q} \left\{\int_\Omega 1_\Omega\, d\mu\right\}^{1/p}$$

$$\leq \mu(\Omega)^{1/p} \cdot \left\{\int_\Omega |f|^q\, d\mu\right\}^{1/q}$$

Hence, if $\mu(\Omega) < \infty$ and $0 \le r \le \infty$ and $0 < q < \infty$ are given numbers, then we have the following scheme of implications for the convergence notions introduced in Section (3.22):

$$\begin{array}{ccccccc} & \mu\text{-a.s.} & \Rightarrow & \mu\text{-measure} & \Leftrightarrow & \text{arctan-mean} & \Leftrightarrow & \text{0-mean} \\ & \Uparrow & & \Uparrow & & & \\ (3.25.12) & \mu\text{-a.s} + (U^q) & \Rightarrow & q\text{-mean} & \Rightarrow & r\text{-mean} & \forall\, 0 \le r \le q \\ & & & \Uparrow & & & \\ & & & \mu\text{-measure} + (U^q) & & & \end{array}$$

where convergence in arctan-mean is convergence in φ-mean for the particular function $\varphi(x) := \arctan(x)$. In particular, if $\mu(\Omega) < \infty$, then convergence in μ-measure is the weakest of the convergence notions from Section (3.22).

Proof: (3.25.1): Suppose that $f_n \to f$ μ-a.s. and that $\{\varphi(|f_n - f|) \mid n \ge 1\}$ is uniformly integrable. Then $\varphi(|f_n - f|) \to 0$ μ-a.s. and so by Dunford-Pettis' theorem (3.23), we have that $f_n \to f$ in φ-mean. Suppose that $f_n \to f$ μ-a.s. and that $\{\psi(|f_n|) \mid n \ge 1\}$ is uniformly integrable. Since ψ is increasing and left continuous, we have $\psi(a) \le \liminf_{n\to\infty} \psi(a_n)$ whenever $a_n \to a$. Hence, by Fatou's lemma (1.5.1) and uniform integrability, we find

$$\begin{aligned} \int_\Omega \psi(|f(\omega)|)\,\mu(d\omega) &\le \int_\Omega \liminf_{n\to\infty} \psi(|f_n(\omega)|)\,\mu(d\omega) \\ &\le \liminf_{n\to\infty} \int_\Omega \psi(|f_n(\omega)|)\,\mu(d\omega) \\ &\le \sup_{n\ge 1} \int_\Omega \psi(|f_n(\omega)|)\,\mu(d\omega) < \infty \end{aligned}$$

Thus, $\psi(|f|) \in L^1(\mu)$. Let $x, y \ge 0$ be given. If $z = x \vee y$, then $\psi(x+y) \le \psi(2z) \le C\,\psi(z) \le C(\psi(x) + \psi(y))$. Hence, $\psi(|f_n - f|) \le C(\psi(|f_n|) + \psi(|f|)$ and so by (3.22.37) and (3.22.39), we conclude that $\{\psi(|f_n - f|) \mid n \ge 1\}$ is uniformly μ-integrable. Hence, by the first part of the proposition, we see that $f_n \to f$ in ψ-mean.

(3.25.2): Let $F_n = \{|f_n - f| > \varepsilon_n\}$, then $\sum \mu(F_n) < \infty$ and so by the first Borel–Cantelli lemma (1.4.6), we have that $\mu(N) = 0$ where $N = \limsup F_n$. If $\omega \in \Omega \setminus N$, then there exists $k \in \mathbf{N}$ such that $\omega \in F_n^c$ for all $n \ge k$ and so $|f_n(\omega) - f(\omega)| \le \varepsilon_n$ for $n \ge k$. Since $\varepsilon_n \to 0$, we see that $f_n(\omega) \to f(\omega)$ for all $\omega \in \Omega \setminus N$.

(3.25.3): By Beppo Levi's theorem (3.7), we have that $\sum \varphi(|f_n - f|) < \infty$ μ-a.s. Hence, $\varphi(|f_n - f|) \to 0$ μ-a.s., and since φ is increasing and $\varphi(x) > 0$ for $x > 0$, we find that $|f_n - f| \to 0$ μ-a.s.

(3.25.4): By definition (3.22.1) there exist integers $\sigma(1) < \sigma(2) < \dots$ such that $\mu(|f_n - f| > 2^{-p}) \le 2^{-p}$ for all $n \ge \sigma(p)$. Hence, by (3.25.2) we have $f_{\sigma(n)} \to f$ μ-a.s.

(3.25.5): See (3.22.5).

(3.25.6): Suppose that $f_n \to f$ in μ-measure and that $\{\varphi(|f_n - f|) \mid n \ge 1\}$ is uniformly integrable. If $\{f_n\}$ does not converge to f in φ-mean, then there exists $\varepsilon_0 > 0$ such that $L = \{n \in \mathbf{N} \mid |f_n - f|_\varphi \ge \varepsilon_0\}$ is an infinite subset of $\mathbf{N}$. Hence, by (3.25.4) there exists integers $\sigma(1) < \sigma(2) < \dots$ such that $\sigma(n) \in L$ for all n and $f_{\sigma(n)} \to f$ μ-a.s. But then $|f_{\sigma(n)} - f|_\varphi \ge \varepsilon_0$ for all $n \ge 1$. However, this contradicts (3.25.1) and so $\{f_n\}$ must converge to f in φ-mean. The other part is proved in a similar manner.

(3.25.7): Let $\varepsilon > 0$ be given. Since $\mu(\Omega) < \infty$, then by (1.4.5) we have

$$\begin{aligned} 0 \le \limsup_{n\to\infty} \mu(|f_n - f| > \varepsilon) &\le \mu(\limsup\{|f_n - f| > \varepsilon\}) \\ &\le \mu\big(\omega \mid \{f_n(\omega)\} \text{ does not converge to } f(\omega)\big) = 0 \end{aligned}$$

and so $f_n \to f$ in μ-measure.

(3.25.8): See (3.22.9).

(3.25.9+10): Suppose that $\{f_n\}$ is a Cauchy sequence with respect to the pseudo-metric $\|\cdot\|_0$. Then there exist integers $\sigma(1) < \sigma(2) < \cdots$ such that $\|f_n - f_m\|_0 \le 2^{-k}$ for all $n, m \ge \sigma(k)$. Hence, by (3.22.18) we have $\mu(|f_n - f_m| > 2^{-k}) \le 2^{-k}$ for all $k \ge 1$ and all $n, m \ge \sigma(k)$. Let us define

$$F_k = \{\, |f_{\sigma(k+1)} - f_{\sigma(k)}| > 2^{-k}\} \qquad F^k = \bigcup_{j=k}^{\infty} F_j \qquad N = \limsup F_k$$

Then $\sum \mu(F_k) < \infty$ and so by the first Borel-Cantelli lemma (1.4.6), we have that $\mu(N) = 0$. Let $\omega \in \Omega \setminus N$ be given. Then there exists $m \in \mathbf{N}$ such that $\omega \notin F^m$, and so we have

$$\begin{aligned} |f_{\sigma(n)}(\omega) - f_{\sigma(k)}(\omega)| &\le \sum_{j=k+1}^{n} |\, f_{\sigma(j)}(\omega) - f_{\sigma(j-1)}(\omega)| \\ &\le \sum_{j=k+1}^{n} 2^{-j+1} \le 2^{-k+1} \end{aligned}$$

for all $n > k \ge m$. Hence, $\{f_{\sigma(n)}(\omega)\}$ is a Cauchy sequence on $\mathbf{R}$ and, thus, $h(\omega) = \lim_{k\to\infty} f_{\sigma(k)}(\omega)$ exists for all $\omega \in \Omega \setminus N$. Setting $h(\omega) = 0$ if $\omega \in N$, we see that h is a measurable function and

$$\lim_{j\to\infty} |f_n(\omega) - f_{\sigma(j)}(\omega)| = |f_n(\omega) - h(\omega)| \qquad \forall \omega \in \Omega \setminus N$$

$$\mu(|h - f_{\sigma(k)}| > 2^{-k+1}) \le \mu(F^k) \le \sum_{j=k}^{\infty} \mu(F_j) \le \sum_{j=k}^{\infty} 2^{-j} = 2^{-k+1}$$

for all $n, k \geq 1$. Thus, $||h - f_{\sigma(k)}||_0 \leq 2^{-k+1}$, and so by the choice of $\sigma(k)$ we get

$$||h - f_n||_0 \leq ||h - f_{\sigma(k)}||_0 + ||f_{\sigma(k)} - f_n||_0 \leq 2^{-k+1} + 2^{-k}$$

for all $n \geq \sigma(k)$. Hence, $f_n \to h$ in μ-measure . In particular, we see that $(L^0(\mu), ||\cdot||_0)$ is a complete pseudo-metric space, and so (3.25.10) is proved for $q = 0$.

Suppose that $\lim_{n,m\to\infty} |f_n - f_m|_\varphi = 0$. Then by (3.22.4) we have that $\{f_n\}$ is a $||\cdot||_0$-Cauchy sequence and, thus, we may construct $\sigma(k)$, F_k and h as above. Since φ is increasing and left continuous, we have that $\varphi(a) \leq \liminf_{n\to\infty} \varphi(a_n)$ whenever $a_n \in \mathbf{R}_+$ and $a_n \to a$. Hence, we see that

$$\varphi(|f_n - h|) \leq \liminf_{j\to\infty} \varphi(|f_n - f_{\sigma(j)}|) \quad \mu\text{-a.s.}$$

Thus, by Fatou's lemma (3.5.1) we have

$$\begin{aligned} |f_n - h|_\varphi &= \int_\Omega \varphi(|f_n - h|)\, d\mu \leq \int_\Omega \liminf_{j\to\infty} \varphi\big(|f_n - f_{\sigma(j)}|\big)\, d\mu \\ &\leq \liminf_{j\to\infty} \int_\Omega \varphi\big(|f_n - f_{\sigma(j)}|\big)\, d\mu = \liminf_{j\to\infty} |f_n - f_{\sigma(j)}|_\varphi \end{aligned}$$

and since $|f_n - f_m|_\varphi \to 0$ as $n, m \to \infty$, we conclude that $f_n \to h$ in φ-mean. Thus, (3.22.9) is proved. Setting $\varphi(t) = t^q$ we have proved (3.22.10) for $0 < q < \infty$, and the last case: $q = \infty$, follows easily from (P.6.5). □

3.26 The Lebesgue Integral on R and $\mathbf{R}^k$

Let λ be the Lebesgue measure on $\mathbf{R}$ and let $I = [a, b]$ be a bounded interval. Then we have two integrals for functions $f : I \to \mathbf{R}$, *the Lebesgue integral* and the classical Riemann integral:

(3.26.1) $$\int_I f(x)\, \lambda(dx) \quad \text{and} \quad \int_a^b f(x)\, dx \quad \text{respectively}$$

I claim that we have:

(3.26.2) If the Riemann integral of $f : I \to \mathbf{R}$ exists, then so does the Lebesgue integral and the two integrals are equal.

That is, I claim that the Lebesgue integral is an extension of the classical Riemann integral.

So suppose that the Riemann integral exists. Then f is bounded and if $a_{nj} = \frac{j}{n}(b-a) + a$ and

$$m_{nj} = \inf\{f(t) \mid a_{nj-1} \le t \le a_{nj}\}$$
$$M_{nj} = \sup\{f(t) \mid a_{nj-1} \le t \le a_{nj}\}$$

for $j = 1, \ldots, n$, then we have

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \frac{b-a}{n} \sum_{j=1}^n m_{nj} = \lim_{n\to\infty} \frac{b-a}{n} \sum_{j=1}^n M_{nj}$$

Let $I_{nj} =]a_{nj-1}, a_{nj}]$ and

$$f_n = \sum_{j=1}^n m_{nj}\, 1_{I_{nj}} \quad , \quad g_n = \sum_{j=1}^n M_{nj}\, 1_{I_{nj}}$$

Then $f_n \le f \le g_n$, and by Fatou's lemma (3.5) and (3.8.2), we have

$$\begin{aligned}
\int_a^b f(x)\,dx &= \lim_{n\to\infty} \frac{b-a}{n} \sum_{j=1}^n m_{nj} = \limsup_{n\to\infty} \int_I f_n\,d\lambda \\
&\le \int_I \left(\limsup_{n\to\infty} f_n\right) d\lambda \le \int_I f\,d\lambda \\
&\le \int_I \left(\liminf_{n\to\infty} g_n\right) d\lambda \le \liminf_{n\to\infty} \int_I g_n\,d\lambda \\
&= \lim_{n\to\infty} \frac{b-a}{n} \sum_{j=1}^n M_{nj} = \int_a^b f(x)\,dx
\end{aligned}$$

Hence, we see that the Riemann integral coincides with the Lebesgue integral whenever the Riemann integral exists.

From now on use the following notation for *the Lebesgue integral* on $\mathbf{R}$:

(3.26.3) $$\int_a^b f(x)\,dx := \int_{]a,b[} f\,d\lambda \qquad \forall\, a, b \in \mathbf{R}$$

(3.26.4) $$\int_A f(x)\,dx := \int_A f\,d\lambda \qquad \forall\, A \in \mathcal{B}(\mathbf{R})$$

whenever $1_{]a,b[} f \in L(\lambda)$ or $1_A f \in L(\lambda)$.

Let λ_k be the k-dimensional Lebesgue measure on $\mathbf{R}^k$. Then in exactly the same way one shows that the Lebesgue integral on $\mathbf{R}^k$ is an extension of the classical k-dimensional multiple Riemann integral. We shall use the following two notations for *the Lebesgue integral* on $\mathbf{R}^k$:

$$\int_A f(x)\,dx := \int_A f\,d\lambda_k \qquad \forall A \in \mathcal{B}(\mathbf{R}^k) \tag{3.26.5}$$

$$\int_A f(x_1,\ldots,x_k)\,dx_1\cdots dx_k := \int_A f\,d\lambda_k \qquad \forall A \in \mathcal{B}(\mathbf{R}^k) \tag{3.26.6}$$

whenever $1_A f \in L(\lambda_k)$. Consider the two–dimensional Lebesgue measure λ_2 . Since

$$\lambda_2(\,]a_1,b_1]\times]a_2,b_2]\,) = (b_1-a_1)(b_2-a_2) = (\lambda\otimes\lambda)(]a_1,b_1]\times]a_2,b_2])$$

it follows from the uniqueness theorem (1.7) that $\lambda_2 = \lambda\otimes\lambda$. Similarly, we find

$$\lambda_k = \lambda\otimes\lambda\otimes\cdots\otimes\lambda \qquad (k \text{ times }) \tag{3.26.7}$$

Hence, by the Tonelli-Fubini theorem (3.21), we may compute a k-dimensional Lebesgue integral by integrating with respect to one coordinate at the time in any order we like, provided that the function is non-negative or provided that either the positive or the negative part of the function has a finite integral.

Let me show that the Lebesgue integral is a proper extension of the Riemann integral. Let $I = [0,1]$ be the unit interval and let f be given by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

Then f is not Riemann-integrable [every upper sum of f equals 1 and every lower sum of f equals 0]. But f is Lebesgue integrable because $f = 1_Q$, where Q is the set of rational numbers in $[0,1]$, and we have

$$\int_0^1 f(x)dx = \lambda(Q) = \sum_{q\in Q} \lambda(\{q\}) = 0$$

since Q is countable. Actually, it can be shown [but this is difficult!] that:

(3.26.8) A function $f : [a,b] \to \mathbf{R}$ is Riemann integrable if and only if f is bounded and continuous λ-a.s.

3.27 The Lebesgue-Stieltjes Integral

Let $F : \mathbf{R} \to \mathbf{R}$ be increasing and right continuous. I have defined the Lebesgue-Stieltjes measure λ_F in (1.32), and I shall use the following two notations for *the Lebesgue–Stieltjes integral* on $\mathbf{R}$:

$$\int_A f(x)\,F(dx) = \int_A f dF := \int_A f\,d\lambda_F \qquad \forall\, A \in \mathcal{B}(\mathbf{R}) \tag{3.27.1}$$

whenever $1_A f \in L(\lambda_F)$. Similarly, if $-\infty \le a \le b \le \infty$, then we define

$$\int_a^b f(x)\,F(dx) = \int_a^b f\,dF := \int_A f\,dF \quad \text{where } A =]a,b] \cap \mathbf{R} \tag{3.27.2}$$

provided, of course, that $1_A f \in L(\lambda_F)$.

Suppose that $\varphi : \mathbf{R} \to \overline{\mathbf{R}}_+$ is measurable and that

$$F(b) - F(a) = \int_a^b \varphi(x)\,dx \qquad \forall\, a \le b \tag{3.27.3}$$

Then by (1.7) we have that λ_F is absolutely λ-continuous and φ is the λ-density of λ_F . Hence, by (3.17.7) we have

$$\int_A f(x)\,F(dx) = \int_A f(x)\varphi(x)\,dx \tag{3.27.4}$$

in the sense that if one of the integrals exists, then so does the other and they are equal.

A function F from an interval I into $\mathbf{R}$ satisfying (3.27.3), whenever $a, b \in I$, for some Borel function $\varphi : I \to \overline{\mathbf{R}}$ is called *absolutely continuous*, and φ is called *the density* of F . It can be shown that we have [see Exc.3.37]

(3.27.5) If $F : I \to \mathbf{R}$ is absolute continuous with density φ , then there exists a Lebesgue null set $N \subseteq I$ such that F is differentiable at x with $F'(x) = \varphi(x)$ for all $x \in I \setminus N$. Conversely, if $F : I \to \mathbf{R}$ is differentiable at all $x \in I \setminus T$ where T is countable, and if either $F'(x)^+$ or $F'(x)^-$ is Lebesgue integrable over $I \setminus T$, then F is absolutely continuous with density $\varphi(x) = F'(x)$ for all $x \in I \setminus T$

This result is very deep and it will take us too far into advanced measure theory to prove it here. However, it is fairly easy to show [see Exc.3.3]

(3.27.6) Suppose that F is absolutely continuous with density φ. If φ is continuous at a , then F is differentiable at a with

$F'(a) = \varphi(a)$, and if φ is differentiable at a , then F is twice differentiable at a with $F''(a) = \varphi'(a)$

Let $F : \mathbf{R}^k \to \mathbf{R}$ be right continuous with positive increments [see (1.33.3)]. In Section (1.33), I defined the Lebesgue–Stieltjes measure λ_F on $(\mathbf{R}^k, \mathcal{B}(\mathbf{R}^k))$ and I shall use the following three notations for *the Lebesgue-Stieltjes* integral on $\mathbf{R}^k$:

$$\int_A f dF = \int_A f(x)\, F(dx) := \int_A f\, d\lambda_F \qquad \forall A \in \mathcal{B}(\mathbf{R}^k) \tag{3.27.7}$$

$$\int_A f(x_1, \ldots, x_k)\, F(dx_1, \ldots, dx_k) := \int_A f\, d\lambda_F \qquad \forall A \in \mathcal{B}(\mathbf{R}^k) \tag{3.27.8}$$

whenever $1_A f \in L(\lambda_F)$.

Suppose that $\varphi : \mathbf{R}^k \to \overline{\mathbf{R}}_+$ is measurable and that F satisfies [see (1.33.2)]

$$\Delta_a^b F = \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} \varphi(x_1, \ldots, x_k)\, dx_1 \cdots dx_k \tag{3.27.9}$$

for all $a = (a_1, \ldots, a_k) \le (b_1, \ldots, b_k) = b$. Then, as above, we have that λ_F is absolutely λ_k-continuous, φ is the λ_k-density of λ_F , and

$$\int_A f(x)\, F(dx) = \int_A f(x)\varphi(x)\, dx \tag{3.27.10}$$

in the sense that if one of the integrals exists, then so does the other and they are equal. A function $F : \mathbf{R}^k \to \mathbf{R}$ satisfying (3.27.9) for some Borel function $\varphi : \mathbf{R}^k \to \mathbf{R}$ is called *absolutely continuous*, and φ is called *the density* of F . It can be shown that

(3.27.11) If $F : \mathbf{R}^k \to \mathbf{R}$ is continuous with positive increments and if the k-fold partial derivative

$$\varphi(x) = \frac{\partial^k F}{\partial x_1 \cdots \partial x_k}(x)$$

exists for all $x \in \mathbf{R}^k \setminus N$, where N is some countable subset of $\mathbf{R}^k$, then F is absolutely continuous with density φ

Again the proof is difficult and we shall skip it here.

Let $F : \mathbf{R}^k \to \mathbf{R}$ be right continuous with positive increments and let $g : \mathbf{R}^k \to \overline{\mathbf{R}}_+$ be a non-negative Borel function. Then we define *the Lebesgue-Stieltjes convolution* of g and F , denoted $g \star F$, by

(3.27.12) $$g \star F(v) = \int_{\mathbf{R}^k} g(v-x)\, F(dx) \qquad \forall v \in \mathbf{R}^k$$

If f and g are non-negative Borel functions on $\mathbf{R}^k$, we define *the Lebesgue convolution* of f and g, denoted $f \star g$, by

(3.27.13) $$f \star g(v) = \int_{\mathbf{R}^k} f(x) g(v-x)\, dx$$

We define *the discrete convolution* of f and g, denoted $f \star g$, by

(3.27.14) $$f \star g(v) = \sum_{x \in \mathbf{R}^k} f(x) g(v-x)$$

It is, of course, inconsistent to use the same symbol $f \star g$ for the three different convolutions (3.27.12)–(3.27.14). But this is the tradition, and I shall bow to the tradition, confident of your capability to guess which of the three convolutions I mean when you meet them in the text. It is easy to see that the Lebesgue convolution and the discrete convolution are commutative and associative, that is,

(3.27.15) $$f \star g = g \star f \qquad \text{and} \qquad (f \star g) \star h = f \star (g \star h)$$

Later we shall see that this also holds for the Lebesgue–Stieltjes convolution, provided that f, g and h are distribution functions [see (4.5.10)].

3.28 Integration by Parts

Let F and G be increasing right continuous functions from $\mathbf{R}$ into $\mathbf{R}$ and set

$$G(x-) = \lim_{y \uparrow x} G(y) = \sup_{y<x} G(y) \qquad H(x) = F(x)\, G(x)$$

for all $x \in \mathbf{R}$. Let $-\infty \le a \le b \le \infty$. Then we have the following formulas of integration by parts:

(3.28.1) $$\int_a^b (F(x) - F(a))\, G(dx) = \int_a^b (G(b) - G(y-))\, F(dy)$$

(3.28.2) $$\int_a^b F(x)\, G(dx) + \int_a^b G(y-) F(dy) = H(b) - H(a) \qquad \textit{if} \quad a, b \in \mathbf{R}$$

Moreover, if $b = \infty$ *and/or* $a = -\infty$ *and the right-hand side of (3.28.2) exists [that is, the two integrals exist and their sum is* **not** *of the form:* $\infty - \infty$ *], then the limit(s)*

$$(3.28.3) \qquad H(\infty) := \lim_{x\to\infty} H(x) \qquad \textit{and/or} \qquad H(-\infty) := \lim_{x\to-\infty} H(x)$$

exists and (3.28.2) holds.

Remark: Suppose that F and G are absolutely continuous with densities f and g. Then by (3.27.5), we have that $f(x) = F'(x)$ and $g(x) = G'(x)$ for Lebesgue a.a. $x \in \mathbf{R}$. Hence, by (3.27.4) and (3.28.2) we obtain the usual *rule of integration by parts*

$$(3.28.4) \qquad \int_a^b F(x)\,G'(x)\,dx + \int_a^b F'(x)\,G(x)\,dx = H(b) - H(a)$$

Proof: Let $D = \{(x,y) \in \mathbf{R}^2 \mid a < x \le y \le b\}$, then $1_D(x,\cdot) = 1_{[x,b]}$ and $1_D(\cdot,y) = 1_{]a,y]}$. Hence, by Tonelli's theorem (3.21.1) and Exc.2.34 we have

$$\int_a^b F(dy) \int_a^b 1_D(x,y)\,G(dx) = \int_a^b (G(b) - G(y-))\,F(dy)$$
$$= \int_a^b G(dx) \int_a^b 1_D(x,y)\,F(dy) = \int_a^b (F(x) - F(a))\,G(dx)$$

Thus, (3.28.1) is proved. If a and b are finite, then by Exc.2.34 we get

$$G(b)\,(F(b) - F(a)) = \int_a^b G(b)\,F(dy)$$
$$F(a)(G(b) - G(a)) = \int_a^b F(a)\,G(dx)$$

Hence, moving around with the terms in (3.28.1), we see that (3.28.2) follows. If $a = -\infty$ and/or $b = \infty$, then (3.28.2+3) follows from (3.28.2) applied to an appropriate sequence of finite intervals tending to $I =]a,b]$. □

3.29 Substitution

Let $F : \mathbf{R} \to \mathbf{R}$ *be increasing and right continuous, and set* $\alpha = F(-\infty)$, $\beta = F(\infty)$ *and*

$$\hat{F}(y) = \inf\{v \in \mathbf{R} \mid F(v) \ge y\} \qquad \forall\, \alpha < y < \beta$$
$$F(x-) = \lim_{v\uparrow x} F(v) = \sup_{v<x} F(v) \qquad \forall\, x \in \mathbf{R}$$

Then $\hat{F}$ *is an increasing, left continuous function from* $]\alpha,\beta[$ *into* $\mathbf{R}$ *such that for all* $x \in \mathbf{R}$ *and* $y \in]\alpha,\beta[$ *, we have*

$$F(x) \geq y \iff x \geq \hat{F}(y) \tag{3.29.1}$$

$$F(\hat{F}(y)-) \leq y \leq F(\hat{F}(y)) \tag{3.29.2}$$

$$\int_{-\infty}^{\infty} h(x)\, F(dx) = \int_{\alpha}^{\beta} h(\hat{F}(u))\, du \tag{3.29.3}$$

$$\int_{-\infty}^{\infty} g(F(x))\, F(dx) = \int_{\alpha}^{\beta} g(v)\, dv \qquad \text{if } F \text{ is continuous} \tag{3.29.4}$$

whenever $h : \mathbf{R} \to \mathbf{R}$ *and* $g : [\alpha,\beta] \to \mathbf{R}$ *are Borel measurable, and where (3.29.3) and (3.29.4) are to be understood in the usual sense: If one the integrals exists, then so does the other and they are equal. Moreover, we have the following:*

(1): If F *is absolutely continuous with density function* f *[see (3.27)], then* $f(x) = F'(x)$ *Lebesgue a.s. and*

$$\int_{-\infty}^{\infty} g(F(x))F'(x)\, dx = \int_{\alpha}^{\beta} g(u)\, du \tag{3.29.5}$$

(2): If $g : [\alpha,\beta] \to \mathbf{R}$ *is increasing and* $h : [\alpha,\beta] \to \mathbf{R}$ *is decreasing, then we have*

$$\int_{-\infty}^{\infty} g(F(x-))\, F(dx) \leq \int_{\alpha}^{\beta} g(u)\, du \leq \int_{-\infty}^{\infty} g(F(x))\, F(dx) \tag{3.29.6}$$

$$\int_{-\infty}^{\infty} h(F(x))\, F(dx) \leq \int_{\alpha}^{\beta} h(u)\, du \leq \int_{-\infty}^{\infty} h(F(x-))\, F(dx) \tag{3.29.7}$$

Remark: Note that (3.29.5) is the well known *substitution rule* for ordinary integrals, and we may consider formula (3.29.4) as a *substitution rule* for Lebesgue-Stieltjes integrals.

Proof: (3.29.1+2): Evidently, we have that $\hat{F}$ is increasing and finite on $]\alpha,\beta[$. By right continuity of F , it follows easily that $\hat{F}$ is left continuous and satisfies (3.29.1+2).

(3.29.3): Let λ be the Lebesgue measure on $]\alpha,\beta[$ and let ν be the image measure of λ under $\hat{F}$, that is, $\nu(B) = \lambda(\hat{F}^{-1}(B))$. If $B =]a,b]$, then by (3.29.1) and (3.29.2) we have

$$\begin{aligned} \hat{F}^{-1}(B) &= \{y \in]\alpha,\beta[\mid a < \hat{F}(y) \leq b\} \\ &= \{y \in]\alpha,\beta[\mid F(a) < y \leq F(b)\} \end{aligned}$$

Hence, $\nu(]a,b]) = F(b) - F(a)$, and so $\nu = \lambda_F$. Thus, (3.29.3) follows from the first transformation theorem (3.15).

(3.29.4): Set $h(x) = g(F(x))$. Since F is continuous, then by (3.29.2) we have that $h(\hat{F}(u)) = g(u)$. Thus, (3.29.4) follows from (3.29.2) and (3.29.3).

(3.29.5): Follows directly from (3.29.4) and (3.27.4).

(3.29.6+7): Since $F(x-) \le F(x)$ and g is increasing, then by (3.29.2) we have $g(F(\hat{F}(x)-)) \le g(x) \le g(F(\hat{F}(x)))$. Thus, (3.29.6) follows from (3.29.3), and (3.29.7) follows in exactly the same manner. □

3.30 Abstract Integrals as Ordinary Integrals

Let $(\Omega, \mathcal{F}, \mu)$ *be a measure space, let* f *be a measurable function from* Ω *into* $\overline{\mathbf{R}}$ *, and let* $F : \mathbf{R} \to \mathbf{R}$ *be an increasing right continuous function. If* $R(x)$ *and* $\tilde{R}(x)$ *are defined by*

$$R(x) = \begin{cases} \mu(f > x) & \text{if } x > 0 \\ -\mu(f \le x) & \text{if } x \le 0 \end{cases}$$

$$\tilde{R}(x) = \begin{cases} \mu(f \ge x) & \text{if } x > 0 \\ -\mu(f < x) & \text{if } x \le 0 \end{cases}$$

Then we have

$$\int_\Omega f\, d\mu = \int_{-\infty}^{\infty} R(t)\, dt = \int_{-\infty}^{\infty} \tilde{R}(t)\, dt \tag{3.30.1}$$

$$\int_\Omega \{F(f(\omega)) - F(0)\}\, \mu(d\omega) = \int_{-\infty}^{\infty} \tilde{R}(x)\, F(dx) \tag{3.30.2}$$

where (3.30.1+2) are to be understood in the usual sense: If one of the integrals exists, then so does the other and they are equal. Moreover, if μ *is sum-finite, and* f *is non-negative, then we have*

$$\mu \otimes \lambda(G) = \int_\Omega f\, d\mu \tag{3.30.3}$$

where $G = \{(\omega, x) \mid 0 \le x \le f(\omega)\}$ *is the region in* $\Omega \times \mathbf{R}_+$ *below the graph of* f *, and* $\mu \otimes \lambda$ *denotes the product measure of* μ *and the Lebesgue measure* λ *on* $\mathbf{R}$ *.*

Remark: Formulas (3.30.1) and (3.30.2) show how to compute abstract integral as ordinary Lebesgue or Lebesgue-Stieltjes integrals. Note that formula (3.30.3) simply states, that the integral of a non-negative function is the "area" of the region below the graph. Suppose that F is absolutely continuous with density f. Then $f(x) = F'(x)$ Lebesgue a.s. by (3.27.5). Hence, by (3.30.2) we have

$$\int_\Omega (F(f(\omega)) - F(0))\mu(d\omega) = \int_{-\infty}^{\infty} \tilde{R}(x)\, F'(x)\, dx \tag{3.30.4}$$

Specializing to some concrete functions F, we find

$$\int_\Omega |f(\omega)|^q\, \mu(d\omega) = \int_0^\infty q x^{q-1}\, \mu(|f| \geq x)\, dx \tag{3.30.5}$$

$$\int_\Omega \left(e^{\alpha f(\omega)} - 1\right)\mu(d\omega) = \int_0^\infty \left(e^{\alpha x}\mu(f \geq x) - e^{-\alpha x}\mu(f < -x)\right)dx \tag{3.30.6}$$

Proof: (3.30.1): Note that $R(t) = \mu(f > t) = \mu(f^+ > t)$ for $t > 0$. Hence, by (3.1.1) we have

$$s_n(f^+) = 2^{-n} \sum_{j=1}^{\infty} R(j2^{-n})$$

Let $I_{nj} =]j2^{-n}, (j+1)2^{-n}]$ and $\psi_n(t) = 2^{-n}\, int^+(2^n t)$ where $int^+(x)$ is the smallest integer, which is strictly larger than x. Then, evidently, we have that ψ_n is increasing and $t < \psi_n(t) \leq t + 2^{-n}$ for all $t \in \mathbf{R}$. In particular, we see that $\psi_n(t) \downarrow t$, and if

$$R_n(t) = \sum_{j=1}^{\infty} R(j2^{-n})\, 1_{I_{n,j-1}}(t)$$

then $R_n(t) = R(\psi_n(t))$ for all $t > 0$. Note that $\{f > \psi_n(t)\} \uparrow \{f > t\}$. Hence, by (1.4.2) we have that $R_n(t) \uparrow R(t)$ for all $t > 0$. Thus, by (3.8.2), (3.1.2) and the monotone convergence theorem (3.4) we get

$$\int_0^\infty R(t)\, dt = \lim_{n\to\infty} \int_0^\infty R_n(t)\, dt = \lim_{n\to\infty} \left\{ 2^{-n} \sum_{j=1}^{\infty} R(j2^{-n}) \right\}$$

$$= \lim_{n\to\infty} s_n(f^+) = \int_\Omega f^+\, d\mu$$

Similarly, if $t \leq 0$, then $R(t) = -\mu(f \leq t) = -\mu(f^- \geq -t)$; and exactly as above, one shows that

$$\int_{-\infty}^0 R(t)\, dt = -\int_\Omega f^-\, d\mu$$

Hence, the first equality in (3.30.1) follows from definition (3.2.2). To prove the second equality, note that R is decreasing and $R(t) \le \tilde{R}(t) \le R(s)$ for all $s < t$. Hence, $R(t) \le \tilde{R}(t) \le R(t-)$, and since the decreasing function R has at most countably many discontinuity points, we conclude that $R(t) = \tilde{R}(t)$ Lebesgue a.s., and so the second equality in (3.30.1) follows from (3.3.4).

(3.30.2): Let $F_0(x) = F(x) - F(0)$ and let $\hat{F}_0(y)$ be defined as in (3.29) for $\alpha = F_0(-\infty) < x < F_0(\infty) = \beta$. Set $g(\omega) = F_0(f(\omega))$ and let $\tilde{Q}$ be defined as $\tilde{R}$ with f replaced by g. Then by (3.29.1) it follows easily that $\tilde{Q}(x) = \tilde{R}(\hat{F}_0(x))$ for all $\alpha < x < \beta$, and $\tilde{Q}(u) = \tilde{Q}(v) = 0$ for all $u < \alpha$ and all $v > \beta$. Thus, by (3.30.1) and (3.29.3) we have:

$$\begin{aligned}\int_\Omega g\,d\mu &= \int_{-\infty}^{\infty} \tilde{Q}(x)\,dx = \int_\alpha^\beta \tilde{R}(\hat{F}_0(x))\,dx \\ &= \int_{-\infty}^{\infty} \tilde{R}(x)\,F_0(dx) = \int_{-\infty}^{\infty} \tilde{R}(x)\,F(dx)\end{aligned}$$

because $F_0(b) - F_0(a) = F(b) - F(a)$ and so $\lambda_{F_0} = \lambda_F$.

(3.30.3): Follows directly from the definition of the product measure (3.20.6). □

3.31 Continuity and Differentiability of Integrals

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let (S, d) be a metric space. Suppose that $s_0 \in S$ and $f : S \times \Omega \to \overline{\mathbf{R}}$ satisfies

(3.31.1) $f(s,\cdot)$ *is μ-integrable* $\forall s \in S$

(3.31.2) $f(\cdot,\omega)$ *is continuous at s_0 for μ-a.a. $\omega \in \Omega$*

(3.31.3) $\{f(s_n,\cdot) \mid n = 1,2,\cdots\}$ *is uniformly μ-integrable whenever $\{s_n\}$ is a sequence in S so that $s_n \to s_0$*

Then the function $F : S \to \mathbf{R}$ given by

(3.31.4) $$F(s) = \int_\Omega f(s,\omega)\,\mu(d\omega)$$

is continuous at s_0.

Suppose that S is an interval on $\mathbf{R}$ and that μ is sum-finite. Let g and φ be measurable functions from $S \times \Omega$ into $\mathbf{R}$ such that

(3.31.5) $g(c,\cdot)$ *is μ-integrable for some $c \in S$*

(3.31.6) $$g(b,\omega) - g(a,\omega) = \int_a^b \varphi(s,\omega)ds \qquad \forall \omega \in \Omega\ \forall a,b \in S\ ,\ a \le b$$

(3.31.7) $$\int_\Omega \mu(d\omega) \int_a^b |\varphi(s,\omega)|\, ds < \infty \qquad \forall\, a,b \in S\ ,\ a \le b$$

Then $g(s,\cdot) \in L^1(\mu)$ *for all* $s \in S$, *and* $\varphi(s,\cdot) \in L^1(\mu)$ *for* λ*-a.a.* $s \in S$, *where* λ *is the Lebesgue measure. Moreover, if* G *and* Φ *are given by*

(3.31.8) $$G(s) = \int_\Omega g(s,\omega)\, \mu(d\omega) \qquad \forall\, s \in S$$

(3.31.9) $$\Phi(s) = \int_\Omega \varphi(s,\omega)\ \mu(d\omega) \quad \textit{for } \lambda\textit{-a.a. } s \in S$$

Then we have

(3.31.10) $$G(b) - G(a) = \int_a^b \Phi(s)ds \qquad \forall\, a,b \in S\ ,\ a \le b$$

(3.31.11) *If* Φ *is continuous at* s_0 , *then* G *is differentiable at* s_0 *with* $G'(s_0) = \Phi(s_0)$, *and if* Φ *is differentiable at* s_0 , *then* G *is twice differentiable at* s_0 *with* $G''(s_0) = \Phi'(s_0)$

(3.31.12) G *is differentiable at* s *with* $G'(s) = \Phi(s)$ *for* λ *-a.a.* $s \in S$

Proof The first part of the theorem follows directly from Dunford-Pettis' theorem (3.23). So suppose that S is an interval and that (3.31.5)–(3.31.7) hold. By (3.31.6+7) we have that $g(a,\cdot) - g(b,\cdot)$ is μ-integrable. Hence, by (3.31.5) we see that $g(s,\cdot) \in L^1(\mu)$ for all $s \in S$. By (3.31.7) and Tonelli's theorem, we see that $\varphi(s,\cdot) \in L^1(\mu)$ for λ-a.a. $s \in S$. Formula (3.31.10) follows from Fubini's theorem (3.21.2) in the following way:

$$\begin{aligned} G(b) - G(a) &= \int_\Omega (g(b,\omega) - g(a,\omega))\, \mu(d\omega) \\ &= \int_\Omega \mu(d\omega) \int_a^b \varphi(s,\omega)\, ds \\ &= \int_a^b \Phi(s)\, ds \end{aligned}$$

But then (3.31.11) and (3.31.12) follow from (3.27.5) and (3.27.6). □

3.32 Integrals of Complex Functions

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let f be a measurable function from Ω into the complex numbers $\mathbf{C}$. Then we can write f in the form $f(\omega) = g(\omega) + ih(\omega)$, where $g = \Re f$ is *the real part of* f and $h = \Im f$ is *the imaginary part* of f. We say that f is *μ-integrable* if $g = \Re f$ and $h = \Im f$ are both μ-integrable and if so, we define *the μ-integral* of f as follows:

$$\int_\Omega f d\mu = \int_\Omega (\Re f)\, d\mu + i \int_\Omega (\Im f)\, d\mu \tag{3.32.1}$$

$$L^1(\mu, \mathbf{C}) = \{\, f \in M(\Omega, \mathcal{F}, \mathbf{C}) \mid f \text{ is } \mu\text{-integrable} \,\} \tag{3.32.2}$$

where $M(\Omega, \mathcal{F}, \mathbf{C})$ denotes the set of all measurable functions from Ω into $\mathbf{C}$. If $0 < q < \infty$, then, as in (3.22), we let $L^q(\mu, \mathbf{C})$ denote the set of all measurable functions $f : \Omega \to \mathbf{C}$ such that $|f(\cdot)|^q$ is μ-integrable. It is now an easy (but lengthy) exercise to show that all the theorems for real-valued integrals, which make sense for complex-valued integrals, can be extended to complex-valued functions. For instance, Lebesgue's dominated convergence theorem (3.6), Markov's inequality (3.9), Hölder's inequality (3.11), Minkowski's inequality (3.12), etc. holds for complex functions. We have the following version of Beppo Levi's theorem for complex–valued functions:

(3.32.3) If $\sum_{n=1}^{\infty} \int_\Omega |f_n|\, d\mu < \infty$ then $f := \sum_{n=1}^{\infty} f_n \in L^1(\mu, \mathbf{C})$ and

$$\int_\Omega \left\{ \sum_{n=1}^{\infty} f_n \right\} d\mu = \sum_{n=1}^{\infty} \left\{ \int_\Omega f_n\, d\mu \right\}$$

Moreover, by definition we have

$$f \in L^q(\mu, \mathbf{C}) \iff f \in M(\Omega, \mathcal{F}, \mathbf{C}) \text{ and } |f| \in L^q(\mu) \tag{3.32.4}$$

$$\int_\Omega (\Re f) d\mu = \Re\left(\int_\Omega f d\mu \right) \qquad \forall f \in L^1(\mu, \mathbf{C}) \tag{3.32.5}$$

$$\int_\Omega (\Im f)\, d\mu = \Im\left(\int_\Omega f\, d\mu \right) \qquad \forall f \in L^1(\mu, \mathbf{C}) \tag{3.32.6}$$

$$\int_\Omega \bar{f}\, d\mu = \overline{\int_\Omega f\, d\mu} \qquad \forall f \in L^1(\mu, \mathbf{C}) \tag{3.32.7}$$

(3.32.8) $L^q(\mu, \mathbf{C})$ is a complex vector space and: $f \curvearrowright \int f\, d\mu$ is a complex linear function: $L^1(\mu, \mathbf{C}) \to \mathbf{C}$

where $\bar{z}$ is *the complex conjugated* of z, that is, if $z = x+iy$, then $\bar{z} = x-iy$. Let $f \in L^1(\mu, \mathbf{C})$ and set $c = \int f d\mu$, since $|\Re z| \leq |z|$ we get

$$|c|^2 = \int_\Omega (\bar{c}f)\, d\mu = \int_\Omega \Re(\bar{c}f)\, d\mu \leq \int_\Omega |\bar{c}f|\, d\mu = |c| \int |f|\, d\mu$$

Thus, we have proved the following extension of (3.3.9):

$$\left| \int_\Omega f\, d\mu \right| \leq \int_\Omega |f|\, d\mu \tag{3.32.9}$$

3.33 Examples of Convergence and Divergence

Consider the Lebesgue measure λ on $[0,1]$. Let $0 < \alpha < \infty$ and set $f_n(x) = n^\alpha (1-x)^n$ for all $x \in [0,1]$ and all $n = 1, 2, \ldots$. Then $f_n(x) \to 0$ for all $x \in]0,1[$ and

$$\int_0^1 f_n(x)\, dx = \frac{n^\alpha}{n+1} \longrightarrow \begin{cases} 0 & \text{if } \alpha < 1 \\ 1 & \text{if } \alpha = 1 \\ \infty & \text{if } \alpha > 1 \end{cases}$$

Moreover, a straightforward computation shows that

$$0 \leq f_n(x) \leq \alpha^\alpha x^{-\alpha} \quad \forall x \in]0,1[\ , \ \sup_{n \geq 1} f_n(x) \geq \tfrac{1}{8} x^{-\alpha} \quad \forall x \in]0, \tfrac{1}{2}[$$

Hence, $\{f_n\}$ is dominated by an integrable function if and only if $\alpha < 1$, and exactly in this case we have that the limit of the integrals equals the integral of the limit. Now let

$$g_k(x) = \begin{cases} 1 & \text{if } j\, 2^{-n} \leq x \leq (j+1)\, 2^{-n} \\ 0 & \text{otherwise} \end{cases}$$

where n and j are given by $n = int(\log_2 k)$ and $j = k - 2^{-n}$. Then $\int_0^1 g_k(x) dx = 2^{-n} \to 0$. Hence, $g_k \to 0$ in 1-mean and in λ-measure. However, a straightforward computation shows that

$$\limsup_{k \to \infty} g_k(x) = 1 \quad \text{and} \quad \liminf_{k \to \infty} g_k(x) = 0 \quad \forall 0 \leq x \leq 1$$

Thus, $\{g_k\}$ is nowhere convergent. Finally, let $I_k = [\frac{1}{k+1}, \frac{1}{k}]$ and $h_k = \frac{1}{x} 1_{I_k}(x)$ for all $x \in [0,1]$ and all $k = 1, 2, \ldots$. Then we have

$$\int_0^1 h_k(x)^2\, dx = 1 \quad , \quad \int_0^1 h_k(x)\, dx = \log\left(\tfrac{k+1}{k}\right) \to 0$$

Hence, by (3.24.5) we see that $\{h_k\}$ is uniformly λ-integrable, but $\sup_k h_k(x) = 1/x$ and thus, $\{h_k\}$ is not dominated by an integrable function. Moreover, $h_k \to 0$ a.s., in 1-mean and in λ-measure, but $\{h_k\}$ does not converge to 0 in 2-mean.

3.34 Failure of Tonelli-Fubini's Theorem

Tonelli's theorem fails if we drop the assumption that the measures are sum-finite. To see this, let λ be the Lebesgue measure on $[0,1]$ and let $\#$ be the counting measure on $[0,1]$. Let $D = \{(x,x) \mid 0 \le x \le 1\}$ be the diagonal in the unit square $[0,1] \times [0,1]$. Then we have

$$\int_0^1 dx \int_0^1 1_D(x,y)\,\#(dy) = \int_0^1 1\,dx = 1$$
$$\int_0^1 \#(dy) \int_0^1 1_D(x,y)\,dx = \int_0^1 0\,\#(dy) = 0$$

showing that Tonelli's theorem (3.21.1) fails. Note that λ is a finite measure and $\#$ is a finitely founded measure, but of course $\#$ is not sum-finite.

If all the six integrals in (3.21) are infinite, then Fubini's theorem (3.21.2) fails. To see this, let λ be the Lebesgue measure on $]0,1[$ and set

$$h(x,y) = \frac{x^2 - y^2}{(x^2+y^2)^2} \qquad \text{if } 0 < x < 1 \text{ and } 0 < y < 1$$

Then h is continuous, and $h(x,\cdot)$ and $h(\cdot,y)$ are bounded for any fixed $x,y \in]0,1[$; thus, the two partial integrals exist and we have

$$\int_0^1 h(x,y)\,dy = \frac{1}{1+x^2} \quad , \quad \int_0^1 h(x,y)\,dx = \frac{-1}{1+y^2}$$

which are both integrable functions, and

$$\int_0^1 dx \int_0^1 h(x,y)\,dy = \frac{\pi}{4} \quad , \quad \int_0^1 dy \int_0^1 h(x,y)\,dx = -\frac{\pi}{4}$$

Thus, even though the two iterated integrals exist, they are different! However, an easy computation shows that all the six integrals in (3.21) are infinite.

3.35 A Singular Integral

Let f and g be complex-valued Borel functions from $\mathbf{R}_+$ into $\mathbf{C}$, and let $p \in \mathbf{R}$ be a real number and $\lambda \in \mathbf{C}$ be a complex number satisfying

(3.35.1) $\quad \exists\, c > 0$ *so that* $\displaystyle\int_0^c \frac{|f(s) - \lambda s^p|}{s^{p+1}}\,ds < \infty$

(3.35.2) $\quad \displaystyle\int_0^\infty \frac{|f(s) - g(s)|}{s^{p+1}}\,ds < \infty \quad , \quad \int_\varepsilon^\infty \frac{|f(s)|}{s^{p+1}}\,ds < \infty \quad \forall \varepsilon > 0$

Then $(f(as) - a^p g(s))s^{-p-1}$ *is Lebesgue-integrable over* $]0,\infty[$ *for all* $a > 0$ *and we have*

$$\int_0^\infty \frac{f(as) - a^p g(s)}{s^{p+1}} ds = a^p \int_0^\infty \frac{f(s) - g(s)}{s^{p+1}} ds - \lambda \log a \qquad \forall a > 0 \tag{3.35.3}$$

Remark It is easily seen that there exists at most one complex number $\lambda \in \mathbf{C}$ satisfying (3.35.1). If $s^{-p-1} f(s)$ is Lebesgue-integrable over $]0,\infty[$, then (3.35.1) and (3.35.2) hold with $\lambda = 0$. Moreover, in this case, (3.35.3) follows directly from linearity of the integral. However, if $\lambda \neq 0$, then a formal computation of the integral on the left-hand side of (3.35.3) just gives the first term on the right-hand side which is wrong. The reason being that $s^{-p-1} f(s)$ is not integrable. In this case, we say that f has a singularity, and you may consider $\lambda \log a$ as a term compensating for the singularity.

Proof Let $a > 0$ be given, and set $h_a(s) = (f(as) - a^p g(s))s^{-p-1}$, then we have

$$|h_a(s)| \leq s^{-p-1} (|f(as)| + a^p|f(s)| + a^p|f(s) - g(s)|)$$
$$|h_a(s)| \leq s^{-p-1} (|f(as) - \lambda(as)^p| + a^p|f(s) - \lambda s^p| + a^p|f(s) - g(s)|)$$

Hence, we see that (3.35.1+2) imply that h_a is Lebesgue-integrable over $]0,\infty[$. Let $0 < \varepsilon < \frac{c}{1+a}$ be given where c is the positive constant from (3.35.1). Then we have

$$\begin{aligned}
\int_\varepsilon^\infty h_a(t)\, dt &= \int_\varepsilon^\infty s^{-p-1} f(as) ds - \int_\varepsilon^\infty a^p g(s) s^{-p-1} ds \\
&= a^p \left\{ \int_{\varepsilon a}^\infty s^{-p-1} f(s) ds - \int_\varepsilon^\infty g(s) s^{-p-1} ds \right\} \\
&= a^p \left\{ \int_{\varepsilon a}^\varepsilon s^{-p-1} f(s) ds + \int_\varepsilon^\infty h_1(s) ds \right\} \\
&= a^p \left\{ \int_{\varepsilon a}^\varepsilon \lambda s^{-1} ds + \int_{\varepsilon a}^\varepsilon s^{-p-1}(f(s) - \lambda s^p) ds + \int_\varepsilon^\infty h_1(s) ds \right\} \\
&= a^p \left\{ -\lambda \log a + \int_{\varepsilon a}^\varepsilon s^{-p-1}(f(s) - \lambda s^p) ds + \int_\varepsilon^\infty h_1(s) ds \right\}
\end{aligned}$$

where all the integrals exists and are finite because $\varepsilon < c$ and $\varepsilon a < c$. Letting $\varepsilon \downarrow 0$, we see that (3.35.3) holds. □

3.36 The Mellin Transform

Let g be a Borel function from $\mathbf{R}_+$ into $\overline{\mathbf{R}}$ or $\mathbf{C}$. Then we let $\mathcal{E}(g)$ denote the set of all $p \in \mathbf{R}$ satisfying $s^{p-1}\, g(s) \in L^1(]0,\infty[,\lambda)$, where λ as usual denotes the Lebesgue measure. Moreover, if g takes values in $\overline{\mathbf{R}}$, then we let $\mathcal{E}^*(g)$ denote the set of all $p \in \mathbf{R}$ satisfying $s^{p-1}\, g(s) \in L(]0,\infty[,\lambda)$; and if g is complex-valued but not real-valued, then we set $\mathcal{E}^*(g) := \mathcal{E}(g)$. We may then define *the Mellin transform*, denoted M_g, also called *the Euler integral* of g, as follows:

$$M_g(p) := \int_0^\infty t^{p-1} g(t)\, dt \qquad \forall p \in \mathcal{E}^*(g) \tag{3.36.1}$$

The set $\mathcal{E}(g)$ is called *the domain* of M_g, and the set $\mathcal{E}^*(g)$ is called *the extended domain* of M_g. Note that $\mathcal{E}(g) \subseteq \mathcal{E}^*(g)$ and $M_g(p) = \pm\infty$ for all $p \in \mathcal{E}^*(g) \setminus \mathcal{E}(g)$. Let $u, v \in \mathbf{R}$ and $0 < \alpha < 1$ be given and set $w = \alpha u + (1-\alpha)v$. Applying Hölder's inequality (3.11) with $p = \frac{1}{\alpha}$, $q = \frac{1}{1-\alpha}$ and $r = 1$, we find

$$\begin{aligned}\int_0^\infty s^{w-1}|g(s)|\, ds &= \int_0^\infty |s^{u-1}g(s)|^{1/p}\, |s^{v-1}g(s)|^{1/q}\, ds \\ &\le \left\{\int_0^\infty s^{u-1}|g(s)|\, ds\right\}^\alpha \left\{\int_0^\infty s^{v-1}|g(s)|\, ds\right\}^{1-\alpha}\end{aligned}$$

and since $e^{f(t)}$ is convex whenever f is convex, then taking logarithms we get

(3.36.2) $\mathcal{E}(g)$ is an interval, and if $g \ge 0$ then M_g and $\log M_g$ are convex functions on $\mathcal{E}(g)$. Moreover, if $p \in \mathcal{E}(f) \cap \mathcal{E}(g)$, then $M_{a\,f+b\,g}(p) = a\, M_f(p) + b\, M_g(p)$ for all $a, b \in \mathbf{C}$

Table 3.1

$g(t)$, $t > 0$	$\mathcal{E}(g)$	$M_g(p)$, $p \in \mathcal{E}(g)$
e^{-t}	$]0,\infty[$	$\Gamma(p)$
$1-e^{-t}$	$]-1,0[$	$-\Gamma(p)$
$1-\cos t$	$]-2,0[$	$C(-p)$
$\sin t$	$]-1,0[$	$\Gamma(p)\sin\frac{p\pi}{2}$
$e^{it}-1$	$]-1,0[$	$-\Gamma(p)\,e^{ip\pi/2}$
$e^{it}-1-it$	$]-2,-1[$	$-\Gamma(p)\,e^{ip\pi/2}$
$1_{]0,1[}(t)\,(1-t)^{\alpha-1}$, $\alpha > 0$	$]0,\infty[$	$\frac{\Gamma(p)\,\Gamma(\alpha)}{\Gamma(p+\alpha)}$
$1_{]1,\infty[}(t)\,t^{\beta}$, $\beta > 0$	$]-\infty,-\beta[$	$-\frac{1}{\beta+p}$

The functions $\Gamma(p)$ and $C(-p)$ are defined by (3.37.1+6) and (3.37.14)

Let $g_a(t) = g(at)$. Then substituting $s = at$ in (6.28.1), we find

(3.36.3) If $p \in \mathcal{E}^*(g)$, then $M_{g_a}(p) = a^{-p}\,M_g(p)$ in either of the following three cases. (i): $a > 0$. (ii): $a = 0$, $p < 0$ and $g(0) = 0$. (iii): $a = 0$, $p > 0$, $M_g(p) > 0$ and $g(0) > 0$

Suppose that g is differentiable with derivative $g'(t)$. Then a simple partial integration shows

(3.36.4) $M_{g'}(p+1) = -p\,M_g(p)$ if $p \in \mathcal{E}(g) \setminus 0$ and $p+1 \in \mathcal{E}(g')$

In Table 3.1 you will find the most important Mellin transforms and their domains.

3.37 The Gamma Function and Related Functions

Note that if $g(s) = e^{-s}$, then $\mathcal{E}(g) =]0,\infty[$ and we define *the gamma function*, denoted $\Gamma(\lambda)$, to be the Mellin transform of e^{-s} ; that is,

$$(3.37.1) \qquad \Gamma(\lambda) := M_g(\lambda) = \int_0^\infty s^{\lambda-1}\, e^{-s}\, ds \qquad \forall \lambda > 0$$

Since $(\frac{d}{d\lambda})^n t^\lambda = t^\lambda (\log t)^n$, it follows from (3.31), (3.36.2) and (3.36.4), that we have

$$(3.37.2) \qquad \Gamma(\lambda+1) = \lambda\Gamma(\lambda) \qquad \forall \lambda > 0$$

(3.37.3) $\Gamma(\lambda)$ and $\log\Gamma(\lambda)$ are a convex function on $]0,\infty[$

$$(3.37.4) \qquad \Gamma^{(n)}(\lambda) = \int_0^\infty t^{\lambda-1}(\log t)^n e^{-t} dt \qquad \forall \lambda > 0 \quad \forall n \in \mathbf{N}$$

$$(3.37.5) \qquad \Gamma(\lambda+n) = \lambda(\lambda+1)\cdots(\lambda+n-1)\,\Gamma(\lambda) \qquad \forall \lambda > 0\, \forall n \in \mathbf{N}$$

Observe that equation (3.37.5) may be used to ***define*** $\Gamma(\lambda)$ when $\lambda < 0$ and λ is not an integer, as follows:

$$(3.37.6) \qquad \Gamma(\lambda) := \frac{\Gamma(\lambda+n)}{\lambda(\lambda+1)\cdots(\lambda+n-1)} \quad \text{if } -n < \lambda < 1-n\,,\ n \in \mathbf{N}$$

Substituting $s = ct$, we have $\int_0^\infty t^{p-1}e^{-ct}dt = c^{-p}\,\Gamma(p)$ for all $p, c > 0$. Suppose that $c = a + ib$ is a complex number with $a = \Re c > 0$. Then $|t^{p-1}\, e^{-ct}| = t^{p-1}\, e^{-at}$ and so $t^{p-1}\, e^{-ct}$ is Lebesgue-integrable on $]0,\infty[$, and by methods from complex analysis, it can be shown that the complex substitution $s = ct$ remains valid and that we have

$$(3.37.7) \qquad \int_0^\infty t^{p-1}e^{-ct}\, dt = c^{-p}\,\Gamma(p) \qquad \text{if } p > 0 \text{ and } \Re c > 0$$

where c^q is *the main q-th root* of c ; that is, if we write c in polar form $c = |c|e^{i\theta}$ with $-\pi < \theta \le \pi$, then

$$(3.37.8) \qquad c^q = |c|^q e^{iq\theta} \qquad \forall q \in \mathbf{R}$$

Let $h(t) = e^{-ct} - 1$. Then $h'(t) = -ce^{-ct}$ and $\mathcal{E}(h) =]-1, 0[$. Hence, if $-1 < p < 0$ and $\Re c > 0$ then by (3.36..4) and (3.37.7) we find

$$\int_0^\infty t^{p-1}\left(e^{-ct} - 1\right) dt = M_h(p) = -\tfrac{1}{p}\, M_{h'}(p+1)$$

$$= -\tfrac{c}{p}\,\Gamma(p+1)\, c^{-p-1} = c^{-p}\,\Gamma(p)$$

Since $|h(t)| \le 2 \wedge |ct|$. Then by Lebesgue's dominated convergence theorem (3.6), this also holds for $\Re c = 0$. Thus, if $I_1 =]-1, 0[$ we have

$$(3.37.9) \qquad \int_0^\infty t^{p-1}(e^{-ct} - 1)\, dt = c^{-p}\,\Gamma(p) \quad \text{if } p \in I_1 \ , \ \Re c \ge 0$$

Moreover, since $\frac{d}{dt}(e^{-ct} - 1 + ct) = -c(e^{-ct} - 1)$, then, as above, we find:

$$(3.37.10) \qquad \int_0^\infty t^{p-1}(e^{-ct} - 1 + ct)\, dt = c^{-p}\,\Gamma(p) \quad \text{if } p \in I_2 \ , \ \Re c \ge 0$$

where $I_2 =]-2, -1[$. Similarly, if $I_n =]-n, 1-n[$ for $n \in \mathbf{N}$, then repeating the argument n times, we get

$$(3.37.11) \qquad \int_0^\infty t^{p-1}\left(e^{-ct} - \sum_{j=1}^{n-1} \frac{(-ct)^j}{j!}\right) dt = c^{-p}\Gamma(p) \quad \text{if } p \in I_n \ , \ \Re c \ge 0$$

Suppose that $c = -iv$ where $v \in \mathbf{R}$. Then $\Re c = 0$ and $c = |v|e^{i\pi/2}$ if $v \le 0$, and $c = |v|e^{-i\pi/2}$ if $v \ge 0$. Hence, if $v \in \mathbf{R}$, then by (3.37.9+10) we find

$$(3.37.12) \qquad \int_0^\infty t^{p-1}(e^{ivt} - 1)\, dt = \frac{C(-p)(i\mathcal{K}(-p, v) - 1)}{|v|^p} \quad \text{if } p \in I_1$$

$$(3.37.13) \qquad \int_0^\infty t^{p-1}(e^{ivt} - 1 - ivt)\, dt = \frac{C(-p)(i\mathcal{K}(-p, v) - 1)}{|v|^p} \quad \text{if } p \in I_2$$

where $\frac{1}{|0|^p} = 0$ for all $p < 0$ and $C(q)$ and $\mathcal{K}(q, v)$ are given by

$$(3.37.14) \qquad C(q) = \begin{cases} -\Gamma(-q)\cos\frac{q\pi}{2} & \text{if } 0 < q < 2 \ , \ q \ne 1 \\ \frac{\pi}{2} & \text{if } q = 1 \end{cases}$$

$$(3.37.15) \qquad \mathcal{K}(q, v) = \begin{cases} (\operatorname{sign} v)\,\tan\frac{q\pi}{2} & \text{if } 0 < q \le 2 \ , \ q \ne 1 \ , \ v \in \mathbf{R} \\ \frac{2}{\pi}(\operatorname{sign} v)\log|v| & \text{if } q = 1 \ , \ v \ne 0 \\ 0 & \text{if } q = 1 \ , \ v = 0 \end{cases}$$

Note that we only use $C(q)$ and $\mathcal{K}(q, v)$ in (3.37.12+13) for $q \ne 1$ and $0 < q < 2$. The reason for setting $C(1) = \frac{\pi}{2}$ is simply to make $C(q)$ continuous at 1 . Note that

$$C(q) = -\Gamma(-q)\cos\frac{q\pi}{2} = \frac{\Gamma(3-q)}{q(2-q)}\,\frac{\cos\frac{q\pi}{2}}{1-q}$$

and thus, $C(q) \to \frac{\pi}{2}$ as $q \to 1$. Taking real parts in (3.37.12+13) we get

$$(3.37.16) \qquad C(q) = \int_0^\infty \frac{1-\cos t}{t^{q+1}}\,dt \qquad \forall\, 0 < q < 2$$

[Formally we only obtain (3.37.16) for $q \neq 1$, but if $q = 1$ then (3.37.16) follows from continuity of $C(q)$ and Lebesgue's theorem on dominated convergence (5.6)].

The reason for defining $\mathcal{K}(1,v) = \frac{2}{\pi}(\text{sign } v)\log|v|$ is a bit more involved. Note that the integrals in (3.37.12+13) does not exist for $p = -1$. But let $f(t) = e^{it} - 1$ and $g(t) = i\sin t$. Then the two pairs $(f(t), g(t))$ and $(f(-t), g(-t))$ satisfy (3.35.1) and (3.35.2) with $p = 1$ and $\lambda = \pm i$, and since $f(t) - g(t) = \cos t - 1$ and $C(1) = \frac{\pi}{2}$, then by (3.35.3) we get

$$(3.37.17) \qquad \int_0^\infty t^{-2}(e^{ivt} - 1 - iv\sin t)\,dt = -|v|C(1)(1 - i\mathcal{K}(1,v)) \qquad \forall\, v \in \mathbf{R}$$

3.38 Sums as Integrals

Let Ω be a set, and let $\#$ be the counting measure on $(\Omega, 2^\Omega)$. It is customary to use the notation

$$\ell(\Omega) := L(\#) \qquad \text{and} \qquad \ell^1(\Omega) := L^1(\#)$$

After a straightforward computation, which I shall leave to zealous readers, we get

$$(3.38.1) \qquad f \in \ell(\Omega) \iff \text{either } \sum_{\omega\in\Omega} f^+(\omega) < \infty \text{ or } \sum_{\omega\in\Omega} f^-(\omega) < \infty$$

$$(3.38.2) \qquad f \in \ell^1(\Omega) \iff \sum_{\omega\in\Omega} |f(\omega)| < \infty$$

$$(3.38.3) \qquad \int_\Omega f\,d\# = \sum_{\omega\in\Omega} f^+(\omega) - \sum_{\omega\in\Omega} f^-(\omega) = \sum_{\omega\in\Omega} f(\omega) \qquad \forall\, f \in \ell(\Omega)$$

Hence, sums are integrals with respect to the counting measure, and so all results we have proved for integrals can be applied to sums.

3.39 The Space $L^{(\infty)}(\mu, \mathbf{C})$

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then we let $L^{(\infty)}(\mu, \mathbf{C})$ denote the set of all complex-valued measurable functions $f : \Omega \to \mathbf{C}$ satisfying $f \in L^q(\mu, \mathbf{C})$ for all $0 < q < \infty$. Moreover, we define

(3.39.1) $$R_\mu(f) := \limsup_{n\to\infty} \left\{ \frac{1}{n!} \int_\Omega |f(\omega)|^n \, \mu(d\omega) \right\}^{1/n} \quad \forall f \in M(\Omega, \mathcal{F}, \mathbf{C})$$

(3.39.2) $$\rho_\mu(f) := \frac{1}{R_\mu(f)} \quad \forall f \in M(\Omega, \mathcal{F}, \mathbf{C})$$

Let $f_1, \ldots, f_k : \Omega \to \mathbf{C}$ be measurable functions. Then by Minkowski's inequality (3.12) and Hölder's inequality (3.11), we get

(3.39.3) $$R_\mu\left(\sum_{j=1}^k c_j f_j \right) \le \sum_{j=1}^k |c_j| \, R_\mu(f_j) \quad \forall c_1, \ldots, c_k \in \mathbf{C}$$

(3.39.4) $L^{(\infty)}(\mu, \mathbf{C})$ is a complex vector space such that $\bar{f} \in L^{(\infty)}(\mu, \mathbf{C})$ and $f \cdot g \in L^{(\infty)}(\mu, \mathbf{C})$ for all $f, g \in L^{(\infty)}(\mu, \mathbf{C})$

Moreover, by Hölder's inequality (3.11) we have

$$\int_\Omega |f|^{2n+1} \, d\mu = \int_\Omega |f|^n \, |f|^{n+1} \, d\mu \le \left\{ \int_\Omega |f|^{2n} \, d\mu \right\}^{1/2} \left\{ \int_\Omega |f|^{2n+2} \, d\mu \right\}^{1/2}$$

$$\int_\Omega |f|^q \, d\mu \le \mu(\Omega)^{(r-q)/r} \left\{ \int_\Omega |f|^r \, d\mu \right\}^{q/r}$$

for all $n \in \mathbf{N}$ and all $0 < q < r$, from which it follows easily that

(3.39.5) $$R_\mu(f) = \limsup_{n\to\infty} \left\{ \frac{1}{(2n)!} \int_\Omega |f(\omega)|^{2n} \, \mu(d\omega) \right\}^{1/2n}$$

(3.39.6) If $\mu(\Omega) < \infty$ and $R_\mu(f) < \infty$, then $f \in L^{(\infty)}(\mu, \mathbf{C})$

Moreover, since $(n!)^{-1/n} \to 0$, we have

(3.39.7) $R_\mu(f) = 0$ if f is bounded and $\mu(\Omega) < \infty$

Suppose that $\mu(\Omega) < \infty$ and let $f \in L^{(\infty)}(\mu, \mathbf{C})$ and $r \ge 0$ be given. Then by Beppo Levi's theorem (3.7), we have

$$\int_\Omega e^{r|f|} \, d\mu = \sum_{n=0}^\infty \frac{r^n}{n!} \int_\Omega |f|^n \, d\mu$$

Note that the right-hand side is a power series in r with radius of convergence equal to $\rho_\mu(f)$. Hence, if $\mu(\Omega) < \infty$, $f \in L^{(\infty)}(\mu, \mathbf{C})$ and $r \ge 0$, then we have

(3.39.8) $$r < \rho_\mu(f) \quad \Rightarrow \quad \int_\Omega e^{r|f|} \, d\mu < \infty \quad \Rightarrow \quad r \le \rho_\mu(f)$$

(3.39.9) $$\int_\Omega e^{z\,f(\omega)}\,\mu(d\omega) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \int_\Omega f(\omega)^n\,\mu(d\omega) \quad \forall\, z \in \mathbf{C} \;:\; |z| < \rho_\mu(f)$$

3.40 The Uniqueness Theorem III

Let μ and ν be two finite measures on $(\Omega, \mathcal{F})$ such that $\mu(\Omega) = \nu(\Omega) < \infty$, and let $\mathcal{M}$ be a set of measurable functions from $(\Omega, \mathcal{F})$ into the complex plane $\mathbf{C}$ satisfying the following conditions:

(3.40.1) $$R_\mu(f) = \limsup_{n\to\infty} \left\{ \frac{1}{n!} \int_\Omega |f(\omega)|^n\,\mu(d\omega) \right\}^{1/n} < \infty \quad \forall\, f \in \mathcal{M}$$

(3.40.2) *If $f \in \mathcal{M}$, then $\bar{f} \in \mathcal{M}$*

(3.40.3) *If $f_1, \ldots, f_n \in \mathcal{M}$, then $\prod_{j=1}^{n} f_j \in \mathcal{U}$*

where $\mathcal{U}$ is the set of all functions $f \in L^1(\mu, \mathbf{C}) \cap L^1(\nu, \mathbf{C})$ satisfying

(3.40.4) $$\int_\Omega f d\mu = \int_\Omega f\, d\nu$$

Then $\mu(B) = \nu(B)$ for all $B \in \sigma(\mathcal{M})$.

Remarks: (1): If all functions in $\mathcal{M}$ are bounded, then (3.40.1) holds, and if all functions in $\mathcal{M}$ are real-valued, then (3.40.2) holds. In most applications we have that $\sigma(\mathcal{M}) = \mathcal{B}$; if so, then under the assumptions of the theorem above we have $\mu = \nu$.

(2): Suppose that (Ω, d) is a metric space and $\mathcal{F} = \mathcal{B}(\Omega)$. Let $\mathcal{M}$ be a set of complex-valued functions on Ω satisfying (3.40.1)–(3.40.3) together with the following two conditions:

(3.40.5) Every $f \in \mathcal{M}$ is continuous

(3.40.6) $\forall \omega' \neq \omega''\ \exists f \in \mathcal{M}$ so that $f(\omega') \neq f(\omega")$

If μ is a Radon measure and ν is τ-smooth, then by the uniqueness theorem (1.37), we have that $\mu = \nu$. Note that if (Ω, d) is separable, then every finite Borel measure ν is τ-smooth, and if (Ω, d) is complete and separable, then every finite Borel measure μ on S is a Radon measure.

(3): By Stirling's formula [see Exc.1.7.(xxiii)], we have

$$n! = \Theta_n \sqrt{2\pi}\, n^{n+1/2} e^{-n}$$

where $\exp\left(\frac{1}{12n+1}\right) < \Theta_n < \exp\left(\frac{1}{12n}\right)$. Hence, we find that (3.40.1) is equivalent to

$$\text{(3.40.7)} \qquad \limsup_{n\to\infty} \frac{1}{n}\left\{\int_\Omega |f|^n \, d\mu\right\}^{1/n} < \infty \qquad \forall f \in \mathcal{M}$$

It is well-known that $\sum_1^\infty \frac{1}{n} = \infty$, and it can be shown that if $\{a_n\}$ is any given sequence of non-negative numbers such that $\sum_1^\infty a_n = \infty$, then the conclusion of the uniqueness theorem III still holds if we replace (3.40.1) by

$$\text{(3.40.8)} \qquad \limsup_{n\to\infty} \left(a_n\left\{\int_\Omega |f|^n \, d\mu\right\}^{1/n}\right) < \infty \qquad \forall f \in \mathcal{M}$$

However, this is difficult, and the improvement is slight, so I shall not present the proof here.

Proof Note that $\mathcal{U}$ is a complex vector space, and let $\mathcal{L}$, $\mathcal{P}$ and $\mathcal{L}_0$ be the function spaces

$$\mathcal{L} = \left\{ c_0 + \sum_{j=1}^n c_j f_j \;\middle|\; c_0, \ldots, c_n \in \mathbf{C} \;,\; f_1, \ldots, f_n \in \mathcal{M} \right\}$$

$$\mathcal{P} = \left\{ c_0 + \sum_{i=1}^n c_i \prod_{j=1}^{m(i)} f_{ij} \;\middle|\; f_{ij} \in \mathcal{L} \quad \forall i,j \;,\; c_i \in \mathbf{C} \quad \forall i \right\}$$

$$\mathcal{L}_0 = \{ f \in \mathcal{L} \mid f(\omega) \in \mathbf{R} \;\; \forall \omega \in \Omega \}$$

Then $\mathcal{L}$ and $\mathcal{P}$ are complex vector spaces containing all constant functions, and since $|f|^{2n} = f^n \bar{f}^n$, then by (3.40.2), (3.39.3) and (3.39.5) we have that $\bar{f} \in \mathcal{L}$ and $R_\mu(f) = R_\nu(f) < \infty$ for all $f \in \mathcal{L}$. Since every function in $\mathcal{P}$ is a linear combination of a constant function and finite products of functions from $\mathcal{M}$, then by (3.40.3) and (3.39.6) we get

$$\text{(i)} \qquad \mathcal{M} \subseteq \mathcal{L} \subseteq \mathcal{P} \subseteq \mathcal{U} \cap L^{(\infty)}(\mu) \cap L^{(\infty)}(\nu)$$

Let $f \in \mathcal{L}_0$ be given and choose an integer $r \geq 1$ such that $\frac{1}{r} < \frac{1}{2}\rho_\mu(f)$. Set $f_0(\omega) = 1$, $g(\omega) = e^{|f(\omega)|/r}$ and

$$g_n(\omega) = \left(1 + \frac{if(\omega)}{rn}\right)^n \quad , \quad f_n(\omega) = \exp\left(\frac{inf(\omega)}{r}\right)$$

for $n \geq 1$. Then $g_n \to f_1$ and $|g_n| \leq g$. Moreover, by (3.39.8) we have that $g \in L^2(\mu) \cap L^2(\nu)$. Note that $f_{k+1} = f_1 \cdot f_k$. Hence, by (3.39.8) and Hölder's inequality (3.11) we get

$$\text{(ii)} \qquad |h \cdot g_n \cdot f_k| \leq |h| \cdot g \in L^1(\mu) \cap L^1(\nu) \qquad \forall h \in \mathcal{P} \; \forall k = 0, 1, 2, \ldots$$

(iii) $$\lim_{n\to\infty}(h\cdot g_n\cdot f_k) = h\cdot f_{k+1} \qquad \forall h\in\mathcal{P}\ \forall k=0,1,2\ldots$$

Let $h\in\mathcal{P}$ be given. Then by (i) we have that $h\,g_n\in\mathcal{P}\subseteq\mathcal{U}$. Thus, setting $k=0$ in (ii)+(iii) and using Lebesgue's theorem on dominated convergence we conclude that $h\,f_1\in\mathcal{U}$. Since $h\,g_n\in\mathcal{P}$ we have that $h\,g_n\,f_1\in\mathcal{U}$; hence, applying (ii)+(iii) with $k=1$ we see that $h\,f_2\in\mathcal{U}$. Continuing like this r times and setting $h(\omega)=1$, we see that $f_r=e^{if}\in\mathcal{U}$ for every function $f\in\mathcal{L}_0$. Since $\mathcal{U}$ is a complex vector space, we have that $\mathcal{A}\subseteq\mathcal{U}$ where

$$\mathcal{A}=\left\{\sum_{j=1}^{n}c_je^{if_j(\cdot)}\ \middle|\ c_1,\ldots,c_n\in\mathbf{C},\quad f_1,\ldots,f_n\in\mathcal{L}_0\right\}$$

Let $\mathcal{W}$ be the set of all non-negative functions in $\mathcal{A}$, and let $\mathcal{V}$ be the set of all non-negative measurable functions h satisfying

$$\int_\Omega h\,d\mu=\int_\Omega h\,d\nu$$

Then $\mathcal{V}$ satisfies (1.49.1)–(1.49.3) by linearity of the integral and the monotone convergence theorem (3.4). If $f,g\in\mathcal{W}$, then clearly $f\cdot g\in\mathcal{W}$, and if $g\in\mathcal{W}$ then g is bounded. Hence there exists $0<b<\infty$ so that $g(\omega)\le b$ for all $\omega\in\Omega$, but then $b-g(\cdot)\in\mathcal{W}$ and, thus, $\mathcal{W}$ satisfies (1.49.4) and (1.49.5).

Hence, by the functional Sierpińsky lemma (1.49) we have that $\mathcal{V}$ contains all non-negative $\mathcal{B}$-measurable functions, where $\mathcal{B}=\sigma(\mathcal{W})$. Let $f\in\mathcal{M}$ and let $g=\Re f=\frac{1}{2}(\bar{f}+f)$ and $h=\Im f=\frac{1}{2}(\bar{f}-f)$. Then $a\,g\in\mathcal{L}_0$ for all $a\in\mathbf{R}$ and

$$1+\sin(ag)=1+\tfrac{i}{2}(e^{-iag}-e^{iag})\in\mathcal{W}$$

Hence, $\sin(ag)$ is $\mathcal{B}$-measurable for all $a\in\mathbf{R}$, and since

$$g(\omega)=\lim_{n\to\infty}n\sin(\tfrac{g(\omega)}{n})$$

we see that g is $\mathcal{B}$-measurable. Similarly, we see that h is $\mathcal{B}$-measurable, and since $f=g+ih$ we conclude that f is $\mathcal{B}$-measurable for all $f\in\mathcal{M}$. Thus, $\sigma(\mathcal{M})\subseteq\mathcal{B}$, and since $\mu(B)=\nu(B)$ for all $B\in\mathcal{B}$, the theorem is proved. □

3.41 Measure Transforms

Let T be a set, let $(S,\mathcal{A})$ be a measurable space, and let $\Gamma:S\times T\to\mathbf{C}$ be a function such that $\Gamma(\cdot,t)$ is measurable for all $t\in T$. Suppose that $t\curvearrowright t^*$ is a function from T into T and that $(t_0,t_1)\curvearrowright t_0\nabla t_1$ is a function from $T\times T$ into T satisfying

(3.41.1) $$\Gamma(s,t^*)=\overline{\Gamma(s,t)}\qquad\forall s\in S\ \forall t\in T$$

(3.41.2) $$\Gamma(s,t_0\nabla t_1)=\Gamma(s,t_0)\Gamma(s,t_1)\qquad\forall s\in S\quad\forall t_0,t_1\in T$$

where $\bar{z}$ *is the complex conjugated of* z *whenever* $z \in \mathbf{C}$. *Let* μ *and* ν *be two measures on* $(S, \mathcal{A})$ *with* $\mu(S) = \nu(S) < \infty$, *and let* $T_0 \subseteq T_1 \subseteq T$ *such that*

$$t_0^* \in T_0 \quad \text{and} \quad t_0 \nabla t_1 \in T_1 \qquad \forall t_0 \in T_0 \qquad \forall t_1 \in T_1 \tag{3.41.3}$$

$$\limsup_{n\to\infty} \left\{ \frac{1}{n!} \int_S |\, \Gamma(s,t)\, |^n \; \mu(ds) \right\}^{1/n} < \infty \qquad \forall t \in T_0 \tag{3.41.4}$$

$$\int_S |\Gamma(s,t)|\, \mu(ds) < \infty \;\; \text{and} \;\; \int_S |\Gamma(s,t)|\, \nu(ds) < \infty \qquad \forall t \in T_1 \tag{3.41.5}$$

$$\int_S \Gamma(s,t)\, \mu(ds) = \int_S \Gamma(s,t)\, \nu(ds) \qquad \forall t \in T_1 \tag{3.41.6}$$

Then $\mu(H) = \nu(H)$ *for all* $H \in \mathcal{H}$ *where* $\mathcal{H} = \sigma(\Gamma(\cdot, t) \mid t \in T_0)$.

Remarks: (1): Let $(S, \mathcal{A})$ and $(T, \mathcal{B})$ be two measurable spaces, then a *kernel* on $S \times T$ is a function Γ from $S \times T$ into the complex plane such that $\Gamma(s, \cdot)$ is $\mathcal{B}$-measurable for all $s \in S$ and $\Gamma(\cdot, t)$ is $\mathcal{A}$-measurable for all $t \in T$. By (1.48) we have that an $\mathcal{A} \otimes \mathcal{B}$-measurable function $\Gamma : S \times T \to \mathbf{C}$ is a kernel and if so, we say that Γ is a *measurable kernel* on $S \times T$. If the kernel Γ satisfies (3.41.1+2) for some functions $t \curvearrowright t^*$ and $(t_0, t_1) \curvearrowright t_0 \nabla t_1$, then we say that Γ is a *multiplicative kernel.* Note that if Γ is real-valued, then (3.41.1) holds with $t^* = t$ for all $t \in T$.

(2): Let Γ be a kernel on $S \times T$ and let μ be a measure on $(S, \mathcal{A})$. Then we define *the* Γ*-transform*, denoted $\Gamma_\mu(t)$, and *the domain* of $\Gamma_\mu(t)$, denoted $\mathcal{D}(\Gamma_\mu)$, as follows:

$$\mathcal{D}(\Gamma_\mu) := \left\{ t \in T \;\middle|\; \int_S |\Gamma(s,t)|\, \mu(ds) < \infty \right\} \tag{3.41.7}$$

$$\Gamma_\mu(t) = \int_S \Gamma(s,t)\, \mu(ds) \qquad \forall t \in \mathcal{D}(\Gamma_\mu) \tag{3.41.8}$$

If Γ is a kernel on $S \times T$, we define *the transposed kernel* Γ^* on $T \times S$ by $\Gamma^*(t,s) := \Gamma(s,t)$ for all $(t,s) \in T \times S$. If ν is a measure on $(T, \mathcal{B})$, then we define the Γ^**-transform* $\Gamma_\nu^*(s)$ and its *domain* $\mathcal{D}(\Gamma_\nu^*)$ as above.

(3): Let Γ be a measurable kernel on $S \times T$ and let μ and ν be sum-finite measures on $(S, \mathcal{A})$ and $(T, \mathcal{B})$. Then by (3.20.3), (3.20.4) and Fubini's theorem (3.21.2) we have

$$\mathcal{D}(\Gamma_\mu) \in \mathcal{B} \;\text{ and }\; \Gamma_\mu \;\text{ is } \mathcal{B}\text{-measurable on } \mathcal{D}(\Gamma_\mu) \tag{3.41.9}$$

$$\mathcal{D}(\Gamma_\nu^*) \in \mathcal{A} \;\text{ and }\; \Gamma_\nu^* \;\text{ is } \mathcal{A}\text{-measurable on } \mathcal{D}(\Gamma_\nu) \tag{3.41.10}$$

$$(3.41.11) \qquad \int_S \Gamma_\nu^* \, d\mu = \int_T \Gamma_\mu \, d\nu = \int_{S\times T} \Gamma \, d(\mu \otimes \nu) \qquad \text{if } \Gamma \in L^1(\mu \otimes \nu)$$

Equation (3.41.11) is called *Parseval's equation.*

(4): Note that if $\Gamma(\cdot, t)$ is bounded for all $t \in T_0$, then (3.41.4) holds. Also observe that in the notation introduced above, conditions (3.41.5) and (3.41.6) takes the form

$$(3.41.12) \qquad T_1 \subseteq \mathcal{D}(\Gamma_\mu) \cap \mathcal{D}(\Gamma_\nu) \text{ and } \Gamma_\mu(t) = \Gamma_\nu(t) \qquad \forall t \in T_1$$

By (3.39.5) and (3.40.7) we have that (3.41.4) is equivalent to any of the following two conditions:

$$(3.41.13) \qquad \limsup_{n\to\infty} \left\{ \frac{1}{(2n)!} \int_S |\Gamma(s,t)|^{2n} \, \mu(ds) \right\}^{1/2n} < \infty \qquad \forall t \in T_0$$

$$(3.41.14) \qquad \limsup_{n\to\infty} \frac{1}{n} \left\{ \int_S |\Gamma(s,t)|^{n} \, \mu(ds) \right\}^{1/n} < \infty \qquad \forall t \in T_0$$

It can be shown [see Remark (3) to (3.40)] that the theorem above remains valid if we replace (3.41.4) by

$$(3.41.15) \qquad \limsup_{n\to\infty} a_n \left\{ \int_S |\Gamma(s,t)|^{n} \, \mu(ds) \right\}^{1/n} < \infty \qquad \forall t \in T_0$$

where $\{a_n\}$ is any given sequence of non-negative numbers with $\sum a_n = \infty$.

(5): In Chapter 4 we shall meet 10 kernels, all of which have significant applications to probability theory. Let me just mention the two most important kernels. Let $S = T = \mathbf{R}^k$ and let us define *the Fourier kernel*, denoted φ, and *the Laplace kernel*, denoted by L, by

$$\varphi(s,t) = \exp(i(s \bullet t)) \qquad \text{and} \qquad L(s,t) = \exp(s \bullet t)$$

where $s \bullet t = \sum s_j t_j$ denotes the inner product on $\mathbf{R}^k$. Then φ is a bounded multiplicative kernel with $t^* = -t$ and $t_0 \nabla t_1 = t_0 + t_1$, and the φ-transform of a Borel measure μ on $\mathbf{R}^k$ is called *the Fourier transform* of μ and is denoted $\hat{\mu}(t)$. Similarly, L is a multiplicative kernel with $t^* = t$ and $t_0 \nabla t_1 = t_0 + t_1$, and the L-transform of a Borel measure μ on $\mathbf{R}^k$ is called *the Laplace transform* of μ and is denoted $\tilde{\mu}(t)$.

(6): In Chapter 1 we saw that measures are complicated objects which rarely can be given in a simple explicit form. However, in many cases the Γ-transform can be computed for some multiplicative kernel Γ, such as the Fourier kernel, the Laplace kernel or some other explicit kernel. Note that the above theorem

gives a necessary condition for a finite measure to be uniquely determined by its Γ-transform. More precisely, let Γ be a kernel on $S \times T$, and let μ be a measure on $(S, \mathcal{A})$. If D is a subset of T, then we say that μ is Γ-*determined* on D if $D \subseteq \mathcal{D}(\Gamma_\mu)$ and for every measure ν on $(S, \mathcal{A})$, we have

$$(3.41.16) \qquad \left.\begin{array}{ll} \nu(S) = \mu(S) \ , & D \subseteq \mathcal{D}(\Gamma_\nu) \\ \Gamma_\mu(t) = \Gamma_\nu(t) & \forall t \in D \end{array}\right\} \quad \Rightarrow \quad \mu = \nu$$

Thus, if μ is Γ-determined on D, then the Γ-transform of μ on the set D describes the measure completely, and the theorem above gives a criterion for a finite measure to be Γ-determined on the set T_0.

Proof Let $\mathcal{M} = \{\Gamma(\cdot, t) \mid t \in T_0\}$, then by (3.41.1), (3.41.3) and (3.41.4) we see that $\mathcal{M}$ satisfies (3.40.1+2). Let $t_1, \ldots, t_n \in T_0$ be given and set $u_1 = t_1$ and $u_{j+1} = t_{j+1} \nabla u_j$ for $1 \leq j < n$. Then by (3.41.2+3) and induction in j we have that $u_j \in T_1$ and

$$\Gamma(s, u_n) = \prod_{i=1}^{n} \Gamma(s, t_i) \qquad \forall s \in S$$

Hence, by (3.41.3), (3.41.5) and (3.41.6) we have that $\mathcal{M}$ satisfies (3.40.3). Thus, the theorem follows from (3.40). □

3.42 Integrals of Vector and Matrix Functions

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $0 \leq q \leq \infty$. Then we let $L_k^q(\mu)$ denote the set of all measurable functions $f(\omega) = (f_1(\omega), \cdots, f_k(\omega))$ from Ω into $\mathbf{R}^k$ such that $f_j \in L^q(\mu)$ for $j = 1, \cdots, k$. Similarly, we let $L_{n\times k}^q(\mu)$ denote the set of all measurable functions $F(\omega) = \{F_{ij}(\omega)\}$ from Ω into $\mathbf{R}^{n\times k}$ such that $F_{ij} \in L^q(\mu)$ for $i = 1, \cdots, n$ and $j = 1, \cdots, k$. If $F(\omega) = \{F_{ij}(\omega)\}$ is a measurable function: $\Omega \to \mathbf{R}^{n\times k}$, then by (P.8.8) we have

$$(3.42.1) \qquad F \in L_{n\times k}^q(\mu) \qquad \Longleftrightarrow \qquad ||F(\cdot)|| \in L^q(\mu)$$

where $||F(\omega)||$ is the usual euclidean norm of $F(\omega)$ [see (P.8.6)]. Since vectors in $\mathbf{R}^k$ are considered as row vectors, that is, as $(1 \times k)$-matrices, we have

$$(3.42.2) \qquad L_k^q(\mu) = L_{1\times k}^q(\mu)$$

If $f(\omega) = (f_1(\omega), \cdots, f_k(\omega)) \in L^1_k(\mu)$, we define *the μ-integral of* f by

$$\int_\Omega f d\mu := \left(\int_\Omega f_1\, d\mu, \cdots, \int_\Omega f_k\, d\mu\right) \in \mathbf{R}^k \tag{3.42.3}$$

Similarly, if $F(\omega) = \{F_{ij}(\omega)\} \in L^1_{n\times k}(\mu)$, we define *the μ-integral* of F by

$$\int_\Omega F\, d\mu := \left\{\int_S F_{ij}\, d\mu\right\} \in \mathbf{R}^{n\times k} \tag{3.42.4}$$

Thus, $\int F d\mu$ is an $(n \times k)$-matrix.

Let $0 \le q \le \infty$ and let $f \in L^1_k(\mu)$, $F \in L^1_{n\times k}(\mu)$, $A \in \mathbf{R}^{k\times m}$, $B \in \mathbf{R}^{p\times n}$ and $v \in \mathbf{R}^k$ be given. Then as easy computation shows that

(3.42.5) $L^q_{n\times k}(\mu)$ is a vector space and $G \curvearrowright \int G\, d\mu$ is a linear function from $L^1_{n\times k}(\mu)$ into $\mathbf{R}^{n\times k}$

$$\int_\Omega (F(\omega)\cdot A)\, \mu(d\omega) = \left(\int_\Omega F\, d\mu\right)\cdot A \tag{3.42.6}$$

$$\int_\Omega (B\cdot F(\omega))\, \mu(d\omega) = B\cdot\left(\int_\Omega F d\mu\right) \tag{3.42.7}$$

$$\int_\Omega (v \bullet f(\omega))\, \mu(d\omega) = v \bullet \left(\int_\Omega f d\mu\right) \tag{3.42.8}$$

where $C\cdot D$ denotes the usual matrix product, and $x \bullet y$ denotes the usual inner product. Let $F = \{F_{ij}\} \in L^1_{n\times k}(\mu)$ and set $A = \{A_{ij}\} = \int F\, d\mu$, then by Cauchy-Schwartz' inequality we have

$$\begin{aligned} ||A||^2 = \sum_{ij} A_{ij}^2 &= \int_\Omega \Big(\sum_{ij} A_{ij}\, F_{ij}(\omega)\Big)\, \mu(d\omega) \\ &\le \int_\Omega \sqrt{\sum_{ij} A_{ij}^2}\, \sqrt{\sum F_{ij}(\omega)^2}\, \mu(d\omega) \\ &= ||A|| \int_\Omega ||F(\omega)||\, \mu(d\omega) \end{aligned}$$

Dividing through with $||A||$ we get

$$\left|\left|\int_\Omega F\, d\mu\right|\right| \le \int_\Omega ||F(\omega)||\, \mu(d\omega) \qquad \forall F \in L^1_{n\times k}(\mu) \tag{3.42.9}$$

Let $F = \{F_{ij}\}$ and $G = \{G_{ij}\}$ be measurable functions from Ω into $\mathbf{R}^{n\times k}$ and $\mathbf{R}^{k\times m}$, respectively. If $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, then by (P.8.7) and Hölder's inequality (3.11.1) we get

$$(3.42.10) \qquad \left\{\int_\Omega ||FG||^r \, d\mu\right\}^{1/r} \leq \left\{\int_\Omega ||F||^p d\mu\right\}^{1/p} \left\{\int_\Omega ||G||^q \, d\mu\right\}^{1/q}$$

In particular, if f and g are measurable functions from Ω into $\mathbf{R}^k$ and $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, then we have

$$(3.42.11) \qquad \left\{\int_\Omega |f \bullet g|^r \, d\mu\right\}^{1/r} \leq \left\{\int_\Omega ||f||^p \, d\mu\right\}^{1/p} \left\{\int_\Omega ||g||^q d\mu\right\}^{1/q}$$

3.43 References to Chapter 3

The integral in its modern form was invented by Henri Lebesgue in 1903 and his excellent book [5] (first ed. 1903, second ed. 1923) served for many years as the standard text book of integration. The text books [4] and [6] contains a wealth of information about the connection between integration and differentiation; for instance, the results described in (3.27.5), (3.27.11) and Exc.3.37. In Section 3.42, I have shown how to integrate functions taking values in a finite dimensional vector space. In [2] you will find an extension of the theory of integration to the case where either the function or the measure takes values in an infinite dimensional vector space. Moreover, in [2] it shown that uniform integrability is equivalent to compactness in a certain topology on $L^1(\mu)$, called the weak topology. In Sections 3.36 and 3.37, I have computed several integrals and introduced the gamma function together with some related functions. These functions belongs to class of *special functions*; that is, explicitly defined functions which cannot be expressed in terms of elementary functions, such as powers, trigonometric functions, exponential functions and logarithmic functions. In the handbooks [1] and [3] you will find properties of special functions together with tables of integrals and series which are very useful in the actual computations of explicit integrals.

[1] Abramowitz, M. and Segun, I. A. (1968): *Handbook of Mathematical Functions*, Dover Publications, Inc., New York

[2] Dunford, N. and Schwartz, J. T. (1958): *Linear Operators*, Interscience Publishers, Inc., New York

[3] Gradsheyn, I. S. and Rhyzhik, I. M. (1965): *Tables of Integrals, Series and Products*, (fourth ed.), Academic Press, New York and London

[4] Kestelman, H. (1959): *Modern Theories of Integration*, Dover Publications, Inc., New York

[5] Lebesgue, H. L. (1973): *Leçons sur l'Integration et la Recherche des Functions Primitive*, (reprint of the second edition from 1923) Chelsea Publishing Co., Bronx, New York

[6] Saks, S. (1937): *Theory of the Integral*, (second ed.), Hafner Publishing Co., New York

3.44 Exercises to Chapter 3

Exc. 3.1: Find the following two limits:

(i) $$\lim_{n\to\infty} \int_0^n \left(1 - \tfrac{x}{n}\right)^n e^{x/2}\, dx \quad , \quad \lim_{n\to\infty} \int_0^n \left(1 + \tfrac{x}{n}\right)^n e^{-2x}\, dx$$

Exc. 3.2: Let $I \subseteq \mathbf{R}$ be an interval, let $f : I \to \mathbf{R}$ be a Borel function, and let $a \in \operatorname{int}(I)$ such that f is continuous at a. Show that

(i) $$\lim_{\varepsilon\downarrow 0} \left\{ \tfrac{1}{2\varepsilon} \int_{a-\varepsilon}^{a+\varepsilon} |f(t) - f(a)|\, dt \right\} = 0$$

(ii) $$\lim_{\varepsilon\downarrow 0} \left\{ \tfrac{1}{2\varepsilon} \int_{a-\varepsilon}^{a+\varepsilon} f(t)\, dt \right\} = f(a)$$

Remark: A Borel function $f : I \to \mathbf{R}$ is said to be *approximately continuous* at a if $a \in \operatorname{int}(I)$ and (i) holds. Thus, the exercise shows that every continuity point of f is an approximate continuity point of f. The converse is far from being true; actually, it can be shown (difficult!) that we have

(iii) If $f : I \to \mathbf{R}$ is Lebesgue integrable on J for every compact interval $J \subseteq I$, then f is approximately continuous at Lebesgue a.a. $x \in \operatorname{int}(I)$

Exc. 3.3: Let $I \subseteq \mathbf{R}$ be an interval, and let $f, \varphi : I \to \mathbf{R}$ be a Borel function such that φ is λ-integrable on $[a, b]$ for all $a, b \in I$ with $a \le b$, where λ as usual denotes the Lebesgue measure on $\mathbf{R}$. Suppose that we have

$$f(b) - f(a) = \int_a^b \varphi(t)\, dt \quad \forall\, a, b \in I \quad \text{with} \quad a \le b$$

and let $a \in \operatorname{int}(I)$. Show that the following:

(i) f is continuous on I

(ii) If φ is continuous at a , then f is differentiable at a with $f'(a) = \varphi(a)$

(iii) If φ is differentiable at a , then f is twice differentiable at a with $f'(a) = \varphi(a)$ and $f''(a) = \varphi'(a)$

Remark: Recall that f is said to be *twice differentiable* at a with derivatives $f'(a)$ and $f''(a)$ if there exists a function $\rho : I \to \mathbf{R}$ such that $\rho(a) = \lim_{x \to a} \rho(x) = 0$ and

$$(*) \qquad f(x) = f(a) + (x-a)\, f'(a) + \tfrac{1}{2}\,(x-a)^2\, f''(a) + |x-a|^2\, \rho(x)$$

for all $x \in I$.

Hint: (iii): By assumption there exists a Borel function $\hat{\rho} : I \to \mathbf{R}$ such that $\hat{\rho}(a) = \lim_{x \to a} \hat{\rho}(x) = 0$ and

$$\varphi(t) = \varphi(a) + (t-a)\, \varphi'(a) + |t-a|\, \hat{\rho}(t) \qquad \forall\, t \in I$$

Then show that (*) holds with $f'(a) = \varphi(a)$, $f''(a) = \varphi'(a)$ and

$$\rho(x) = \begin{cases} (x-a)^{-2} \displaystyle\int_a^x |t-a|\, \hat{\rho}(t)\, dt & \text{if } x \neq a \\ 0 & \text{if } x = a \end{cases}$$

Finally, verify that $\rho(x) \to 0$ as $x \to a$.

Exc. 3.4: Let $(\Omega, \mathcal{F}, \mu)$ be measure space, let $\{F_n\} \subseteq \mathcal{F}$ be a sequence of measurable sets, and let $N(\omega) = \#\{j \in \mathbf{N} \mid \omega \in F_j\}$ for all $\omega \in \Omega$. Show that N is a measurable function from Ω into $\overline{\mathbf{R}}_+$ such that

(i) $$\int_\Omega N(\omega)\, \mu(d\omega) = \sum_{n=1}^{\infty} \mu(F_n)$$

(ii) $$\mu(N \geq k) \leq \frac{1}{k} \sum_{n=1}^{\infty} \mu(F_n) \qquad \forall\, 0 < k < \infty$$

Hint: (i): Note that $N = \sum 1_{F_n}$ and apply (3.8.2).

Exc. 3.5: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $\{f_n\}$ be a sequence of measurable functions from Ω into $\overline{\mathbf{R}}_+$ such that $f := \liminf_{n\to\infty} f_n \in L^1(\mu)$ and $\int_\Omega f_n \, d\mu \to \int_\Omega f \, d\mu$. Show that

(i) $$\lim_{n\to\infty} \int_F f_n(\omega)\, \mu(d\omega) = \int_F f(\omega)\, \mu(d\omega) \qquad \forall F \in \mathcal{F}$$

Hint: Apply Fatou's lemma (3.5) on the two sequences $\{1_F f_n\}$ and $\{1_{F^c} f_n\}$.

Exc. 3.6: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $f \in L^1(\mu)$, and let $\{f_n\}$ be a sequence of measurable functions such that $f_n \geq 0$ μ-a.s. for all $n \geq 1$, $f_n \to f$ μ-a.s., and $\int_\Omega f_n \, d\mu \to \int_\Omega f \, d\mu$. Show that

(i) $$\lim_{n\to\infty} \int_\Omega |f_n(\omega) - f(\omega)|\, \mu(d\omega) = 0$$

(ii) $$\lim_{n\to\infty} \int_F f_n(\omega)\, \mu(d\omega) = \int_F f(\omega)\, \mu(d\omega) \qquad \forall F \in \mathcal{F}$$

(iii) $$\lim_{n\to\infty} \int_\Omega f_n(\omega)\, g(\omega)\, \mu(d\omega) = \int_\Omega f(\omega)\, g(\omega)\, \mu(d\omega) \qquad \forall g \in L^\infty(\mu)$$

where $L^\infty(\mu)$ denotes the set of all measurable functions $g : \Omega \to \mathbf{R}$ which are bounded μ-a.s., that is, $|g(\omega)| \leq c$ μ-a.s. for some finite constant c [see (3.22)].

Hint: (i): Note that $(f - f_n)^+ \leq f$, and apply Lebesgue's dominated convergence theorem (3.6) to show that

$$\lim_{n\to\infty} \int_\Omega (f(\omega) - f_n(\omega))^+ \, \mu(d\omega) = 0$$

Note that $|f - f_n| = 2\,(f - f_n)^+ + (f_n - f)$ and $\int_\Omega (f_n - f)\, d\mu \to 0$ as $n \to \infty$.

Exc. 3.7: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\{f_n\}$ be a μ-uniformly integrable sequence of μ-integrable functions. Let $f, h \in L^1(\mu)$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a $(\cap f)$-stable paving such that

$$\lim_{n\to\infty} \int_G f_n(\omega)\, \mu(d\omega) = \int_G f(\omega)\, \mu(d\omega) \qquad \forall G \in \mathcal{G}$$

Show that

(i) $$\lim_{n\to\infty} \int_A f_n(\omega)\, \mu(d\omega) = \int_A f(\omega)\, \mu(d\omega) \qquad \forall A \in \sigma_0(\mathcal{G})$$

where $\sigma_0(\mathcal{G})$ denotes the smallest σ-ring containing $\mathcal{G}$ [see Exc.1.21].

Hint: Let $\mathcal{D}$ be the set of all $A \in \mathcal{F}$ satisfying (i) and let $\mathcal{H}$ be the smallest $(\sum c, -, \cap f)$-stable paving containing $\mathcal{G}$. Show that the paving $\mathcal{D}$ is $(\sum c, -)$-stable. Apply Sierpińsky's stability lemma (1.6.1) to show that $\mathcal{H} \subseteq \mathcal{D}$. Then apply (1.2.3+6) to show that $\mathcal{H}$ is a σ-ring.

Exc. 3.8: Let the intervals I_{jn} and the function $f_n : [0,1] \to \mathbf{R}$ be defined by

$$I_{jn} = [0,1] \cap \left[\tfrac{2j-1}{2n}, \tfrac{2j+1}{2n}\right] \qquad \forall j = 0, 1, \ldots, n$$

$$f_n(x) = \sum_{j=0}^{n} 4n\left(\tfrac{1}{2n} - |x - \tfrac{j}{n}|\right) 1_{I_{jn}}(x) \qquad \forall 0 \le x \le 1$$

Show that

(i) $$\lim_{n\to\infty} \int_0^x f_n(t)\,dt = x = \int_0^x 1\,dt \qquad \forall 0 \le x \le 1$$

(ii) $$\lim_{n\to\infty} \int_B f_n(t)\,dt = \lambda(B) = \int_B 1\,dt \qquad \forall B \in \mathcal{B}([0,1])$$

(iii) $$\int_0^1 |f_n(t) - 1|\,dt = \tfrac{1}{2} \qquad \forall n \ge 1$$

where λ denotes the Lebesgue measure on $\mathbf{R}$.

Hint: (i)+(iii): Make a decent drawing of the graph of f_n and give a geometric argument for (i) and (iii). (ii): Apply (i) and Exc.3.7 with $\mathcal{G} = \{[0,x] \mid 0 \le x \le 1\}$.

Exc. 3.9: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $||f||_q$ be the q-norm of f as defined in (3.22). Let $A, B \in \mathcal{F}$ and show that

(i) $$||1_A - 1_B||_0 = 1 \wedge \mu(A \triangle B)$$

(ii) $$||1_A - 1_B||_q = \mu(A \triangle B) \qquad \forall 0 < q \le 1$$

(iii) $$||1_A - 1_B||_q = \mu(A \triangle B)^{1/q} \qquad \forall 1 \le q < \infty$$

(iv) $$||1_A - 1_B||_\infty = \begin{cases} 1 & \text{if } \mu(A \triangle B) > 0 \\ 0 & \text{if } \mu(A \triangle B) = 0 \end{cases}$$

Thus, if $0 \le q < \infty$, then the $||\cdot||_q$-metric on the indicator functions is in essence the pseudo-metric d on $\mathcal{F}$ from Exc.1.9.

Exc. 3.10: Let $\mathcal{F}_0 = \{F \in \mathcal{F} \mid \mu(F) < \infty\}$ where $(\Omega, \mathcal{F}, \mu)$ is a given measure space. If $\mathcal{G} \subseteq \mathcal{F}$, we define

$$\mathcal{S}(\Omega, \mathcal{G}) := \Big\{ \sum_{j=1}^{n} a_j \, 1_{G_j} \;\Big|\; n \in \mathbf{N}\,,\; a_j \in \mathbf{R}\,,\; G_j \in \mathcal{G} \;\; \forall j = 1, \ldots, n \Big\}$$

Let $\sigma_0(\mathcal{G})$ denote the σ-ring generated by $\mathcal{G}$ [see Exc.1.21]. Show the following:

(i) $\mathcal{G} \subseteq \mathcal{F}_0 \quad \Rightarrow \quad \mathcal{S}(\Omega, \mathcal{G}) \subseteq L^q(\mu) \quad \forall 0 \le q \le \infty$

(ii) If $\mathcal{G}$ is $(\cap f)$-stable and $\mathcal{G} \subseteq \mathcal{F}_0 \subseteq \sigma_0(\mathcal{G})$, then $\mathcal{S}(\Omega, \mathcal{G})$ is dense in $(L^q(\mu), \|\cdot\|_q)$ for all $0 < q < \infty$

(iii) If μ is sum-finite and $\mathcal{F}$ is separable [see Exc.1.13], then $(L^q(\mu), \|\cdot\|_q)$ is a separable pseudo-metric space for all $0 < q < \infty$

Hint: (ii): First note that by $(\cap f)$-stability of $\mathcal{G}$ we have that $\mathcal{S}(\Omega, \mathcal{G})$ is a linear space such that $f \cdot g \in \mathcal{S}(\Omega, \mathcal{G})$ whenever $f, g \in \mathcal{S}(\Omega, \mathcal{G})$. Conclude that the paving $\mathcal{A} = \{A \subseteq \Omega \mid 1_A \in \mathcal{S}(\Omega, \mathcal{G})\}$ is a ring [see Exc.1.21]. Then apply Exc.1.47 to show that $\mathcal{A}$ is dense in $(\mathcal{F}_0, d)$, where d is the pseudo-metric defined in Exc.1.9. Apply Exc.3.9 to show that $\mathcal{S}(\Omega, \mathcal{G})$ is dense in $(\mathcal{S}(\Omega, \mathcal{F}_0), \|\cdot\|_q)$ for all $0 < q < \infty$. Finally, apply (1.46.1)–(1.46.4) and (3.25.1) to show that $\mathcal{S}(\Omega, \mathcal{F}_0)$ is dense in $(L^q(\mu), \|\cdot\|_q)$ for all $0 < q < \infty$. (iii): Note that if μ is a finite measure, then (iii) follows easily from (ii) and Exc.1.13. In the general case, let μ be sum-finite. Then by (3.16.5) and the Radon-Nikodym-Lebesgue theorem (3.19) there exist a finite measure ν and measurable function $\varphi : \Omega \to]0, \infty]$ such that

$$\int_\Omega f(\omega)\, \mu(d\omega) = \int_\Omega f(\omega)\, \varphi(\omega)\, \nu(d\omega)$$

for all measurable functions $f : \Omega \to \overline{\mathbf{R}}$ in the sense that if one the integrals exists, then so does the other and they are equal. Let $\mathcal{Q}$ be countable dense subset of $(L^q(\nu), \|\cdot\|_q)$. Then show that the set $\mathcal{Q}^q = \{\, g(\omega)\, \varphi(\omega)^{-1/q} \mid g \in \mathcal{Q} \,\}$ is a countable dense subset of $L^q(\mu)$.

Exc. 3.11: *(Egoroff's theorem)* Let $(\Omega, \mathcal{F}, \mu)$ be a finite measure space, and let $\{f_n\}$ be a sequence of measurable functions from Ω into $\mathbf{R}$ such that $f_n \to f$ μ-a.s. Let $\varepsilon > 0$ be given. Show that there exists a set $A \in \mathcal{F}$ such that $\mu(A) < \varepsilon$ and $f_n(\omega) \rightrightarrows f(\omega)$ uniformly on the set $\Omega \setminus A$.

Hint: Let $N \in \mathcal{F}$ be a μ-null set such that $f_n(\omega) \to f(\omega)$ for all $\omega \in \Omega \setminus N$. Set:

$$F_{kn} = \bigcup_{j=n}^{\infty} \{\, \omega \in \Omega \setminus N \;\big|\; |f_j(\omega) - f(\omega)| \ge 2^{-k} \,\}$$

Show that for $k \in \mathbf{N}$ fixed, then $F_{kn} \downarrow \emptyset$ as $n \uparrow \infty$ and conclude that there exists an integer $\sigma(k) \geq 1$ such that $\mu(F_{k\sigma(k)}) < \varepsilon\, 2^{-k}$. Set $A = N \cup \cup_{k=1}^{\infty} F_{k\sigma(k)}$ and show that this set works.

Exc. 3.12: *(Lusin's theorem)* Let (Ω, ρ) and (M, d) be metric spaces, let μ be a finite Borel measure on Ω , and let $f : \Omega \to M$ be a Borel function such that the image measure μ_f is τ-smooth [see the remark to (1.34)]. Show that for every $\varepsilon > 0$ there exists a closed set $F \subseteq \Omega$ such that $\mu(\Omega \setminus F) < \varepsilon$ and the restriction of f to the set F is continuous from F into M .

Hint: By assumption there exists a separable Borel set $M_0 \subseteq M$ such that $\mu(\Omega \setminus f^{-1}(M_0)) = 0$. Let $\{a_1, a_2, \ldots\}$ be a countable dense subset of M_0 and set $B_{nk} = f^{-1}(b(a_n, 2^{-k}))$ and

$$D_{1k} = B_{1k} \quad , \quad D_{nk} = B_{nk} \setminus \bigcup_{j=1}^{n-1} B_{jk} \qquad \forall n \geq 2\, \forall k \geq 1$$

$$A_{nk} = \bigcup_{j=1}^{n} D_{jk} = \bigcup_{j=1}^{n} B_{jk} \quad , \quad A_k = \bigcup_{n=1}^{\infty} D_{nk} = \bigcup_{n=1}^{\infty} B_{nk} \qquad \forall n, k \geq 1$$

see (1.2.11)–(1.2.13). Then $A_k \supseteq f^{-1}(M_0)$ and $A_{nk} \uparrow A_k$ whenever $n \uparrow \infty$ and k is fixed. Show that there exists an integer $\sigma(k) \geq 1$ such that $\mu(\Omega \setminus A_{\sigma(k),k}) < \varepsilon\, 2^{-k-1}$. Using (1.34.1), show that there exist closed sets $F_{jk} \subseteq D_{jk}$ such that $\mu(D_{jk} \setminus F_{jk}) < \varepsilon 2^{-k-j-1}$. Set

$$F = \bigcap_{k=1}^{\infty} \bigcup_{j=1}^{\sigma(k)} F_{jk}$$

and show that $\mu(\Omega \setminus F) < \varepsilon$ and that $\{F \cap F_{1k}, \ldots, F \cap F_{\sigma(k),k}\}$ is a disjoint partition of F . Hence, $f_k(\omega) := a_j$ if $\omega \in F \cap F_{jk}$ for some $1 \leq j \leq \sigma(k)$, is a well-defined function from F into M . Show that f_k is continuous on F and that $f_k \rightrightarrows f$ uniformly on F . Argue that this proves Lusin's theorem.

Exc. 3.13: Let $(\Omega, \mathcal{F})$ be a measurable space and let (S, d) be a metric space. Suppose that $I \subseteq \mathbf{R}$ is an interval, and let $f(\omega, t)$ be a function from $\Omega \times I$ into S such that $f(\cdot, t)$ is measurable for all $t \in I$ and $f(\omega, \cdot)$ is right continuous on I for all $\omega \in \Omega$. Show that f is $\mathcal{F} \otimes \mathcal{B}(I)$-measurable [cf. Exc.1.90].

Remark: There exists a non-measurable function $\varphi : \mathbf{R}^2 \to \mathbf{R}$ such that $\varphi(s, \cdot)$ and $\varphi(\cdot, t)$ are measurable (even upper semicontinuous) for all $s, t \in \mathbf{R}$.

Hint: Let $I_{jn} =](j-1)2^{-n}, j2^{-n}] \cap I$ and let b be the right end point of I . If $b \in I$, then set $c = b$; otherwise, let $c \in I$ be an arbitrary but fixed point

in I . Note that if $t \in I$ and $n \geq 1$ is an integer, then there exists a unique integer $j \in \mathbf{Z}$ with $t \in I_{jn}$. Hence, we may define

$$f_n(\omega, t) = \begin{cases} f(\omega, j2^{-n}) & \text{if } t \in I_{jn} \text{ and } j2^{-n} \in I \\ f(\omega, c) & \text{if } t \in I_{jn} \text{ and } j2^{-n} \notin I \end{cases}$$

Use the measurability of $f(\cdot, v)$ to show that f_n is $\mathcal{F} \otimes \mathcal{B}(I)$-measurable, and use the right continuity of $f(\omega, \cdot)$ to show that $f_n(\omega, t) \to f(\omega, t)$ for all $(\omega, t) \in \Omega \times I$.

Exc. 3.14: *(Riesz' convexity theorem I)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f : \Omega \to \mathbf{R}$ be a measurable function. Set

$$\mathcal{I} = \left\{ s \in \mathbf{R} \;\middle|\; \int_\Omega |f(\omega)|^s \, \mu(d\omega) < \infty \right\}$$

$$m(s) = \int_\Omega |f(\omega)|^s \, \mu(d\omega) \quad , \quad \vartheta(s) = \log m(s) \qquad \forall s \in \mathcal{I}$$

with the convention: $0^s = \infty$ if $s < 0$ and $0^0 = 1$. Show that $\mathcal{I}$ is an interval, and that m and ϑ are convex functions on $\mathcal{I}$.

Hint: Apply Hölder's inequality (3.11) to prove, that $\mathcal{I}$ is an interval and that ϑ is a convex function. Apply the equality: $m(s) = e^{\vartheta(s)}$ to prove convexity of m .

Exc. 3.15: *(Riesz' convexity theorem II)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f : \Omega \to \mathbf{R}$ be a measurable function. Define $m(s)$ and $\mathcal{I}$ as in Exc.3.14, and let $\mathcal{I}_+ = \mathcal{I} \cap]0, \infty[$ and $\mathcal{J} = \{t > 0 \mid \frac{1}{t} \in \mathcal{I}_+\}$. Let $s \in \operatorname{int}(\mathcal{I}_+)$ and let $r > 0$ be chosen such that $s \pm r \in \mathcal{I}_+$. Show that m is infinitely often differentiable at s and

(i) $$m^{(j)}(s) = \int_{\{|f|>0\}} |f(\omega)|^s \, (\log |f(\omega)|)^j \, \mu(d\omega) \quad \forall j \geq 1$$

(ii) $$m(t) = \sum_{n=0}^{\infty} \frac{(t-s)^n}{n!} \int_{\{|f|>0\}} |f(\omega)|^s \, (\log |f(\omega)|)^n \, \mu(d\omega) \quad \text{if } |t-s| \leq r$$

(iii) The function $\Lambda(t) = t \log m(\frac{1}{t})$ is a convex function on $\mathcal{J}$

Hint: (ii): Note that $m(s) = \int_A |f|^s \, d\mu$ where $A = \{|f| > 0\}$ for all $s \in (\mathcal{I}_+)$. Hence, it is no loss of generality to assume that $|f(\omega)| > 0$ for all $\omega \in \Omega$. Suppose that $s \pm r \in \mathcal{I}_+$ and that $|t - s| \leq r$. Then show that

$$|f(\omega)|^t = |f(\omega)|^s \sum_{n=0}^{\infty} \frac{(t-s)^n}{n!} (\log |f(\omega)|)^n$$

$$\sum_{n=0}^{\infty} \frac{r^n}{n!} \, |\log |f(\omega)| \, |^n = \exp\left(r \, |\log |f(\omega)| \, |\right) \leq |f(\omega)|^{-r} + |f(\omega)|^{+r}$$

and apply Beppo Levi's theorem (3.7). (i): Follows from (ii). (iii): Note that if $I \subseteq]0, \infty[$ is an interval, and $\varphi : I \to \mathbf{R}$ is convex, then so is $\psi(t) = t\,\varphi(\frac{1}{t})$.

Exc. 3.16: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $f : \Omega \to \mathbf{R}$ be a measurable function such that $f \in L^q(\mu)$ for some $0 < q < \infty$. Show that

(i) $$\lim_{s \downarrow 0} \left\{ \int_\Omega |f|^s d\mu \right\}^{1/s} = \begin{cases} \exp\left\{ \int_\Omega \log |f| d\mu \right\} & \text{if } \mu(|f| > 0) = 1 \\ 0 & \text{if } \mu(|f| > 0) < 1 \\ \infty & \text{if } \mu(|f| > 0) > 1 \end{cases}$$

(ii) $$\lim_{s \to \infty} \left\{ \int_\Omega |f(\omega)|^s \, \mu(d\omega) \right\}^{1/s} = \|f\|_\infty$$

Exc. 3.17: Let $\Gamma(\lambda)$ for $\lambda > 0$ be the gamma function [see (3.37)]. Apply Exc.3.15 and Exc.3.16 to prove (3.37.4) and

(i) $$\Gamma(\lambda) = \sum_{n=0}^{\infty} \frac{(\lambda - \alpha)^n}{n!} \int_0^\infty t^{\alpha - 1} e^{-t} (\log t)^n \, dt \qquad \forall 0 < \alpha < \lambda < 2\alpha$$

(ii) $$\lim_{\lambda \downarrow 0} \Gamma(\lambda)^{1/\lambda} = \lim_{\lambda \to \infty} \Gamma(\lambda)^{1/\lambda} = \infty$$

Exc. 3.18: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f \in L^1(\mu)$. Show that there exists a sequence $\{F_n\} \subseteq \mathcal{F}$ such that $\mu(F_n) < \infty$ and $F_n \uparrow \{f \neq 0\}$. [Set $F_n = \{|f| > 2^{-n}\}$].

Exc. 3.19: Let $(\Omega, \mathcal{F}, \mu)$ be a sum-finite measure space and let $f : \Omega \to \overline{\mathbf{R}}_+$ be a measurable function. Prove (3.30.3).

Exc. 3.20: *(Cavalieri's principle)* Let $(\Omega, \mathcal{F}, \mu)$ and $(\Lambda, \mathcal{G}, \nu)$ be two sum-finite measure spaces. Suppose that $H \in \mathcal{F} \otimes \mathcal{G}$ such that $\nu(H(\omega)) = 0$ for μ-a.a. $\omega \in \Omega$, where $H(\omega) = \{\lambda \in \Lambda \mid (\omega, \lambda) \in H\}$. Show that $\mu \otimes \nu(H) = 0$. Suppose that $\mathcal{G}$ is countably separating [see Exc.1.83 and Exc.1.89], and that ν is diffuse [see Exc.1.60]. Show that $\mu \otimes \nu(\Gamma(f)) = 0$ for every measurable function $f : \Omega \to \Lambda$, where $\Gamma(f) := \{(\omega, f(\omega)) \mid \omega \in \Omega\}$ is the graph of f.

Exc. 3.21: Consider the measure space $(\mathbf{N}, 2^{\mathbf{N}}, \#)$. By (3.38) we know that the $\#$-integral of f equals the infinite sum: $\sum_{n=1}^{\infty} f(n)$. Reformulate the four convergence theorems (monotone convergence, Fatou's lemma, Lebesgue's dominated convergence, and Beppo Levi's theorem) for sums, and compare them with the convergence theorems for sums that you have learned in your calculus course. Let $\ell^q := L^q(\#)$ and let $f, f_1, f_2, \ldots$ be a sequence of functions from $\mathbf{N}$ into $\mathbf{R}$ and show that

(i) $\quad f_k \to f \quad \#\text{-a.s.} \quad \Leftrightarrow \quad \lim_{k\to\infty} f_k(n) = f(n) \quad \forall n \in \mathbf{N}$

(ii) $\quad f_k \to f$ in $\#$-measure $\quad \Leftrightarrow \quad f_k \rightrightarrows f$ uniformly on $\mathbf{N}$

(iii) $$\|f\|_q = \begin{cases} \sum_{n=1}^{\infty} |f(n)|^q & \text{if } 0 < q \le 1 \\ \left\{ \sum_{n=1}^{\infty} |f(n)|^q \right\}^{1/q} & \text{if } 1 \le q < \infty \\ \sup_{n\ge 1} |f(n)| & \text{if } q = \infty \end{cases}$$

(iv) $\quad 1 \wedge \left\{ \sup_{n\ge 1} |f(n)| \right\} \le \|f\|_0 \le \sup_{n\ge 1} |f(n)|$

(v) $\quad \|f\|_0 \ge \limsup_{n\to\infty} |f(n)|$

(vi) $\quad \ell^\infty = \ell^0 = \{f : \mathbf{N} \to \mathbf{R} \mid f \text{ is bounded}\}$

(vii) $\quad \ell^q \subseteq \ell^p \subseteq \ell^\infty \qquad \forall 0 < q \le p \le \infty$

Let $\mathcal{V}$ be a set of functions from $\mathbf{N}$ into $\mathbf{R}$ and show that $\mathcal{V}$ is uniformly $\#$-integrable if and only if

(viii) $\quad \sup_{f\in\mathcal{V}} \left\{ \sum_{j=k}^{\infty} |f(j)| \right\} \to 0 \quad$ as $k \to \infty$

(ix) $\quad \sup_{f\in\mathcal{V}} \{ |f(n)| \} < \infty \qquad \forall n \in \mathbf{N}$

Reformulate Dunford-Pettis' theorem for sums. Finally, let $\#^2 = \# \otimes \#$ denote the product measure on $\mathbf{N}^2 = \mathbf{N} \times \mathbf{N}$. Show that $\#^2$ is the counting measure on $\mathbf{N}^2$ and conclude that we have

$$\text{(x)} \qquad \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{nk} = \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{nk} \quad \text{if } a_{nk} \geq 0 \text{ for all } n, k \geq 1$$

Moreover, show that the interchange of the order of summation in (x) remains valid if $a_{nk} \in \mathbf{R}$ and just one of the following four infinite sums are finite:

$$\sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{nk}^{+} \quad , \quad \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{nk}^{-} \quad , \quad \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{nk}^{+} \quad , \quad \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{nk}^{-}$$

Remark: Recall that if μ is a finite measure, then the $L^q(\mu)$-spaces satisfies the converse inclusions of (vii). A set $\mathcal{V}$ of functions from $\mathbf{N}$ into $\mathbf{R}$ satisfying (viii) and (ix) is said to be *uniformly summable.*

Exc. 3.22: Let $\mu = \sum_1^{\infty} a_n \mu_n$ where μ_n is a measure on $(\Omega, \mathcal{F})$ and $a_n \in \overline{\mathbf{R}}_+$ for all $n \geq 1$. Prove (3.15.3)

Exc. 3.23: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space such that $\mu(F) = 0$ or ∞ for all $F \in \mathcal{F}$. Show that $f \in L^1(\mu)$ if and only if $f = 0$ μ-a.s.

Exc. 3.24: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $(\Lambda, \mathcal{G})$ be a measurable space, and let $\xi(G, \omega)$ be a function from $\mathcal{G} \times \Omega$ into $[0, \infty]$ such that $\xi(\cdot, \omega)$ is a measure on $(\Lambda, \mathcal{G})$ for all $\omega \in \Omega$ and $\xi(G, \cdot)$ is $\mathcal{F}$-measurable for all $G \in \mathcal{G}$. Let

$$\nu(G) = \int_{\Omega} \xi(G, \omega)\, \mu(d\omega) \qquad \forall\, G \in \mathcal{G}$$

Show that ν is a measure on $(\Lambda, \mathcal{G})$ such that

$$\text{(i)} \qquad \int_{\Lambda} f(x)\, \nu(dx) = \int_{\Omega} \left\{ \int_{\Lambda} f(x)\, \xi(dx, \omega) \right\} \mu(d\omega)$$

for every measurable function $f : \Lambda \to [0, \infty]$. [Apply Beppo Levi's theorem (3.7) and the standard proof (3.14.1)–(3.14.4)].

Suppose that $\xi(G, \omega) = \sum_{n=1}^{\infty} \xi_n(G, \omega)$, where $\xi_n(\cdot, \omega)$ is a finite measure on $(\Lambda, \mathcal{G})$ for all $\omega \in \Omega$ and all $n \geq 1$, and $\xi_n(G, \cdot)$ is measurable for all $G \in \mathcal{G}$ and all $n \geq 1$. Show that

$$\vartheta(A) := \int_{\Omega} \left\{ \int_{\Lambda} 1_A(\omega, x)\, \xi(dx, \omega) \right\} \mu(d\omega) \qquad \forall\, A \in \mathcal{F} \otimes \mathcal{G}$$

is a measure on $(\Omega \times \Lambda, \mathcal{F} \otimes \mathcal{G})$ and that

(ii) $$\int_{\Omega\times\Lambda} g(\omega, x)\, \vartheta(d\omega, dx) = \int_{\Omega} \left\{ \int_{\Lambda} g(\omega, x)\, \xi(dx, \omega) \right\} \mu(dx)$$

whenever $g : \Omega \times \Lambda \to [0, \infty]$ is $\mathcal{F} \otimes \mathcal{G}$-measurable. [Mimic the proof of Tonelli's theorem; see (3.20).]

Exc. 3.25: Let $(\Omega, \mathcal{F}, \mu)$ be a sum-finite measure space and let $(M, \mathcal{M})$ be a measurable space. Let $0 < q < \infty$ be given, then $(L^q(\mu), ||\cdot||_q)$ is a complete pseudo-metric space [see (3.25.10)]. Show that $\Phi : M \to L^q(\mu)$ is Bochner measurable [see Exc.1.75] if and only if there exists a $\mathcal{M} \otimes \mathcal{F}$-measurable function $\varphi(x, \omega)$ from $M \times \Omega$ into $\mathbf{R}$ satisfying

(i) $\Phi(x) = \varphi(x, \cdot)$ μ-a.s.

(ii) $$\int_{\Omega} |\varphi(x, \omega)|^q\, \mu(d\omega) < \infty \quad \forall x \in M$$

Hint: Let $\varphi : M \times \Omega \to \mathbf{R}$ be a $\mathcal{M} \otimes \mathcal{F}$-measurable function satisfying (ii) and let $\Phi(x) = \varphi(x, \cdot)$. Apply Exc.1.10 on the sets $\{\varphi \le r\}$, where r is rational, to show that there exist a separable σ-algebra $\mathcal{F}_0 \subseteq \mathcal{F}$ [see Exc.1.13] such that φ is $\mathcal{M} \otimes \mathcal{F}_0$-measurable. Apply Exc.3.10 to show that $\Phi(M)$ is a separable subset of $L^q(\mu)$. Apply (3.20.3) to show that the function $\Phi_f(x) = ||\Phi(x) - f||_q$ is measurable from $(M, \mathcal{M})$ into $\mathbf{R}$ for every given $f \in L^q(\mu)$. Apply (1.9) with $\mathcal{H}$ as the paving of all open balls to show that Φ is measurable. Argue that this shows the "if" part. To show the "only if" part, apply (ii) in Exc.1.75, (3.25.3) and (1.40.4).

Exc. 3.26: Let $(\Omega, \mathcal{F}, \mu)$ be a sum-finite measure space, let $(M, \mathcal{M})$ be a measurable space, and let $f : M \times \Omega \to \mathbf{R}$ be a $\mathcal{M} \otimes \mathcal{F}$-measurable function satisfying

$$\int_{\Omega} |f(x, \omega)|^q\, \mu(d\omega) < \infty \quad \forall x \in M$$

for some $0 < q < \infty$. Show that there exist sets $N_x \in \mathcal{F}$ and $M_{jn} \in \mathcal{M}$, and functions $f_{jn} \in L^q(\mu)$ for all $n, j \in \mathbf{N}$ and all $x \in M$ satisfying [apply Exc.3.25, Exc.1.75 (ii) and (3.25.3)]

(i) $M = \bigcup_{j=1}^{\infty} M_{jn}$ and $M_{jn} \cap M_{in} = \emptyset \quad \forall 1 \le i \ne j < \infty$

(ii) $\mu(N_x) = 0 \quad \forall x \in M$

(iii) $f_n(x,\omega) := \sum_{j=1}^{\infty} 1_{M_{jn}}(x)\, f_{jn}(\omega) \underset{n\to\infty}{\longrightarrow} f(x,\omega) \quad \forall x \in M\ \forall \omega \in \Omega \setminus N_x$

(iv) $\int_\Omega |f(x,\omega) - f_n(x,\omega)|^q\, \mu(d\omega) \le 2^{-n} \qquad \forall x \in M\ \forall n = 1,2,\ldots$

Exc. 3.27: *(Riesz' convexity theorem III)* Let $(S,\mathcal{A},\mu)$ and $(T,\mathcal{B},\nu)$ be sum-finite measure spaces, and let $K(s,t)$ be a $\mathcal{A}\otimes\mathcal{B}$-measurable complex-valued function satisfying

$$\int_A \mu(ds) \int_B |K(s,t)|\, \nu(dt) < \infty \qquad \forall A \in \mathcal{A}_0\ \forall B \in \mathcal{B}_0$$

where $\mathcal{B}_0 = \{ B \in \mathcal{B} \mid \nu(B) < \infty\}$ and $\mathcal{A}_0 = \{ A \in \mathcal{A} \mid \mu(A) < \infty\}$. Let $\mathcal{D}_\mu$ and $\mathcal{D}_\nu$ denote the set of all μ-integrable respectively ν-integrable, simple, complex-valued functions, and let $\mathcal{V}_K$ denote the set of all pairs (f,g) satisfying

(*) $f : S \to \mathbf{C}$ is $\mathcal{A}$-measurable, and $g : T \to \mathbf{C}$ is $\mathcal{B}$-measurable

(**) $\int_S \mu(ds) \int_T |f(s)\, g(t)\, K(s,t)|\, \nu(dt) < \infty$

If $(f,g) \in \mathcal{V}_K$, then we define

$$\tilde{K}(f,g) := \int_S \mu(ds) \int_T f(s)\, g(t)\, K(s,t)\, \nu(dt)$$

Show that $\mathcal{D}_\mu \times \mathcal{D}_\nu \subseteq \mathcal{V}_K$, and let

$$M_K(p,q) := \sup\{ \tilde{K}(f,g) \mid f \in \mathcal{D}_\mu\ ,\ g \in \mathcal{D}_\nu\ ,\ \|f\|_p \le 1\ ,\ \|g\|_q \le 1 \}$$

for all $0 < p,q \le \infty$. Let $(f,g) \in \mathcal{V}_K$, and show that

(i) $|\tilde{K}(f,g)| \le M_K(p,q)\, (\|f\|_p)^{\frac{1}{p\wedge 1}}\, (\|g\|_q)^{\frac{1}{q\wedge 1}} \qquad \forall 0 < p,q \le \infty$

Remark: If $\mu\otimes\nu(\{K \ne 0\}) > 0$, then by methods from complex analysis it can be shown that we have

(ii) The function $\psi(\alpha,\beta) := \log M_K(\frac{1}{\alpha},\frac{1}{\beta})$ is a convex function from $[0,\infty[\times[0,\infty[$ into $]-\infty,\infty]$

This result is called *the Riesz-Thorin's convexity theorem.*

Exc. 3.28: *(The extended Minkowski inequality)* Let $(\Omega, \mathcal{F}, \mu)$ and $(\Gamma, \mathcal{G}, \nu)$ be two sum-finite measure spaces and let $f : \Omega \times \Gamma \to [0, \infty]$ be an $\mathcal{F} \otimes \mathcal{G}$-measurable function. Let $0 < s \leq r < \infty$ and show that

$$\text{(i)} \quad \begin{aligned} &\left\{ \int_\Omega \left\{ \int_\Gamma f(\omega, \gamma)^s \nu(d\gamma) \right\}^{r/s} \mu(d\omega) \right\}^{1/r} \\ &\qquad \leq \left\{ \int_\Gamma \left\{ \int_\Omega f(\omega, \gamma)^r \mu(d\omega) \right\}^{s/r} \nu(d\gamma) \right\}^{1/s} \end{aligned}$$

Hint: Let $p = \frac{r}{s}$. Note that if $r = s$, then (i) follows directly from Tonelli's theorem (3.21.1). So suppose that $r > s$, then $1 < p < \infty$; and if $q = \frac{p}{p-1}$, then $\frac{1}{p} + \frac{1}{q} = 1$. Set

$$G(\omega) = \int_\Gamma f(\omega, \gamma)^s \, \nu(d\gamma)$$

Apply Hölder's inequality (3.11) to show that

$$\int_\Omega G(\omega)^{p-1} f(\omega, \gamma)^s \mu(d\omega) \leq \left\{ \int_\Omega G(\omega)^p \mu(d\omega) \right\}^{\frac{1}{q}} \left\{ \int_\Omega f(\omega, \gamma)^r \mu(d\omega) \right\}^{\frac{1}{p}}$$

Note that

$$\int_\Omega G(\omega)^p \, \mu(d\omega) = \int_\Omega \mu(d\omega) \int_\Gamma G(\omega)^{p-1} f(\omega, \gamma)^s \, \nu(d\gamma)$$

and show that (i) follows from Tonelli's theorem (3.21.1) and the inequality above.

Exc. 3.29: *(Hardy-Littlewood-Polya's inequality I)* Let $f, g : [0, \infty[\to [0, \infty]$ be Borel functions and let $1 \leq p < \infty$. Show that

$$\text{(i)} \quad \begin{aligned} &\int_0^\infty x^{-p} \left\{ \int_0^\infty g\left(\frac{y}{x}\right) f(y) \, dy \right\}^p dx = \int_0^\infty \left\{ \int_0^\infty g(y) \, f(xy) \, dy \right\}^p dx \\ &\qquad \leq \left\{ \int_0^\infty g(y) y^{-\frac{1}{p}} \, dy \right\}^p \cdot \int_0^\infty f(y)^p \, dy \end{aligned}$$

Remark: When $g = 1_{[0,1]}$ and $1 < p < \infty$, the inequality above is called *Hardy's inequality.*

Hint: Apply Exc.3.28 with $s = 1$, $r = p$, $\mu = \nu =$ the Lebesgue measure, and $f(x, y) = g(y) \, f(x \cdot y)$.

Exc. 3.30: *(Hilbert-Schur-Hardy-Riesz' inequality)* Let $f : [0,\infty[\to [0,\infty]$ be a Borel function and let $1 < p < \infty$. Show that

(i) $$\int_0^\infty \left\{ \int_0^\infty \frac{f(y)}{x+y}\, dy \right\}^p dx \le \left(\frac{\pi}{\sin(\pi/p)} \right)^p \int_0^\infty f(y)^p\, dy$$

Hint: Apply Exc.3.29 with $f = f$ and $g(x) = \frac{1}{1+x}$ and note that

$$\int_0^\infty (1+x)^{-1} x^{-\alpha}\, dx = \int_0^1 t^{\alpha-1} (1-t)^{-\alpha}\, dt = \frac{\pi}{\sin(\alpha\pi)}$$

for all $0 < \alpha < 1$.

Exc. 3.31: *(Hardy-Littlewood-Polya's inequality II)* Let $f : [0,\infty[\to [0,\infty]$ be a Borel function and let $1 < p < \infty$. Show that [apply Exc.3.29 with $f = f$ and $g(x) = \frac{1}{1\vee x}$]

(i) $$\int_0^\infty \left\{ \int_0^\infty \frac{f(y)}{x\vee y}\, dy \right\}^p dx \le \left(\frac{p^2}{p-1} \right)^p \int_0^\infty f(y)^p\, dy$$

Exc. 3.32: Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f, g : \Omega \to \overline{\mathbf{R}}$ be measurable functions such that $f \le g$ μ-a.s. and either f or g is μ-integrable. Show that

(i) $f \in L(\mu)$ and $g \in L(\mu)$

(ii) $$\int_\Omega f(\omega)\, \mu(d\omega) = \int_\Omega g(\omega)\, \mu(d\omega) \quad \Leftrightarrow \quad f = g \ \ \mu\text{-a.s.}$$

Exc. 3.33: Let $I \subseteq \mathbf{R}$ be an interval, then a function $\varphi : I \to \mathbf{R}$ is said to be *convex* if

(*) $$\varphi(\alpha s + (1-\alpha)t) \le \alpha\,\varphi(s) + (1-\alpha)\,\varphi(t) \qquad \forall\, s,t \in I\ \forall\, 0 \le \alpha \le 1$$

Let a and b be the end points of the interval I , and let $a < u < b$. Then the right and left difference quotients at u are given by

$$\Delta_u^r(s) := \frac{\varphi(u+s) - \varphi(u)}{s} \quad , \quad \Delta_u^l(t) := \frac{\varphi(u) - \varphi(u-t)}{t}$$

for $0 < s < b-u$ and $0 < t < u-a$. Let $a < u < v < b$ be given and show that

(i) $\Delta_u^r(s)$ is increasing and $\Delta_u^l(t)$ is decreasing

(ii) $\Delta_u^l(t) \le \Delta_u^r(s) \quad \forall\, 0 < s < b-u\ \forall\, 0 < t < u-a$

(iii) $\Delta_u^r(s) \le \Delta_v^l(t) \qquad \forall\, s > 0\ \forall\, t > 0$ with $u + s < v - t$

Deduce that the one-sided derivatives

$$(**) \quad D_r\varphi(v) := \lim_{s\downarrow 0} \Delta_v^r(s) \quad , \quad D_l\varphi(u) := \lim_{t\downarrow 0} \Delta_u^l(t)$$

exist and satisfy

$$\text{(iv)} \quad -\infty < D_l\varphi(u) \le D_r\varphi(u) \le D_l\varphi(v) \le D_r\varphi(v) < \infty$$

Hint: Let $0 < s < r < b-u$ be given. Set $\alpha = \frac{r-s}{r}$. Show that $0 < \alpha < 1$ and $u + s = \alpha u + (1-\alpha)(u+r)$. Apply (*) to show that

$$\varphi(u+s) \le \tfrac{r-s}{r}\,\varphi(u) + \tfrac{s}{r}\,\varphi(u+r)$$

Moving a around the terms in this inequality, we obtain $\Delta_u^r(s) \le \Delta_u^r(r)$. The other statements follow in a similar manner.

Exc. 3.34: *(Jensen's inequality)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $g : \Omega \to [0,\infty]$ be a measurable function such that $0 < c < \infty$ where $c = \int_\Omega g\, d\mu$. Let $I \subseteq \mathbf{R}$ be an interval, let $f : \Omega \to I$ be measurable such that $f \cdot g \in L^1(\mu)$, and let

$$m = \frac{1}{c}\int_\Omega f(\omega)\, g(\omega)\, \mu(d\omega)$$

Suppose that $\varphi : I \to \mathbf{R}$ is a convex function and show that

$$\text{(i)} \quad m \in I \quad \text{and} \quad \varphi(m) \le \frac{1}{c}\int_\Omega \varphi(f(\omega))\, g(\omega)\, \mu(d\omega)$$

Observe that Jensen's inequality (3.10) follows from the result above by setting $g = 1_F$. Let us define $L_m = R_m = \{m\}$ if m is one of the end points of I , and define

$$L_m = \{x \in I \mid \varphi(x) = \alpha \cdot (x-m) + \varphi(m)\}$$

$$R_m = \{x \in I \mid \varphi(x) = \beta \cdot (x-m) + \varphi(m)\}$$

if $m \in \operatorname{int}(I)$ where α and β are the left-hand and the right-hand derivatives of φ at m [see the previous exercise]. Show that

$$\text{(ii)} \quad \varphi(m) = \frac{1}{c}\int_\Omega \varphi(f(\omega))\, g(\omega)\, \mu(d\omega) \quad \Leftrightarrow \quad \mu(g > 0\,,\ f \notin L_m \cap R_m) = 0$$

Remark: A convex function $\varphi : I \to \mathbf{R}$ is said to be *strictly convex* at $m \in I$ if and only if $L_m \cap R_m = \{m\}$. Thus, if φ is strictly convex at m , then we have equality in Jensen's inequality if and only if $f(\omega) = m$ for μ-a.a. $\omega \in \{g > 0\}$.

Hint: (i): Apply Jensen's inequality (3.10) with μ replaced by the measure $\nu(A) = \int_A g\, d\mu$. (ii): Examine the proof of Jensen's inequality (3.10) and apply Exc.3.32.

Exc. 3.35: *(Hölder's inequality)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let $f, g : \Omega \to \overline{\mathbf{R}}$ be measurable functions, and let $0 < p, q < \infty$ such that

$$\alpha = \int_\Omega |f(\omega)|^p \, \mu(d\omega) < \infty \quad , \quad \beta = \int_\Omega |g(\omega)|^q \, \mu(d\omega) < \infty$$

Let $0 < r < \infty$ be defined by $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, and show that we have equality in Hölder's inequality (3.11.1); that is,

$$\text{(i)} \qquad \left\{ \int_\Omega |f \cdot g|^r d\mu \right\}^{1/r} = \left\{ \int_\Omega |f|^p d\mu \right\}^{1/p} \cdot \left\{ \int_\Omega |g|^q d\mu \right\}^{1/q}$$

if and only if $\beta\, |f(\omega)|^p = \alpha\, |g(\omega)|^q$ μ-a.s.

Hint: First note that the result is true if either $\alpha = 0$ or $\beta = 0$. So suppose that $0 < \alpha, \beta < \infty$, and examine the proof of Hölder's inequality noting that $x^t y^{1-t} < tx + (1-t)y$ whenever $x, y \geq 0$, $x \neq y$ and $0 < t < 1$.

Exc. 3.36: *(Minkowski's inequality)* Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f, g \in L^q(\mu)$ where $0 < q < \infty$. Show the following:

(i) If $0 < q < 1$, then $||f+g||_q = ||f||_q + ||g||_q$ if and only if $f(\omega)\, g(\omega) = 0$ μ-a.s.

(ii) If $q = 1$, then $||f+g||_q = ||f||_q + ||g||_q$ if and only if $f(\omega)\, g(\omega) \geq 0$ μ-a.s.

(iii) If $1 < q < \infty$, then $||f+g||_q = ||f||_q + ||g||_q$ if and only if there exists numbers $a \geq 0$ and $b \geq 0$ satisfying $a + b > 0$ and $a\, f(\omega) = b\, g(\omega)$ μ-a.s.

Hint: Examine the proof of Minkovski's inequality (3.12), and apply Exc.3.35 for the case that $1 < q < \infty$.

Exc. 3.37: Let $I \subseteq \mathbf{R}$ be an interval, and suppose that $g : I \to \mathbf{R}$ is continuously differentiable. Show that g is absolutely continuous with density $g'(x)$ [see (3.27)]. Let $f : I \to \mathbf{R}$ be an absolutely continuous function. Let $a, b \in I$ such that $a < b$. Show that for every $\varepsilon > 0$, there exists $\delta > 0$ such that

(i) $\sum_{j=1}^{n} |f(b_j) - f(a_j)| < \varepsilon$ whenever $a_1, b_1, \ldots, a_n, b_n \in [a, b]$ satisfies: $a_1 \le b_1 \le a_2 \le b_2 \le \cdots \le a_n \le b_n$ and $\sum_{j=1}^{n} (b_j - a_j) < \delta$

Remark: Let $f : [a, b] \to \mathbf{R}$ be a given function, and let $D_r^* f(x)$ and $D_r^\circ f(x)$ be *the right upper* respectively *the right lower derivative* of f, that is,

$$D_r^* f(x) = \limsup_{y \downarrow x} \frac{f(y) - f(x)}{y - x} \quad , \quad D_r^\circ f(x) = \liminf_{y \downarrow x} \frac{f(y) - f(x)}{y - x}$$

whenever $x \in [a, b[$. If $D_r^* f(x) = D_r^\circ(x) \neq \pm\infty$, then we say that f is *right differentiable* at x with *right derivative* $D_r f(x) := D_r^* f(x) = D_r^\circ(x)$. Suppose that $f : [a, b] \to \mathbf{R}$ is continuous. Then it can be shown (difficult!) that we have the following:

(ii) f is absolutely continuous on $[a, b]$ if and only if f satisfies (i); and if so, then f is differentiable at x for Lebesgue a.a. $x \in]a, b[$, and the density of f equals $f'(x)$ for Lebesgue a.a. $x \in]a, b[$

(iii) If $D_r^* f(x)^+$ is Lebesgue integrable on $[a, b[$ and $D_r^* f(x) < \infty$ for all $x \in [a, b[\setminus T$ for some finite or countable set $T \subseteq [a, b[$, then we have

$$f(\beta) - f(\alpha) \le \int_\alpha^\beta D_r^* f(x)\, dx \qquad \forall\, a \le \alpha \le \beta \le b$$

(iv) If $D_r^\circ f(x)^-$ is Lebesgue integrable on $[a, b[$ and $D_r^\circ f(x) > -\infty$ for all $x \in [a, b[\setminus T$ for some finite or countable set $T \subseteq [a, b[$, then we

$$f(\beta) - f(\alpha) \ge \int_\alpha^\beta D_r^\circ f(x)\, dx \qquad \forall\, a \le \alpha \le \beta \le b$$

(v) If f is right differentiable at all $x \in [a, b[\setminus T$ for some finite or countable set $T \subseteq [a, b[$, and if either $D_r f(x)^+$ or $D_r f(x)^-$ is Lebesgue integrable on $[a, b[$, then f is absolutely continuous on $[a, b]$ with density $D_r f(x)$ for Lebesgue a.a. $x \in [a, b[$

Exc. 3.38: Let $f(x)$ be defined by

$$f(x) = \begin{cases} x^2 \sin\left(\frac{\pi}{x^2}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Show that f is differentiable everywhere on $\mathbf{R}$, but f is **not** absolute continuous on $[0,1]$. Why doesn't this contradict (v) in Exc.3.37 ?

Exc. 3.39: Let $\varphi : [0,1] \to [0,1]$ and $\psi : [0,1] \to [0,1]$ be continuous functions satisfying

$$\int_0^1 \varphi(x)^n \, \psi(x)^m \, dx = \frac{1}{n+1} \cdot \frac{1}{m+1} \qquad \forall n, m \in \mathbf{N}_0$$

Show that $\gamma(x) = (\varphi(x), \psi(x))$ is a *Peano curve*; that is, show that γ is continuous and $\gamma([0,1]) = K$, where $K = [0,1] \times [0,1]$ is the unit square.

Remark: A Peano curve is a continuous curve defined on the unit interval which passes through all points in the unit square. The construction of such a curve was first established by Guiseppe Peano (1858–1932) in 1890.

Hint: Let λ be the Lebesgue measure on the unit interval, and let λ_2 be the Lebesgue measure on the unit square. Note that γ is a continuous function from $[0,1]$ into K and let $\nu := \lambda_\gamma$ be the image measure of λ under the function γ. Use the first transformation theorem (3.15) to show that

$$\int_K x^n \, y^m \, \nu(dx, dy) = \int_K x^n \, y^m \, \lambda_2(dx, dy) \qquad \forall n, m \in \mathbf{N}_0$$

Then apply (3.40) with $\mathcal{M} = \{\wp_1, \wp_2\}$, where $\wp_1(x,y) = x$ and $\wp_2(x,y) = y$ are the two projections, to show that $\nu = \lambda_2$. Let $C = \gamma([0,1])$ and show that C is a compact subset of the unit square K with $\lambda_2(C) = 1$. Finally, show that this implies that $C = K$.

Exc. 3.40: Let μ_x be a finite Borel measure on $[0,\infty[$ for all $0 \leq x < \infty$, and let $\Phi : [0,\infty[\to [0,\infty[$ be a given function satisfying

$$\lim_{x\to\infty} \left\{ \frac{\mu_x([0,\infty[)}{\Phi(x)} \right\} = 1 \quad , \quad \lim_{x\to\infty} \left\{ \frac{\mu_x([0,a])}{\Phi(x)} \right\} = 0 \qquad \forall\, 0 < a < \infty$$

Let $f : [0,\infty[\to \mathbf{R}$ be a bounded Borel function. Show that

(i) $$\liminf_{t\to\infty} f(t) \leq \liminf_{t\to\infty} \left\{ \frac{1}{\Phi(t)} \int_{[0,\infty[} f(y) \, \mu_t(dy) \right\}$$

(ii) $$\limsup_{t\to\infty} f(t) \geq \limsup_{t\to\infty} \left\{ \frac{1}{\Phi(t)} \int_{[0,\infty[} f(y) \, \mu_t(dy) \right\}$$

Exc. 3.41: Let $\varphi : [0,\infty[\to [0,\infty[$ be a measurable function satisfying

$$\int_0^\infty \varphi(x)\,dx = \infty \quad \text{but} \quad \Phi(t) = \int_0^t \varphi(x)\,dx < \infty \quad \forall\, 0 \le t < \infty$$

Let $f : [0,\infty[\to \mathbf{R}$ be a bounded Borel function. Show that [apply Exc.3.40]

(i) $$\liminf_{t\to\infty} f(t) \le \liminf_{t\to\infty} \left\{ \frac{1}{\Phi(t)} \int_0^t \varphi(x)\, f(x)\,dx \right\}$$

(ii) $$\limsup_{t\to\infty} f(t) \ge \limsup_{t\to\infty} \left\{ \frac{1}{\Phi(t)} \int_0^t \varphi(x)\, f(x)\,dx \right\}$$

Exc. 3.42: Let $\varphi : [0,\infty[\to [0,\infty[$ be a monotone function satisfying

$$\lim_{t\to\infty} \frac{\varphi(t)}{\Phi(t)} = 0 \quad \text{where} \quad \Phi(t) = \int_0^t \varphi(x)\,dx \quad \forall\, 0 \le t < \infty$$

Let $f : [0,\infty[\to \mathbf{R}$ be a bounded Borel function. Show that [apply Exc.3.40]

(i) $$\liminf_{t\to\infty} f(t) \le \liminf_{t\to\infty} \left\{ \frac{1}{\Phi(t)} \int_0^t f(x)\, \varphi(t-x)\,dx \right\}$$

(ii) $$\limsup_{t\to\infty} f(t) \ge \limsup_{t\to\infty} \left\{ \frac{1}{\Phi(t)} \int_0^t f(x)\, \varphi(t-x)\,dx \right\}$$

Exc. 3.43: Let μ be a Borel measure on $[0,\infty[$ such that $\Phi(x) = \mu([0,x]) < \infty$ for all $0 \le x < \infty$ and $\mu([0,\infty[) = \infty$. Let $f : [0,\infty[\to \mathbf{R}$ be a bounded function. Show that [apply Exc.3.40]

(i) $$\liminf_{t\to\infty} f(t) \le \liminf_{t\to\infty} \left\{ \frac{1}{\Phi(x)} \int_{[0,x]} f(x)\,\mu(dx) \right\}$$

(ii) $$\limsup_{t\to\infty} f(t) \ge \limsup_{t\to\infty} \left\{ \frac{1}{\Phi(x)} \int_{[0,x]} f(x)\,\mu(dx) \right\}$$

Exc. 3.44: Let μ be a Borel measure on $[0,\infty[$, and let $\beta \in \mathbf{R}$ be a number such that $L_\mu(\beta) = \infty$ and $L_\mu(\alpha) < \infty$ for all $\alpha > \beta$ where

$$L_\mu(\alpha) = \int_0^\infty e^{-\alpha x}\,\mu(dx) \quad \forall\, \alpha \in \mathbf{R}$$

Let $f : [0,\infty[\to [0,\infty[$ be a bounded Borel function. Show that the integral

$$L_f(\alpha) = \int_0^\infty f(x)\, e^{-\alpha x}\,\mu(dx)$$

exists and is finite for all $\alpha > \beta$. Moreover, show that [apply Exc.3.40]

(i) $\liminf_{t\to\infty} f(t) \le \liminf_{\alpha\downarrow\beta} \left\{ \frac{L_f(\alpha)}{L_\mu(\alpha)} \right\}$

(ii) $\limsup_{t\to\infty} f(t) \ge \limsup_{\alpha\downarrow\beta} \left\{ \frac{L_f(\alpha)}{L_\mu(\alpha)} \right\}$

Exc. 3.45: Let $f : [0,\infty[\to [0,\infty[$ be a bounded Borel function and let

$$a = \liminf_{t\to\infty} f(t) \quad , \quad b = \limsup_{t\to\infty} f(t)$$

Let $0 < \alpha < \infty$ be given and show that [apply Exc.3.41–3.43]:

(i) $a \le \liminf_{x\to\infty} \int_0^x \frac{f(y)}{x}\, dy \le \limsup_{x\to\infty} \int_0^x \frac{f(y)}{x}\, dy \le b$

(ii) $a \le \liminf_{x\to\infty} \int_0^x \frac{\alpha(x-y)^{\alpha-1}}{x^\alpha} f(y)\, dy \le \limsup_{x\to\infty} \int_0^x \frac{\alpha(x-y)^{\alpha-1}}{x^\alpha} f(y)\, dy \le b$

(iii) $a \le \liminf_{x\downarrow 0} \int_0^\infty x\, e^{-xy} f(y)\, dy \le \limsup_{x\downarrow 0} \int_0^\infty x\, e^{-xy} f(y)\, dy \le b$

(iv) $\frac{a}{2} \le \liminf_{x\downarrow 0} \int_0^\infty \frac{x f(y) e^{-(xy)^2}}{\sqrt{\pi}}\, dy \le \limsup_{x\downarrow 0} \int_0^\infty \frac{x f(y) e^{-(xy)^2}}{\sqrt{\pi}}\, dy \le \frac{b}{2}$

(v) $\frac{a}{2} \le \liminf_{x\to 0} \int_0^\infty \frac{\sin^2(xy) f(y)}{\pi\, x\, y^2}\, dy \le \limsup_{x\to 0} \int_0^\infty \frac{\sin^2(xy) f(y)}{\pi\, x\, y^2}\, dy \le \frac{b}{2}$

Exc. 3.46: Let λ be the Lebesgue measure on $\mathbf{R}$ and let $0 < q < \infty$. Let $C_{00}(\mathbf{R})$ be the set of all continuous functions $f : \mathbf{R} \to \mathbf{R}$ such that the set $\{f \neq 0\}$ is bounded. Let $\mathcal{S}(\mathbf{R})$ be the set of all *step functions*, that is, the set of all functions $f : \mathbf{R} \to \mathbf{R}$ of the form

$$f(x) = \sum_{j=1}^{n} c_j\, 1_{]a_j,b_j]}(x)$$

for some $a_j, b_j, c_j \in \mathbf{R}$ with $a_j < b_j$ for all $j = 1, \ldots, n$. Show that

(i) $C_{00}(\mathbf{R})$ is dense in $(L^q(\lambda), ||\cdot||_q)$

(ii) $\mathcal{S}(\mathbf{R})$ is dense in $(L^q(\lambda), ||\cdot||_q)$

Hint: (ii) follows from (iii) in Exc.3.10, and (i) follows from (ii).

Exc. 3.47: *(Riemann-Lebesgue's lemma)* Let $\Phi : \mathbf{R}^2 \to \mathbf{C}$ be a bounded complex-valued Borel function satisfying

$$(*) \qquad \lim_{x\to\infty} \int_a^b \Phi(x,y)\,dy = 0 \quad \forall\, a,b \in \mathbf{R} \text{ with } a < b$$

Let $B \subseteq \mathbf{R}$ be a Borel set and let $L^1(B)$ be the set of all Lebesgue integrable functions $f : B \to \mathbf{R}$. Show that

$$\text{(i)} \qquad \lim_{x\to\infty} \int_B \Phi(x,y)\,f(y)\,dy = 0 \qquad \forall\, f \in L^1(B)$$

$$\text{(ii)} \qquad \lim_{x\to\infty} \int_{-\infty}^{\infty} e^{ixy}\,f(y)\,dy = 0 \qquad \forall\, f \in L^1(\mathbf{R})$$

$$\text{(iii)} \qquad \lim_{n\to\infty} \int_{-\pi}^{\pi} \sin(n\,y)\,f(y)\,dy = 0 \qquad \forall\, f \in L^1([-\pi,\pi])$$

$$\text{(iv)} \qquad \lim_{n\to\infty} \int_{-\pi}^{\pi} \cos(n\,y)\,f(y)\,dy = 0 \qquad \forall\, f \in L^1([-\pi,\pi])$$

Hint: (i): Let $0 \le C < \infty$ be chosen such that $|\Phi(x,y)| \le C$ for all $(x,y) \in \mathbf{R}^2$ and let $\mathcal{S}(\mathbf{R})$ be defined as in Exc.3.46. Show that

$$\lim_{x\to\infty} \int_{-\infty}^{\infty} \Phi(x,y)\,g(y)\,dy = 0 \qquad \forall\, g \in \mathcal{S}(\mathbf{R})$$

Let $f \in L^1(B)$ and $\varepsilon > 0$ be given. Set $f_B(x) = f(x)$ if $x \in B$, and $f_B(x) = 0$ if $x \in \mathbf{R} \setminus B$. By (ii) in Exc.3.46, there exists $g \in \mathcal{S}(\mathbf{R})$ such that $||f_B - g||_1 < \frac{\varepsilon}{C}$, and note that

$$\left| \int_{-\infty}^{\infty} \Phi(x,y)\,f_B(y)\,dy - \int_{-\infty}^{\infty} \Phi(x,y)\,g(y)\,dy \right| < \varepsilon$$

(ii)–(iv): Note that the functions: e^{ixy}, $\sin(xy)$ and $\cos(xy)$ are bounded Borel functions satisfying (*).

Exc. 3.48: Let $(\Omega, \mathcal{F}, \mu)$ be a sum-finite measure space, and let $\varphi : \Omega \to M$ and $f : \Omega :\to [0,\infty]$ be measurable functions where $(M, \mathcal{B})$ is a measurable space. Show that the image measure $\mu_\varphi(B) = \mu(\varphi^{-1}(B))$ is sum-finite and that the indefinite integral

$$\nu(F) = \int_F f(\omega)\,\mu(d\omega) \qquad \forall\, F \in \mathcal{F}$$

is a sum-finite measure. Moreover, if $\mu_1, \mu_2, \ldots$ is a sequence of sum-finite measures on $(\Omega, \mathcal{F})$, then show that their sum $\sigma(F) = \sum_1^\infty \mu_n(F)$ is a sum-finite measure.

Chapter 4

EXPECTATIONS AND MOMENTS

4.1 Expectations

From now on, $(\Omega, \mathcal{F}, P)$ will denote a probability space. If X is a random variable, for instance, if X is the outcome of a random experiment, then the expectation EX should be that real value which we expect before the experiment has been performed; that is, EX should be that real number which in some sense is as close as possible to the many values that our random variable may take. The first textbook on probability theory (from 1657) was written by Christaan Huyghens (1629-1695) entitled *De Ratiociniis in Ludo Alea* which means "How to reason in the game of dice." In his book, Huyghens take the expectation as the fundamental concept of probability, and probabilities are only expressed in terms of odds. For instance, "To have p chances of winning a and q chances of winning b," means in modern terminology that the probability of winning a equals $\frac{p}{p+q}$ and the probability of winning b equals $\frac{q}{p+q}$. In his three first theorems, Huyghens defines the expectation (in his terminology: "The value of my chance") as follows:

> ***Theorem I:*** If I have even chances of winning a and b, then the value of my chance is $\frac{1}{2}(a+b)$
>
> ***Theorem II:*** If we have even chances of winning a, b and c, then the value of my chance is $\frac{1}{3}(a+b+c)$
>
> ***Theorem III:*** If I have p chances of winning a and q chances of winning b, then the value of my chance equals $\frac{pa+qb}{p+q}$

Thus, Huyghens maintains that the expectation of X is a weighted average of the possible values of X, where we use the probabilities as weights. That is, exactly what the integral $\int_\Omega X(\omega)\,P(d\omega)$ is supposed to do.

Formally speaking, we define *the expectation*, also called *the mean* and denoted EX, of a random variable X to be the P-integral of X, that is,

$$EX := \int_\Omega X \; dP \tag{4.1.1}$$

provided, of course, that the integral exists, that is, $X \in L(P)$. Hence, we have all the integration theory from Chapter 3 available, and with the simplification that our measure is a probability measure and so $P(\Omega) = 1$. We shall adopt the notation from general integration theory with the following modifications:

$$\text{" expectation " or " mean " = " } P\text{-integral "}$$

$$EX = \int_\Omega X\,dP$$

$$\text{" convergence in probability " or " in } pr \text{ " = " convergence in } P\text{-measure "}$$

$$\text{" } X_n \to X \text{ in } pr \text{ " = " } X_n \to X \text{ in } P\text{-measure "}$$

4.2 Computations of Expectations

Let $(S, \mathcal{B}, \mu)$ be a measure space and let $U : (\Omega, \mathcal{F}, P) \to (S, \mathcal{B}, \mu)$ be a random function. Then the image measure of P under U is called the distribution or the law of U and is denoted P_U [see (2.8)]. If P_U is absolutely μ-continuous with μ-density f [see (3.18)], we say that U is *absolutely μ-continuous* and f is called *the μ-density* of U. By the first transformation theorem (3.15.1) we have

$$E\psi(U) = \int_S \psi(s)\,P_U(ds) \qquad \text{if } \psi : S \to \overline{\mathbf{R}} \text{ is measurable} \tag{4.2.1}$$

If U is absolutely μ-continuous with μ-density f_X, then by (4.2.1) and (3.17.2) we have

$$E\psi(U) = \int_S \psi(s)\,f_X(s)\,\mu(ds) \qquad \text{if } \psi : S \to \overline{\mathbf{R}} \text{ is measurable} \tag{4.2.2}$$

where (4.2.1) and (4.2.2) should be understood in the usual sense: If one of the integrals exists, so does the other and they are equal. Suppose that $\mathcal{B}$ contains all singletons and let $\#$ be the counting measure on $(S, \mathcal{B})$ [see (1.13)]. Then by (2.14.10) we have that U is discrete if and only if U is absolutely $\#$-continuous; and if so, then the probability mass function p_X is the $\#$-density of U. Hence, if U is discrete, then by (4.2.2) and (3.38.3) we have

$$E\psi(U) = \sum_{s \in S} \psi(s)\, p_X(s) \qquad \text{if } \psi : S \to \overline{\mathbf{R}} \text{ is measurable} \tag{4.2.3}$$

in the usual sense.

Consider the case where $S = \mathbf{R}^k$ and $\mathcal{B} = \mathcal{B}(\mathbf{R}^k)$. So let $X = (X_1, \ldots, X_k)$ be a k-dimensional random vector and let $\psi : \mathbf{R}^k \to \overline{\mathbf{R}}$ be a Borel function. Then by (4.2.1), (2.8.10) and (3.27.8) we have

$$E\psi(X) = \int_{\mathbf{R}^k} \psi(x_1, \ldots, x_k)\, F_X(dx_1, \ldots, dx_k) \tag{4.2.4}$$

in the usual sense, where F_X is the distribution function of X. Let λ_k be the k-dimensional Lebesgue measure on $\mathbf{R}^k$. Then X is absolutely λ_k-continuous if and only if F_X is absolutely continuous in the sense of (3.27); and if so, we say that X is *absolutely continuous*, and the λ_k-density of X is just called *the density* of X. By (4.2.2) we have that if X is absolutely continuous with density f, then

$$E\psi(X) = \int_{\mathbf{R}^k} \psi(x_1, \ldots, x_k)\, f(x_1, \ldots, x_k)\, dx_1 dx_2 \cdots dx_k \tag{4.2.5}$$

in the usual sense.

Let X be an absolutely continuous real random variable with density function $f_X(t)$ such that $P(X \in I) = 1$ where I is a given interval, and let $Y = \phi(X)$ where ϕ is an bijective function from the interval I onto the interval J with inverse function $\psi : J \to I$. Suppose that ψ is absolutely continuous on J with density ψ'. Then $P(Y \in J) = 1$ and ψ is either increasing or decreasing on J. If ψ is increasing, we find by (4.2.5) and the substitution rule (3.29.5) that

$$\begin{aligned} P(a < Y \le b) = P(\psi(a) < X \le \psi(b)) &= \int_{\psi(a)}^{\psi(b)} f_X(t)\, dt \\ &= \int_a^b f_X(\psi(s))\, \psi'(s)\, ds \end{aligned}$$

for all $a, b \in J$. If ψ is decreasing, then applying the formula above on $-X$ and $\tilde{\phi}(t) := \phi(-t)$ we get

$$P(a < Y \leq b) = -\int_a^b f_X(\psi(s))\,\psi'(s)\,ds$$

for all $a, b \in J$. Thus, if $\phi : I \to J$ is a bijection with an absolute continuous inverse $\psi : J \to I$, and if X is an absolutely continuous random variable with density f_X such that $P(X \in I) = 1$, then $Y = \phi(X)$ is absolutely continuous with density $f_Y(s)$ given by

$$f_Y(s) = \begin{cases} f_X(\psi(s))|\psi'(s)| & \text{if } s \in J \\ 0 & \text{if } s \notin J \end{cases} \tag{4.2.6}$$

Consider a real random variable X . If we define *the tail function* of X to be the function $R_X(t) := P(X > t) = 1 - F_X(t)$ for all $t \in \mathbf{R}$, then by (3.30.1) we have

$$EX = \int_0^\infty R_X(t)\,dt - \int_0^\infty F_X(-t)\,dt \tag{4.2.7}$$

in the usual sense. By the Tonelli-Fubini theorem (3.21), we get

(4.2.8) If $\psi : \mathbf{R} \to \mathbf{R}$ is absolutely continuous with density ψ' [see (3.27)] such that $E|\varphi(X)| < \infty$, where

$$\varphi(x) = \int_0^x |\psi'(t)|\,dt \qquad \forall x \in \mathbf{R}$$

then $\psi(X) \in L^1(P)$, and we have

$$E\psi(X) = \psi(0) + \int_0^\infty \{\psi'(t)\,R_X(t) - \psi'(-t)\,F_X(-t)\}\,dt$$

Suppose that X is a non-negative integer-valued random variable. Then X only assumes the values $0, 1, 2, \ldots$ and so $F_X(-t) = 0$ for $t > 0$ and $R_X(t) = P(X > n)$ for all $t \in [n, n+1[$. Hence, by (4.2.3), (4.2.7) and (4.2.8) we have

$$EX = \sum_{n=0}^\infty P(X > n) = \sum_{n=1}^\infty P(X \geq n) \tag{4.2.9}$$

$$E\psi(X) = \psi(0) + \sum_{n=0}^\infty (\psi(n+1) - \psi(n))P(X > n) \tag{4.2.10}$$

in the usual sense. Note that formulas (4.2.5)–(4.2.10) show how the computation of expectations may be reduced to a computation of ordinary sums and Lebesgue integrals.

4.3 Moments and Metrics

Let X be a random variable. Then the quantity $E|X|^q$ is called *the q–absolute moment* of X and we define

(4.3.1) $$m_X(q) := E|X|^q \quad \forall q \geq 0 \ , \ \mathcal{D}(m_X) := \{ q \geq 0 \mid E|X|^q < \infty \}$$

with the convention: $x^0 = 1$ for all $x \in \mathbf{R}$, that is, $m_X(0) = 1$. The set $\mathcal{D}(m_X)$ is called the domain of m_X , and the function $m_X : \mathcal{D}(m_X) \to \mathbf{R}$ is called *the absolute moment transform* of X . In (4.14)–(4.21) we shall consider a series of other transforms of random variables. Let $0 \leq q \leq \infty$ be given. Recall that we have defined $||X||_q$ as follows [see (3.22)]:

(4.3.2) $$||X||_0 = \inf\{a \in \overline{\mathbf{R}}_+ \mid P(|X| > a) \leq a\}$$

(4.3.3) $$||X||_q = m_X(q) = E|X|^q \qquad \text{if } 0 < q \leq 1$$

(4.3.4) $$||X||_q = m_X(q)^{1/q} = (E|X|^q)^{1/q} \qquad \text{if } 1 \leq q < \infty$$

(4.3.5) $$||X||_\infty = \inf\{a \in \overline{\mathbf{R}}_+ \mid P(|X| > a) = 0\}$$

and we defined $M(P)$ and $L^q(P)$ as follows:

(4.3.6) $$M(P) = \{X \mid X \text{ is a real random variable } \}$$

(4.3.7) $$L^q(P) = \{X \in M(P) \mid ||X||_q < \infty\} \qquad \forall 0 \leq q \leq \infty$$

Note that $M(P) = L^0(P)$ by (3.22.28), and from (3.22.8) we find

(4.3.8) If $\varphi \in \mathcal{L}(0, \infty)$ is bounded, then $X_n \to X$ in *pr* if and only if $E\,\varphi(|X_n - X|) \to 0$, and if and only if $X_n \to X$ in 0–mean

We shall frequently use this result with $\varphi(x) = \arctan(x)$, which is a bounded function in $\mathcal{L}(0, \infty)$. From Hölder's inequality (3.11) and Markov's inequality (3.9) we have

$$(E|X|^r)^{1/r} \leq (E|X|^q)^{1/q}(E\,1^p)^{1/p} = (E|X|^q)^{1/q}$$
$$P(|X| > a) \leq a^{-r}EX^r \leq a \qquad \forall a \geq (EX^r)^{1/(1+r)}$$

whenever $0 < r < q$ and $p = \frac{rq}{q-r}$. Moreover, if $s, t > 0$ and $v = \lambda s + (1-\lambda)t$, where $0 < \lambda < 1$, then setting $p = \frac{1}{\lambda}$ $q = \frac{1}{1-\lambda}$ and $r = 1$ in Hölder's inequality (3.11) we find

$$E|X|^v = E\big(|X|^{\lambda s}|X|^{(1-\lambda)t}\big) \le \big(E|X|^s\big)^\lambda \big(E|X|^t\big)^{1-\lambda}$$

Taking logarithms, we see that $\log(m_X(q))$ is convex and since exp is increasing and convex, we find

(4.3.9) $\quad (\|X\|_0)^{\frac{r+1}{r}} \le (E|X|^r)^{\frac{1}{r}} \le (E|X|^q)^{\frac{1}{q}} \le \|X\|_\infty \qquad \forall 0 < r \le q < \infty$

(4.3.10) $\quad q \curvearrowright \log m_X(q)$ is convex $: \mathcal{D}(M_X) \to \mathbf{R}_+$

(4.3.11) $\quad m_X$ is convex $: \mathcal{D}(m_X) \to \mathbf{R}_+$

Moreover, from (4.3.9) it follows that

(4.3.12) $\quad \mathcal{D}(m_X)$ is an interval in $\mathbf{R}_+$ containing 0

(4.3.13) $\quad L^\infty(P) \subseteq L^q(P) \subseteq L^r(P) \subseteq L^0(P) \qquad \forall 0 \le r \le q \le \infty$

(4.3.14) $\quad$ If $X_n \to X$ in q-mean, then $X_n \to X$ in pr and in r-mean for all $0 \le r \le q$

Let $0 \le q \le \infty$ be given, then by (3.25.10) we have

(4.3.15) $\quad (L^q(P), \|\cdot\|_q)$ is a complete pseudo–metric linear space

Note that *Markov's inequality* (3.9) and *Jensen's inequality* (3.10) take the following forms for random variables X and Y with $X \in L^1(P)$:

(4.3.16) $\quad P(|Y| \ge a) \le \dfrac{1}{\varphi(a)} E\varphi(|Y|) \qquad$ if φ is increasing

(4.3.17) $\quad \psi(EX) \le E\psi(X) \qquad$ if $\psi : I \to \mathbf{R}$ is convex and $P(X \in I) = 1$

If $n \in \mathbf{N}_0$ and $E|X|^n < \infty$, then EX^n exists and is called *the nth moment* of X , and we define

(4.3.18) $\quad \mu_X(n) := EX^n \qquad \forall n \in \mathcal{D}(\mu_X) := \mathbf{N}_0 \cap \mathcal{D}(m_X)$

The set $\mathcal{D}(\mu_X) \subseteq \mathbf{N}_0$ is called *the domain* of μ_X and μ_X is called *the moment transform* of X . Clearly we have

(4.3.19) $\quad \mu_X(0) = 1 \quad , \quad \mu_X(1) = EX \quad , \quad \mu_X(2n) = m_X(2n)$

If $X \in L^1(P)$, we define *the central moment transform*, denoted μ_X^0 , and *the central absolute moment transform*, denoted m_X^0 , as follows

$$\mu_X^0(n) = E(X - EX)^n \qquad \forall n \in \mathcal{D}(\mu_X) \tag{4.3.20}$$

$$m_X^0(q) = E|X - EX|^q \qquad \forall q \geq 0 \tag{4.3.21}$$

The binomial formula gives the following relation between the moments and central moments:

$$\mu_X^0(n) = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} (EX)^{n-j} \mu_X(j) \qquad \forall n \in \mathcal{D}(\mu_X) \tag{4.3.22}$$

$$\mu_X(n) = \sum_{j=0}^{n} \binom{n}{j} (EX)^{n-j} \mu_X^0(j) \qquad \forall n \in \mathcal{D}(\mu_X) \tag{4.3.23}$$

Moreover, (4.2.8) gives the following formulas:

$$m_X(q) = q \int_0^\infty t^{q-1} P(|X| > t)\, dt \qquad \forall 0 < q < \infty \tag{4.3.24}$$

$$\mu_X(n) = n \int_0^\infty t^{n-1} (R_X(t) + (-1)^n F_X(-t))\, dt \quad \forall n \in \mathcal{D}(\mu_X) \setminus \{0\} \tag{4.3.25}$$

Let $0 \leq r \leq \infty$ and $0 < q < \infty$ be given numbers, let $X, X_1, X_2, \ldots$ be given real random variables, and let (U^q) denote the following condition:

(U^q) : $\{\, |X_n|^q \mid n \geq 1 \,\}$ is uniformly P-integrable

Since $P(\Omega) = 1$, then by (3.22.12) we have the following scheme of implications for the convergence notions introduced above:

$$\begin{array}{ccccccc} & P\text{-a.s.} & \Rightarrow & \text{in } pr & \Leftrightarrow & \text{arctan-mean} & \Leftrightarrow \text{ 0-mean} \\ & \Uparrow & & \Uparrow & & & \\ P\text{-a.s} + (U^q) & & \Rightarrow & q\text{-mean} & \Rightarrow & r\text{-mean} & \forall 0 \leq r \leq q \\ & & & \Uparrow & & & \\ & & & \text{in } pr + (U^q) & & & \end{array} \tag{4.3.26}$$

where convergence in arctan-mean is convergence in φ-mean for the particular function $\varphi(x) := \arctan(x)$. In particular, convergence in pr is the weakest of the convergence notions introduced above.

4.4 Variance and Covariance

The second central moment is nothing but the familiar *variance* of X :

(4.4.1) $$\mathrm{var}(X) := E(X - EX)^2 = \mu_X^0(2) = m_X^0(2)$$

and its square root is the *standard deviation* of X :

(4.4.2) $$\sigma(X) := \sqrt{\mathrm{var}(X)} = \sqrt{E(X - EX)^2}$$

so $\sigma^2(X) = \mathrm{var}(X)$. A simple computation shows that

(4.4.3) $$\mathrm{var}(X) = EX^2 - (EX)^2$$

Applying Markov's inequality (4.3.16) with $\varphi(x) = x^2$ on the random variable $X - EX$ gives *Chebyshev's inequality*:

(4.4.4) $$P(|X - EX| > a) \leq a^{-2}\mathrm{var}(X)$$

Chebyshev's inequality gives an upper bound on the probability that the random variable X deviates more than a from its mean. Setting $a = 5\sigma(X)$ in Chebyshev's inequality gives $P(|X - EX| > 5\sigma(X)) \leq 0.04$. Thus, the probability that a random variable deviates more than five times its standard deviation from its mean is at most 4%. Note that by (3.3.8) we have

(4.4.5) $$\mathrm{var}(X) = 0 \iff X \text{ is degenerated}$$

Let $X, Y \in L^2(P)$. Then by Hölder's inequality (3.11) with $p = q = 2$ and $r = 1$, we see that *the covariance:*

(4.4.6) $$\mathrm{cov}(X, Y) := E(X - EX)(Y - EY) = E(XY) - (EX)(EY)$$

exists as a finite number satisfying

(4.4.7) $$|\mathrm{cov}(X, Y)| \leq \sigma(X)\sigma(Y)$$

Hence, if we define *the correlation coefficient* of X and Y by

(4.4.8) $$\rho(X, Y) = \begin{cases} \dfrac{\mathrm{cov}(X, Y)}{\sigma(X)\,\sigma(Y)} & \text{if } \sigma(X) > 0 \text{ and } \sigma(Y) > 0 \\ 0 & \text{if } \sigma(X) = 0 \text{ or } \sigma(Y) = 0 \end{cases}$$

Then we have

(4.4.9) $$-1 \leq \rho(X, Y) \leq 1$$

Let $X = (X_1, \ldots, X_k)$ be a k-dimensional vector such that $X_j \in L^1(P)$ for all $j = 1, \ldots, k$. Then we define the *mean vector* of X by

$$(4.4.10) \qquad EX := (EX_1, \ldots, EX_k)$$

Similarly, if $X = \{X_{ij}\}$ is a random $(m \times k)$-matrix such that $X_{ij} \in L^1(P)$ for all $1 \le i \le m$ and all $1 \le j \le k$, then we define *the mean matrix* of X by

$$(4.4.11) \qquad EX = \begin{Bmatrix} EX_{11} & EX_{12} & \cdots & EX_{1k} \\ EX_{21} & EX_{22} & \cdots & EX_{2k} \\ \vdots & \vdots & & \vdots \\ EX_{m1} & EX_{m2} & \cdots & EX_{mk} \end{Bmatrix}$$

Let $X = \{X_{ij}\}$ be a random $(m \times k)$-matrix and let $A = \{a_{ij}\}$ be a $(n \times m)$-matrix and $B = \{b_{ij}\}$ be a $(k \times n)$-matrix. Then by (3.42.6) and (3.42.7) we have

$$(4.4.12) \qquad E(AX) = A(EX) \qquad \text{and} \qquad E(XB) = (EX)B$$

Let $X = (X_1, \ldots, X_k)$ be a k-dimensional random vector such that $X_j \in L^2(P)$ for all $j = 1, \ldots, k$. Then the *covariance matrix* of X is the $(k \times k)$-matrix defined by

$$(4.4.13) \qquad V(X) = \begin{Bmatrix} \operatorname{cov}(X_1, X_1) & \operatorname{cov}(X_1, X_2) & \cdots & \operatorname{cov}(X_1, X_k) \\ \operatorname{cov}(X_2, X_1) & \operatorname{cov}(X_2, X_2) & \cdots & \operatorname{cov}(X_2, X_k) \\ \vdots & \vdots & & \vdots \\ \operatorname{cov}(X_k, X_1) & \operatorname{cov}(X_k, X_2) & \cdots & \operatorname{cov}(X_k, X_k) \end{Bmatrix}$$

Hence, if we consider $X = (X_1, \ldots, X_k)$ as a $(1 \times k)$-matrix, then we have

$$V(X) = E\big[(X - EX)^t (X - EX)\big]$$

where A^t denotes *the transposed matrix* of A . Similarly, if we have two random vectors $X = (X_1, \cdots, X_k)$ and $Y = (Y_1, \cdots, Y_m)$, then we define *the covariance matrix* of X and Y by

$$(4.4.14) \qquad \operatorname{Cov}(X, Y) := E\big[(X - EX)^t (Y - EY)\big] = \{\operatorname{cov}(X_i, Y_j)\}$$

Note that if A is a $(k \times n)$-matrix and B is an $(m \times p)$-matrix, then by (4.4.12) we have

$$(4.4.15) \qquad V(XA) = A^t (VX) A \quad , \quad \operatorname{Cov}(XA, YB) = A^t \operatorname{Cov}(X, Y)\, B$$

In particular, if $A = (a_1, \ldots, a_k)^t$, we find

(4.4.16) $$\operatorname{var}\left(\sum_{j=1}^{n} a_j X_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \operatorname{cov}(X_i, X_j)$$

Hence, we see that

(4.4.17) VX is a symmetric, non-negative definite matrix

We say that $X_1, \ldots, X_k$ are *uncorrelated* if $\operatorname{cov}(X_i, X_j) = 0$ for all $i \neq j$. Note that by (4.4.16), (4.4.6) and (4.5.5) below, we have

(4.4.18) $$X_1, \ldots, X_n \text{ uncorrelated} \quad \Rightarrow \quad \operatorname{var}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \operatorname{var}(X_j)$$

(4.4.19) If $X_1, \ldots, X_n$ are independent and $X_1, \ldots, X_n \in L^2(P)$, then $X_1, \ldots, X_n$ are uncorrelated

4.5 Expectations of Independent Random Functions

(1): Let U and V be independent random functions with values in the measurable spaces $(S, \mathcal{A})$ and $(T, \mathcal{B})$, let $X = \varphi(U)$ and $Y = \psi(V)$ where $\varphi : S \to \mathbf{C}$ and $\psi : T \to \mathbf{C}$ are measurable functions, and let g be a measurable function from $S \times T$ into either $\overline{\mathbf{R}}$ or $\mathbf{C}$ such that $g(U, V) \in L(P)$ if g takes values in $\overline{\mathbf{R}}$, and $g(U, V) \in L^1(P, \mathbf{C})$ if g takes values in $\mathbf{C}$. Then $Eg(s, V)$ exists for P_U-a.a. $s \in S$ and

(4.5.1) $P_{(U,V)} = P_U \otimes P_V$ *and* X *and* Y *are independent*

(4.5.2) $Eg(U, V) = Eh(U)$ *where* $h(s) = Eg(s, V)$ *for* P_U*-a.a.* $s \in S$

(4.5.3) $$P((U, V) \in C) = \int_S P((s, V) \in C)\, P_U(ds) \qquad \forall C \in \mathcal{A} \otimes \mathcal{B}$$

(4.5.4) $E|XY| = E|X| \cdot E|Y|$

(4.5.5) *If* $X, Y \in L^1(P)$, *then* $XY \in L^1(P)$ *and* $E(XY) = EX \cdot EY$

(2): Suppose that $U_1, \ldots, U_n$ are independent random functions with values in the measurable spaces $(S_1, \mathcal{A}_1), \ldots, (S_n, \mathcal{A}_n)$, and let $X_j = \varphi_j(U_j)$ where φ_j is a measurable real- or complex-valued function. Then we have

(4.5.6) $P_{(U_1,\ldots,U_n)} = P_{U_1} \otimes \cdots \otimes P_{U_n}$ *and* $X_1, \ldots, X_n$ *are independent*

(4.5.7) $$E|X_1 \cdots X_n| = \prod_{j=1}^{n} E|X_j|$$

(4.5.8) *If* X_j *is integrable for all* $j = 1, \ldots, n$ *, then so is* $\prod_{j=1}^{n} X_j$ *and*

$$E\left(\prod_{j=1}^{n} X_j\right) = (EX_1) \cdot (EX_2) \cdots (EX_n)$$

(3): Let $(S_j, \mathcal{B}_j, \mu_j)$ *be a sum-finite measure space and let* U_j *be an* S_j*-valued random function for all* $j = 1, \ldots, n$ *such that* $U = (U_1, \ldots, U_n)$ *is absolutely* μ*-continuous with* μ*-density* $f(s_1, \ldots, s_n)$ *, where* $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ *is the product measure. Then* U_j *has* μ_j*-density given by*

(4.5.9) $$f_j(u) = \int_{S_1} d\mu_1 \cdots \int_{S_{j-1}} d\mu_{j-1} \int_{S_{j+1}} d\mu_{j+1} \cdots \int_{S_n} f_u^{(j)}\, d\mu_n$$

for all $u \in S_j$ *and all* $j = 1, \ldots, n$ *, where* $f_u^{(j)}$ *is the function from* $S_1 \times \cdots \times S_{j-1} \times S_{j+1} \times \cdots \times S_n$ *into* $[0, \infty]$ *given by*

$$f_u^{(j)}(s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_n) = f(s_1, \ldots, s_{j-1}, u, s_{j+1}, \ldots s_n)$$

for all $u \in S_j$ *and all* $j = 1, \ldots, n$ *. Moreover, the following three statements are equivalent:*

(4.5.10) $U_1, \ldots, U_n$ *are independent*

(4.5.11) $$f(s_1, \ldots, s_n) = \prod_{j=1}^{n} f_j(s_j) \quad \mu\text{—a.s.}$$

(4.5.12) *There exist functions* $h_j : S_j \to \overline{\mathbf{R}}$ *for* $j = 1, \ldots, n$ *such that*

$$f(s_1, \ldots, s_n) = \prod_{j=1}^{n} h(s_j) \quad \mu\text{-a.s.}$$

Proof: (4.5.1): By (2.10.6), we have that X and Y are independent and by independence, we have $P_{(U,V)}(A \times B) = P(U \in A\ ,\ V \in B) = P_U(A)P_V(B)$ for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$. Hence, (4.5.1) follows from (3.20.8).

(4.5.2): Suppose that $g \geq 0$. Then by (4.2.1), (4.5.1) and Tonelli's theorem (3.21.1) we find

$$Eg(U,V) = \int_{S \times T} g(s,t)\, P_{(U,V)}(ds, dt) = \int_S P_U(ds) \int_T g(s,t)\, P_V(dt)$$

$$= \int_S Eg(s,V)\, P_U(ds) = \int_S h(s)\, P_U(ds) = Eh(U)$$

Hence, (4.5.2) holds for non-negative functions g , but then (4.5.2) follows easily for general functions g .

Setting $g = 1_C$ we see that (4.5.3) follows from (4.5.2), and setting $g(s,t) = \varphi(s)\psi(t)$, we see that (4.5.4+5) likewise follows from (4.5.2). Moreover, the extension (4.5.6)–(4.5.8) to the case of n independent random function follows in exactly the same manner.

(3): For the sake of simplicity, let me only do the proof for the case $n = 2$. Then by Tonelli's theorem (3.21.1) we have

$$\begin{aligned} P(U_1 \in A_1) = P(U \in A_1 \times S_2) &= \int_{A_1 \times S_2} f(u,v)\,\mu(du,dv) \\ &= \int_{A_1} \mu_1(du) \int_{S_2} f(u,v)\,\mu_2(dv) = \int_{A_1} f_1(u)\,\mu_1(du) \end{aligned}$$

for all $A_1 \in \mathcal{A}_1$, and so U_1 is absolutely μ_1-continuous with μ_1-density f_1 , and similarly, for U_2 .

(4.5.10) $\Rightarrow$ (4.5.11): Suppose that U_1 and U_2 are independent, and let $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$ be given. By (4.5.3), (4.2.2) and Tonelli's theorem (3.21.1) we find

$$\begin{aligned} \int_C f(s,t)\,\mu(ds,dt) = P(U \in C) &= \int_{S_1} P((s,U_2) \in C)\, f_1(s)\,\mu_1(ds) \\ &= \int_{S_1} \mu_1(ds) \int_{S_2} 1_C(s,t)\, f_1(s)\, f_2(t)\,\mu_2(dt) \\ &= \int_C f_1(s)\, f_2(t)\,\mu(ds,dt) \end{aligned}$$

Thus, by (3.17) we conclude that $f(s,t) = f_1(s)f_2(t)$ μ-a.s.

(4.5.11) $\Rightarrow$ (4.5.12): Evident! [Set $h_j = f_j$.]

(4.5.12) $\Rightarrow$ (4.5.10): Let $A_j \in \mathcal{A}_j$ for $j = 1,2$ be given. Then by Fubini's theorem (3.21.2) we get

$$\begin{aligned} P(U_1 \in A_1\ ,\ U_2 \in A_2) &= \int_{A_1 \times A_2} h_1(s)\, h_2(t)\,\mu(ds,dt) \\ &= \int_{A_1} h_1(s)\,\mu_1(ds) \cdot \int_{A_2} h_2(t)\,\mu_2(dt) \end{aligned}$$

Hence, by (2.4) we conclude that U_1 and U_2 are independent. $\square$

4.6 Sums of Independent Random Vectors

Let $X = (X_1, \ldots, X_k)$ *and* $Y = (Y_1, \ldots, Y_k)$ *be independent k-dimensional random vectors, and let* ϕ *be a measurable function from* $\mathbf{R}^k$ *into either* $\overline{\mathbf{R}}$ *or* $\mathbf{C}$ *such that* $\phi(X+Y) \in L(P)$ *if* ϕ *takes values in* $\overline{\mathbf{R}}$, *and* $\phi(X+Y) \in L^1(P, \mathbf{C})$ *if* ϕ *takes values in* $\mathbf{C}$. *Then* $E\phi(s+Y)$ *exists for* P_X*-a.a.* $s \in \mathbf{R}^k$ *and*

$$E\phi(X+Y) = E\psi(X) \text{ where } \psi(s) = E\phi(s+Y) \text{ for } P_X\text{-a.a. } s \in \mathbf{R}^k \tag{4.6.1}$$

$$P(X+Y \in C) = \int_{\mathbf{R}^k} P(X+s \in C)\, F_Y(ds) \qquad \forall C \in \mathcal{B}(\mathbf{R}^k) \tag{4.6.2}$$

$$F_{X+Y}(v) = F_X \star F_Y(v) = \int_{\mathbf{R}^k} F_X(v-s)\, F_Y(ds) \tag{4.6.3}$$

$$p_{X+Y}(v) = p_X \star p_Y(v) = \sum_s p_X(v-s)\, p_Y(s) \tag{4.6.4}$$

$$E|X+Y|^q \geq E|X+EY|^q \quad \text{if } k=1,\ q \geq 1 \text{ and } Y \in L^1(P) \tag{4.6.5}$$

where $F_X \star F_Y$ *is the Lebesgue-Stieltjes convolution [see (3.27.12)] and* $p_X \star p_Y$ *is the discrete convolution [see (3.27.14)]. If* X *is absolutely continuous with density* f_X, *then* $X+Y$ *is absolute continuous with density function given by*

$$f_{X+Y}(v) = f_X \star F_Y(v) = \int_{\mathbf{R}^k} f_X(v-s)\, F_Y(ds) \tag{4.6.6}$$

where $f_X \star F_Y$ *is the Lebesgue-Stieltjes convolution [see (3.27.12)]. Moreover, if* X *and* Y *are absolutely continuous with densities* f_X *and* f_Y, *then* $X+Y$ *is absolutely continuous with density*

$$f_{X+Y}(v) = f_X \star f_Y(v) = \int_{\mathbf{R}^k} f_X(v-s)\, f_Y(s)\, ds \tag{4.6.7}$$

where $f_X \star f_Y$ *is the Lebesgue convolution [see (3.27.13)].*

Proof: Setting $g(s,t) = \phi(s+t)$, we see that (4.6.1+2) follows from (4.5.2+3). Setting $C = \{(s,t) \mid s+t \leq v\}$ and $C = \{(s,t) \mid s+t = v\}$, we see that (4.6.3+4) follows from (4.6.2).

(4.6.5): Let $\phi(t) = |t|^q$ where $q \geq 1$. Then $t \curvearrowright \phi(s+t)$ is convex for every fixed $s \in \mathbf{R}$. Hence, if $\mu = EY$, then by Jensen's inequality (4.3.17) we have $\psi(s) = E|s+Y|^q \geq |s+\mu|^q$ and so (4.6.5) follows from (4.6.1).

(4.6.6): Let $C \in \mathbf{B}(\mathcal{R}^k)$. Then by (4.6.1) and Tonelli's theorem (3.21.1) we get

$$\begin{aligned} P(X+Y \in C) &= \int_{\mathbf{R}^k} F_Y(ds) \int_{\mathbf{R}^k} 1_C(s+t)\, F_X(dt) \\ &= \int_{\mathbf{R}^k} F_Y(ds) \int_{\mathbf{R}^k} 1_C(s+t)\, f_X(t)\, dt \\ &= \int_{\mathbf{R}^k} F_Y(ds) \int_{\mathbf{R}^k} 1_C(v)\, f_X(v-s)\, dv \\ &= \int_C dv \int_{\mathbf{R}^k} f_X(v-s)\, F_Y(ds) \\ &= \int_C (f_X \star F_Y)(v)\, dv \end{aligned}$$

Thus, (4.6.6) holds and (4.6.7) follows from (4.6.6) and (3.27.10). □

4.7 Convergence of Partial Sums

Let $X_1, X_2, \cdots$ be independent random variables and let $S_n = X_1 + \cdots + X_n$ denote their partial sums for $n = 1, 2, \ldots$.

(1): If $S_n \to S$ in pr for some random variable S , then $S_n \to S$ a.s.

(2): Suppose that $a_n \in \mathbf{R}$, $c \in \mathbf{R}_+$ and that $v : \mathbf{N}_0 \to \mathbf{N}$ is an increasing function satisfying $v(0) = 1$, $\lim_{n\to\infty} v(n) = \infty$ and

(4.7.1) $\quad |a_k| \le c\,|a_{v(n)}|$ *if* $n \ge 1$ *and* $v(n-1) \le k < v(n)$

(4.7.2) $\quad a_n S_n \to 0$ *in pr and* $a_{v(n)} S_{v(n)} \to 0$ *a.s.*

Then we have that $a_n S_n \to 0$ a.s.

Remark: Let $\{Z_n\}$ be a sequence of random variables. Then we say that $\sum_1^\infty Z_n$ is *summable a.s.* respectively *summable in pr* if the partial sums $T_n = Z_1 + \cdots + Z_n$ converges a.s. respectively in *pr* to a random variable T with $P(|T| = \infty) = 0$. Similarly, if $0 \le q \le \infty$ we say that $\sum_1^\infty Z_n$ is *summable in q-mean* if T_n converges in q-mean to a random variable $T \in L^q(P)$. Hence, if $X_1, X_2, \ldots$ are independent random variables, then the first part of the theorem above together with (3.25.7) gives

(4.7.3) $$\sum_{n=1}^\infty X_n \text{ is summable in } pr \iff \sum_{n=1}^\infty X_n \text{ is summable a.s.}$$

This result together with the second part of theorem are two of the rare cases, where convergence in probability implies a.s.-convergence.

Proof: Suppose that $S_n \to S$ in pr . Then by (3.25.4) there exists a strictly increasing function $\sigma : \mathbf{N} \to \mathbf{N}$ such that $S_{\sigma(n)} \to S$ P-a.s. Set $\sigma(0) = 1$ and let $k \in \mathbf{N}$. Since $\sigma(n) \uparrow \infty$, there exists a unique integer $q \in \mathbf{N}$ such that $\sigma(q-1) \leq k < \sigma(q)$, and we define $\tau(k) = \sigma(q-1)$ if $\sigma(q-1) \leq k < \sigma(q)$. Then τ is non-decreasing, $\tau(k) \leq k$ for all $k \geq 1$ and $\tau(k) \to \infty$ as $k \to \infty$. Set

$$U_k = |S - S_k| \quad , \quad V_k = |S_{\tau(k)} - S_k| \quad , \quad W_k = |S - S_{\tau(k)}|$$

Then $\sigma(U_k, U_{k+1}, \ldots) \subseteq \sigma(X_{k+1}, X_{k+2}, \ldots)$ and $\sigma(V_k) \subseteq \sigma(X_1, \ldots, X_k)$. Hence, (2.18.3) holds. Moreover, since $U_k \leq V_k + W_k$, then by (2.18.6) we have

$$P(U > \varepsilon) \liminf_{k \to \infty} P(V_k \leq \tfrac{1}{2}\,\varepsilon) \leq P(W \geq \tfrac{1}{2}\,\varepsilon) \qquad \forall \varepsilon > 0$$

where $U = \limsup U_k$ and $W = \limsup W_k$. By assumption we have that $S_k \to S$ in pr and $S_{\tau(k)} \to S$ in pr . Hence, $P(V_k \leq \frac{1}{2}\varepsilon) \to 1$ as $k \to \infty$. Moreover, by definition of σ we have that $W = 0$ a.s. Thus, $P(W \geq \frac{1}{2}\varepsilon) = 0$ and so $P(U > \varepsilon) = 0$ for all $\varepsilon > 0$. But this means that $U = 0$ P-a.s. and so $S_k \to S$ P-a.s.

Suppose that (4.7.1+2) hold and set $w(k) = v(q)$ if $v(q-1) \leq k < v(q)$. Moreover, let

$$U_k = |a_k S_k| \qquad V_k = |a_k(S_k - S_{w(k)})| \qquad W_k = |a_k S_{w(k)}|$$

Then $U_k \leq V_k + W_k$, and since $w(k) \geq k$, then, as above, we see that (2.18.2) holds. Note that by (4.7.1) we have

$$V_k \leq |a_k S_k| + c\,|a_{w(k)} S_{w(k)}| \quad , \quad W_k \leq c\,|a_{w(k)} S_{w(k)}|$$

So by (4.7.2) we have that $V_k \to 0$ in pr , and $W_k \to 0$ P-a.s. Hence, as above, we conclude that $U_k \to 0$ P-a.s. But this means that $a_k\, S_k \to 0$ P-a.s. and the theorem is proved. □

4.8 The Birth of the Law of Large Numbers

About 300 years ago, around 1695, James Bernoulli (1654-1705) found the theorem which should become the destiny of probability theory. James Bernoulli was born in the same year as probability theory (1654). He spent some years in

Paris studying mathematics, and in order to earn money for his studies he worked part time at the printing company that produced Huyghens' book *De Ratiociniis in Ludo Alea.* He became interested in probability theory and in the period 1690–1705 he worked on his excellent book *Ars Conjectandi* ("The art of guessing") which was published in 1713, eight years after his death in 1705. The main result of the book is what James called "the golden theorem" and what nowadays is called "the law of large numbers." In modern language his result goes as follows:

Suppose that we make a series of independent trials of an experiment with two outcomes, say 0 and 1, having probabilities $1-p$ and p, respectively. Let f_n denote the frequency of "1" in the first n trials, that is, $f_n = (X_1 + \cdots + X_n)/n$ where X_j is the random variable that takes the values 0 or 1 according to the outcome of the jth trial. Then James showed that $P(|f_n - p| > \varepsilon)$ tends to zero as n tends to infinity no matter how small $\varepsilon > 0$ is. That is, James proved that $f_n \to p$ in pr and he gave a very precise estimate of the speed of convergence.

James recognized the immense importance of his result. In the period from the birth of probability in 1654 until Kolmogorov's model in 1933 the only solid founded model for probability was the lottery model [see (2.15)]. James realized that his theorem might provide a rigorous definition of real-life probabilities such as: What is the probability that a 25-year-old woman will survive her 65th birthday? Or, what is the probability that a ship crossing the Atlantic Ocean will arrive safely? These were probabilities that were of vital interest for the insurance companies which just had come into existence a few years before the birth of James Bernoulli.

Part IV of *Ars Conjectandi* bears the title "The art of conjecturing, showing the use and application of the preceding treatise in civil, moral, and economic affairs." In this part, James speculates about how probability can be applied to daily-life problems, and he tries to lay down a solid foundation of such probabilities, but in vain as plenty of his successors. It was first with the emergence of measure theory in the beginning of this century such solid foundation was found. The law of large numbers is the main link between reality and probability theory, but the modern model of probability is not based on the law of large numbers; thus, we have to prove it.

Well, what is the law of large numbers? Let $X_1, X_2, \cdots$ be a sequence of random variables and let $\mu_1, \mu_2, \cdots$ be a sequence of real numbers, called *the centering constants.* Consider *the centered averages*:

$$\hat{X}_n = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_j) \tag{4.8.1}$$

The *law of large numbers* states that under such and such conditions, we have that $\hat{X}_n \to 0$ in some sense as $n \to \infty$. If $\hat{X}_n \to 0$ in pr or $\hat{X}_n \to 0$ in q-mean we speak about a *weak law of large numbers*; if $\hat{X}_n \to 0$ P-a.s. we speak about

a *strong law of large numbers*. In general, a strong law of large numbers is much harder to prove than a weak law of large numbers. Set

$$(4.8.2) \qquad s_q(n) = E\Big|\sum_{j=1}^{n}(X_j - \mu_j)\Big|^q \qquad \forall q > 0\ \forall n = 1, 2, \cdots$$

Then we have

$$(4.8.3) \qquad E|\hat{X}_n|^q = n^{-q} s_q(n) \qquad \forall q > 0\ \forall n = 1, 2, \cdots$$

Hence, we find the following prototype of *the weak law of large numbers:*

$$(4.8.4) \qquad \hat{X}_n \to 0 \text{ in } q\text{-mean} \iff \lim_{n\to\infty} n^{-q} s_q(n) = 0$$

Recall that by (4.3.8) and (4.3.14) we have:

$$(4.8.5) \qquad \text{If } \hat{X}_n \to 0 \text{ in } q\text{-mean, then } \hat{X}_n \to 0 \text{ in } r\text{-mean } \forall 0 \le r \le q$$

By (3.25.3) we get the following prototype of *the strong law of large numbers:*

$$(4.8.6) \qquad \text{If } \sum_{n=1}^{\infty} n^{-q} s_q(n) < \infty \text{ then } \hat{X}_n \to 0 \quad P\text{-a.s.}$$

The condition: $\sum n^{-q} s_q(n) < \infty$ is a very strong restriction which fails in most cases. In such cases, the following version, which also follows directly from (3.25.3), is often useful:

(4.8.7) If $v : \mathbf{N} \to \mathbf{N}$ is an increasing function, then we have:

$$\sum_{k=1}^{\infty} \frac{s_q(v(k))}{v(k)^q} < \infty \quad \Rightarrow \quad \hat{X}_{v(n)} \to 0 \ P\text{-a.s.}$$

Note that if $n^{-q} s_q(n) \to 0$ as $n \to \infty$, then we can find an increasing function satisfying the condition of (4.8.7). Thus, in this case we may not have a.s.-convergence of $\{\hat{X}_n\}$, but we do have a.s.-convergence of sufficiently thin subsequences of $\{\hat{X}_n\}$.

The observations above show that the qth moment $s_q(n)$ of the centered averages plays a crucial role in the law of large numbers and much of the work we are going to do consist in estimating the size of $s_q(n)$. Note that by Hölder's inequality (3.11.1) and Minkowski's inequality (3.12.3) we have

$$(4.8.8) \qquad \Big|\sum_{j=1}^{n} u_j\Big|^q \le n^r \sum_{j=1}^{n} |u_j|^q \qquad \text{where} \qquad r = (q-1)^+$$

for all $n \in \mathbf{N}$ and all $u_1, \ldots, u_n \in \mathbf{R}$. Hence, we find

$$(4.8.9) \qquad s_q(n) \le n^r \sum_{j=1}^{n} E|X_j - \mu_j|^q \qquad \text{where} \qquad r = (q-1)^+$$

Khinchine's inequality (4.32.1) gives a considerable improvement of this inequality when $X_1, X_2, \cdots$ are independent.

If the means EX_j exist and are finite, it is customary to use the mean as centering constants in the law of large numbers, but other choices of $\{\mu_j\}$ are possible. Note that if $\mu_j = EX_j$ and $\sigma_{ij} = \operatorname{cov}(X_i, X_j)$, then by (4.4.16) and (4.8.4) we have

$$(4.8.10) \qquad s_2(n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij}$$

$$(4.8.11) \qquad \frac{1}{n} \sum_{j=1}^{n} (X_j - EX_j) \to 0 \text{ in 2-mean} \iff \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} = 0$$

In particular, if $X_1, X_2, \cdots$ are uncorrelated and $\sigma_j^2 = \operatorname{var}(X_j)$, then

$$(4.8.12) \qquad \frac{1}{n} \sum_{j=1}^{n} (X_j - EX_j) \to 0 \text{ in 2-mean} \iff \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = 0$$

In applications, (4.8.7) is particularly useful, when v is rapidly increasing. Precisely speaking, we say that $v : \mathbf{N} \to \mathbf{N}$ is *rapidly increasing* if v is increasing and

$$(4.8.13) \qquad \liminf_{n\to\infty} \frac{v(n+1)}{v(n)} > 1$$

For instance, $v(n) = int(\beta^n)$ is rapidly increasing if $\beta > 1$ since

$$(4.8.14) \qquad \lim_{n\to\infty} \frac{int(\beta^{n+1})}{int(\beta^n)} = \beta$$

But $v(n) = n^2$ is **not** rapidly increasing since $(n+1)^2/n^2 \to 1$. It is an elementary calculus exercise [see Exc.4.43] to show that if $v : \mathbf{N} \to \mathbf{N}$ is rapidly increasing, then there exist constants $c_\alpha \in \mathbf{R}_+$ for every $\alpha > 0$ such that

$$(4.8.15) \qquad \sum_{j=k}^{\infty} v(j)^{-\alpha} \le c_\alpha v(k)^{-\alpha} \qquad \forall \alpha > 0 \qquad \forall k = 1, 2, \cdots$$

It is instructive to compare (4.8.15) to the following well-known (or easily proved) inequalities:

$$\frac{k^{1-\alpha}}{\alpha-1} \le \sum_{j=k}^{\infty} j^{-\alpha} \le \frac{\alpha\, k^{1-\alpha}}{\alpha-1} \qquad \forall\, \alpha > 1 \qquad \forall\, k = 1, 2, \cdots \tag{4.8.16}$$

4.9 Kronecker's and Cesàro's Lemmas

Let $\{a_n\}$ *and* $\{b_n\}$ *be two sequences of real numbers, and suppose that* $b_1 \ge b_2 \ge \cdots > 0$ *and* $b_n \to 0$ *as* $n \to \infty$. *Then we have*

(4.9.1) *If* $\lim\limits_{n\to\infty} a_n = a$, *then* $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{j=1}^{n} a_j = a$

(4.9.2) *If* $\sum\limits_{n=1}^{\infty} a_n b_n$ *converges and is finite, then* $\lim\limits_{n\to\infty} b_n \sum\limits_{j=1}^{n} a_j = 0$

Moreover, if the sequence $\{a_j\}$ *satisfies*

(4.9.3) *either* $\limsup\limits_{j\to\infty} a_j < \infty$ *or* $\liminf\limits_{j\to\infty} a_j > -\infty$

and if for every $\varepsilon > 0$ *there exists an increasing function* $v : \mathbf{N} \to \mathbf{N}$ *such that* $\lim\limits_{n\to\infty} v(n) = \infty$ *and*

$$\lim_{n\to\infty} \frac{1}{v(n)} \sum_{j=1}^{v(n)} a_j = 0 \qquad \textit{and} \qquad \limsup_{n\to\infty} \frac{v(n+1)}{v(n)} \le 1 + \varepsilon \tag{4.9.4}$$

then we have $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{j=1}^{n} a_j = 0$.

Remark: (4.9.1) is called *Cesàro's lemma*, and (4.9.2) is called *Kronecker's lemma.*

Proof: (4.9.1) is easy and I shall leave the proof to the sedulous reader.

(4.9.2): Let $r_n = \sum_{j=n}^{\infty} a_j b_j$. Then $r_n \to 0$ because $\sum_{j=1}^{\infty} a_j b_j$ converges to a finite value. Observe that $a_n = b_n^{-1}(r_n - r_{n+1})$. Hence, we have

$$\left| b_n \sum_{j=1}^{n} a_j \right| = \left| b_n \left(\sum_{j=1}^{n} \frac{r_j}{b_j} - \sum_{j=2}^{n+1} \frac{r_j}{b_{j-1}} \right) \right|$$

$$\leq \frac{b_n |r_1|}{b_1} + |r_{n+1}| + b_n \sum_{j=2}^{n} \left(\frac{1}{b_j} - \frac{1}{b_{j-1}} \right) |r_j|$$

Let $\varepsilon > 0$ be given and choose $p \in \mathbf{N}$ and $a \in \mathbf{R}_+$ such that $|r_j| < \varepsilon$ for all $j \geq p$ and $|r_j| \leq a$ for all $j \geq 1$. Then we have

$$\left| b_n \sum_{j=1}^{n} a_j \right| \leq \frac{b_n a}{b_p} + 2\varepsilon \qquad \forall n \geq p$$

and since $b_n \to 0$ and p is fixed, we see that

$$\limsup_{n\to\infty} \left| b_n \sum_{j=1}^{n} a_j \right| \leq 2\varepsilon \qquad \forall \varepsilon > 0$$

Thus, (4.9.2) follows.

Suppose that (4.9.4) holds and $\limsup a_n < \infty$. Then there exists $a \in \mathbf{R}_+$ such that $a_n \leq a$ for all $n \geq 1$. Let $\varepsilon > 0$ be given and let $v : \mathbf{N} \to \mathbf{N}$ be an increasing function satisfying (4.9.4). If $n \geq v(1)$, then there exists a unique integer $q = q_n$ such that $v(q) \leq n < v(q+1)$, and we set $\sigma(n) = v(q)$ and $\tau(n) = v(q+1)$. Then we have

$$\frac{1}{n} \sum_{j=1}^{n} a_j \leq \frac{1}{n} \left| \sum_{j=1}^{\sigma(n)} a_j \right| + \frac{1}{n} \sum_{\sigma(n)<j\leq n}^{n} a_j$$

$$\leq \left| \frac{1}{\sigma(n)} \sum_{j=1}^{\sigma(n)} a_j \right| + \left(1 - \frac{\sigma(n)}{n} \right) a$$

$$\leq \left| \frac{1}{\sigma(n)} \sum_{j=1}^{\sigma(n)} a_j \right| + \left(1 - \frac{\sigma(n)}{\tau(n)} \right) a$$

because $a_j \leq a$ and $\sigma(n) \leq n < \tau(n)$. Similarly, we have

$$\frac{1}{n} \sum_{j=1}^{n} a_j \geq -\frac{1}{n} \left| \sum_{j=1}^{\tau(n)} a_j \right| - \frac{1}{n} \sum_{j=n+1}^{\tau(n)} a_j$$

$$\geq -\frac{\tau(n)}{\sigma(n)} \left| \frac{1}{\tau(n)} \sum_{j=1}^{\tau(n)} a_j \right| - \left(\frac{\tau(n)}{\sigma(n)} - 1 \right) a$$

and since $q = q_n \to \infty$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \frac{1}{\sigma(n)} \sum_{j=1}^{\sigma(n)} a_j = \lim_{q\to\infty} \frac{1}{v(q)} \sum_{j=1}^{v(q)} a_j = 0$$

$$\lim_{n\to\infty} \frac{1}{\tau(n)} \sum_{j=1}^{\tau(n)} a_j = \lim_{q\to\infty} \frac{1}{v(q+1)} \sum_{j=1}^{v(q+1)} a_j = 0$$

$$\limsup_{n\to\infty} \frac{\tau(n)}{\sigma(n)} \le \limsup_{q\to\infty} \frac{v(q+1)}{v(q)} \le 1+\varepsilon$$

Hence, we find

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} a_j \le \left(1 - \frac{1}{1+\varepsilon}\right) a \quad , \quad \liminf_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} a_j \ge -\varepsilon a$$

for all $\varepsilon > 0$. Thus, letting $\varepsilon \downarrow 0$, we see that $\frac{1}{n} \sum_{j=1}^{n} a_j \to 0$. The case where $\liminf a_n > -\infty$ is proved similarly. □

4.10 The Law of Large Numbers I

Let $\{X_n\}$ be a sequence of random variables and let $\{\mu_n\}$ be a sequence of real numbers. Suppose that there exist $q > r \ge 0$, $c \ge 0$ and $\tau_j \ge 0$ for $j = 1, 2, \cdots$ satisfying

(4.10.1) $$E\left|\sum_{j=1}^{n}(X_j - \mu_j)\right|^q \le c n^r \sum_{j=1}^{n} \tau_j \qquad \forall n \ge 1$$

(4.10.2) $$\sum_{j=1}^{\infty} j^{r-q} \tau_j < \infty$$

and let $\hat{X}_n = \frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j)$ be the centered average.

(1): Suppose that $v : \mathbf{N} \to \mathbf{N}$ is a given rapidly increasing function [see (4.8.13)]. Then we have

(4.10.3) $\hat{X}_n \to 0$ *in q-mean, and* $\hat{X}_n \to 0$ *in pr*

(4.10.4) $\hat{X}_{v(n)} \to 0$ *P-a.s.*

(4.10.5) $\hat{X}_n \to 0$ *P-a.s. if* $X_1, X_2, \cdots$ *are independent*

(4.10.6) $\hat{X}_n \to 0$ *P-a.s. if* $P(L_0 = \infty\ ,\ L_1 = -\infty) = 0$

where $L_0 = \limsup_{j\to\infty} (X_j - \mu_j)$ *and* $L_1 = \liminf_{j\to\infty} (X_j - \mu_j)$.

(2): Suppose that (4.10.1) holds and that there exist finite numbers $\lambda, p, a \geq 0$ *and* $\kappa_j \geq 0$ *for* $j = 1, 2, \ldots$ *satisfying*

(4.10.7) $0 \leq \lambda \leq 1$ *and* $\lambda < q - r$

(4.10.8) $\sum_{j=1}^{\infty} j^{r-q+\lambda} \tau_j < \infty \quad , \quad \sum_{j=1}^{\infty} j^{\beta(1-\lambda)-q} \kappa_j < \infty$

(4.10.9) $E\left\{ \max_{m\leq k<n} \left| \sum_{j=m}^{k} (X_j - \mu_j) \right|^q \right\} \leq a(n-m)^p \sum_{j=m}^{n-1} \kappa_j \qquad \forall 1 \leq m < n$

where $\beta = p \wedge (q-1)^+$. *Then we have*

(4.10.10) $\hat{X}_n \to 0$ *P-a.s. and in q-mean*

Remark: Note that (4.8.9) gives an inequality of the type (4.10.1) with $r = (q-1)^+$ and $\tau_j = E|X_j - \mu_j|^q$. Moreover, note that if $\tau_j = \kappa_j = E\,|X_j - \mu_j|^q$, then by (4.8.9) and Exc.4.46 we have that (4.10.1) and (4.10.9) holds with $a = c = 1$ and $p = r = (q-1)^+$. Khinchine's inequalities (4.32.1) and (4.34.1) gives similar inequalities with $p = r = \frac{1}{2}(q-2)^+$, provided $X_1, X_2, \ldots$ are independent and $\mu_j = E\,X_j$.

Warning: If $\{x_n\}$ is a sequence of ***real*** variables such that $x_{v(n)} \to 0$ for every rapidly increasing function $v : \mathbf{N} \to \mathbf{N}$, then it is easily shown that $x_n \to 0$. However, if $\{\hat{X}_n\}$ is a sequence of ***random*** variables satisfying (4.10.4) for every rapidly increasing function v , then $\{\hat{X}_n\}$ is not necessarily a.s. convergent to zero. The problem being that for each rapidly increasing function v we have an exceptional null set N_v , but as there are uncountably many rapidly increasing functions, the union of the exceptional null sets $\{N_v\}$ may not be a null set.

Proof: By (4.10.1) we have

$$E|\hat{X}_n|^q = n^{-q}\, E\left| \sum_{j=1}^{n} (X_j - \mu_j) \right|^q \leq c\, n^{r-q} \sum_{j=1}^{n} \tau_j$$

and since $q > r$, we have that $n^{r-q} \downarrow 0$; so by (4.10.2) and Kronecker's lemma (4.9.2), we see that $E|\hat{X}_n|^q \to 0$. Hence, (4.10.3) follows from (4.3.14).

(4.10.4): Let $v : \mathbf{N} \to \mathbf{N}$ be rapidly increasing, and set $d = q - r$. Let $\sigma(j) = \min\{n \in \mathbf{N} \mid v(n) \geq j\}$; then by (4.8.15) and (4.10.1) we have

$$\sum_{n=1}^{\infty} E|\hat{X}_{v(n)}|^q \leq c \sum_{n=1}^{\infty} \sum_{j=1}^{v(n)} v(n)^{-d} \tau_j = c \sum_{j=1}^{\infty} \sum_{n=\sigma(j)}^{\infty} v(n)^{-d} \tau_j$$
$$\leq c\, c_d \sum_{j=1}^{\infty} v(\sigma(j))^{-d} \tau_j \leq c\, c_d \sum_{j=1}^{\infty} j^{-d} \tau_j$$

because $v(\sigma(j)) \geq j$ by definition of σ . Hence, (4.10.4) follows from (4.10.2) and (4.8.7).

(4.10.5): By (4.10.3) we have that $\hat{X}_n \to 0$ in pr ; and if $v(n) = 2^n$, then by (4.10.4) we have $\hat{X}_{v(n)} \to 0$ P-a.s. Setting $a_n = \frac{1}{n}$, we find

$$|a_k| = \tfrac{1}{k} \leq 2^{-n+1} = 2|a_{v(n)}| \quad \text{if} \quad 2^{n-1} \leq k \leq 2^n$$

Thus, (4.10.5) follows from (4.7).

(4.10.6): Let $v_k(n) = int(1 + 2^{-k})^n$ for $k, n \in \mathbf{N}$. Then v_k is rapidly increasing for all $k \in \mathbf{N}$, and by (4.8.14) we have

$$\lim_{n\to\infty} \frac{v_k(n+1)}{v_k(n)} = 1 + 2^{-k} \qquad \forall k = 1, 2, \cdots$$

Since $\{v_k \mid k \in \mathbf{N}\}$ is a countable collection of rapidly increasing functions [see the warning above] there exist a P-null set N such that

$$\lim_{n\to\infty} \hat{X}_{v_k(n)}(\omega) = 0 \qquad \forall \omega \in \Omega \setminus N \qquad \forall k \geq 1$$
$$\text{either } L_0(\omega) < \infty \text{ or } L_1(\omega) > -\infty \qquad \forall \omega \in \Omega \setminus N$$

Hence, (4.10.6) follows from the last part of (4.9).

(4.10.10): Let $d = q - r$. Set $\xi(k) = 2^{k-1}$ if $\lambda = 0$ and $\xi(k) = int(k^{1/\lambda})$ if $\lambda > 0$. Then by Exc.4.47 we have that $\xi : \mathbf{N} \to \mathbf{N}$ is strictly increasing, and there exists a constant c_λ satisfying

(i) $$\sum_{j=m}^{\infty} \xi(j)^{-d} \leq c_\lambda\, \xi(m)^{\lambda - d} \qquad \forall m \geq 1$$

(ii) $$\xi(m+1) \leq c_\lambda\, \xi(m) \ , \ \xi(m+1) - \xi(m) \leq c_\lambda\, \xi(m)^{1-\lambda} \quad \forall m \geq 1$$

Let $\hat{X}_0 = 0$, $V_k = \hat{X}_{\xi(k)-1}$ and

$$W_k = \xi(k)^{-1} \left\{ \max_{\xi(k) \le i < \xi(k+1)} \left| \sum_{j=\xi(k)}^{i} (X_j - \mu_j) \right| \right\}$$

Moreover, let $\sigma(j) = \min\{k \ge 1 \mid \xi(k) \ge j\}$. Then $\xi(\sigma(j)) \ge j$ and $\lambda - d = r - q + \lambda < 0$, and so by (i) and (4.10.1) we find

$$\sum_{k=1}^{\infty} E|V_k|^q \le c \sum_{k=1}^{\infty} \sum_{j=1}^{\xi(k)} \xi(k)^{-d} \tau_j = c \sum_{j=1}^{\infty} \sum_{k=\sigma(j)}^{\infty} \xi(k)^{-d} \tau_j$$

$$\le c c_\lambda \sum_{j=1}^{\infty} \xi(\sigma(j))^{\lambda - d} \tau_j \le c c_\lambda \sum_{j=1}^{\infty} j^{\lambda - d} \tau_j$$

Hence, by (4.10.8) and (3.25.3) we have that $V_k \to 0$ P-a.s. Let j be an integer with $\xi(k) \le j < \xi(k+1)$. Then by (ii) we have

$$(\xi(k+1) - \xi(k))^\beta \, \xi(k)^{-q} \le c_\lambda^\beta \, \xi(k)^{\beta(1-\lambda)} \, \xi(k)^{-q} \le c_\lambda^{\beta+q} \, j^{\beta(1-\lambda)-q}$$

Setting $m = j$ and $n = j+1$ in (4.10.9) gives $E\,|X_j - \mu_j|^q \le a\kappa_j$. Hence, by Exc.4.46 we see that (4.10.9) holds with p replaced by β . Thus, we get

$$\sum_{k=1}^{\infty} E|W_k|^q \le a \sum_{k=1}^{\infty} \sum_{\xi(k) \le j < \xi(k+1)} (\xi(k+1) - \xi(k))^\beta \, \xi(k)^{-q} \, \kappa_j$$

$$\le a c_\lambda^{\beta+q} \sum_{j=1}^{\infty} j^{\beta(1-\lambda)-q} \, \kappa_j$$

and as above we see that $W_k \to 0$ P-a.s. Let $n \ge 1$ be given; then there exists a unique $k \ge 1$ such that $\xi(k) \le n < \xi(k+1)$, and so we have:

$$|\hat{X}_n| \le \frac{1}{n} \left| \sum_{j=1}^{\xi(k)-1} (X_j - \mu_j) \right| + \frac{1}{n} \left| \sum_{j=\xi(k)}^{n} (X_j - \mu_j) \right| \le |V_k| + |W_k|$$

Hence, we conclude that $\hat{X}_n \to 0$ P-a.s. Moreover, since (4.10.8) implies (4.10.2), then by (4.10.3) we have that $\hat{X}_n \to 0$ in q-mean. Thus, (4.10.10) is proved. □

4.11 The Law of Large Numbers II

Let $\{X_n\}$ be a sequence of uncorrelated random variables satisfying

(4.11.1) $$\sum_{j=1}^{\infty} j^{-2} \operatorname{var}(X_j) < \infty$$

and let $\hat{X}_n = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_j)$, *where* $\mu_j = EX_j$. *Suppose that* $v : \mathbf{N} \to \mathbf{N}$ *is a rapidly increasing function. Then we have*

(4.11.2) $\hat{X}_n \to 0$ *in 2-mean and* $\hat{X}_n \to 0$ *in pr*

(4.11.3) $\hat{X}_{v(n)} \to 0$ *P-a.s.*

(4.11.4) $\hat{X}_n \to 0$ *P-a.s. if* $X_1, X_2, \cdots$ *are independent*

(4.11.5) $\hat{X}_n \to 0$ *P-a.s. if* $P(L_0 = \infty\ ,\ L_1 = -\infty) = 0$

(4.11.6) $\hat{X}_n \to 0$ *P-a.s. if* $X_i \perp\!\!\!\perp X_j \quad \forall i \neq j$ *and* $\sup_j E|X_j - \mu_j| < \infty$

(4.11.7) $\hat{X}_n \to 0$ *P-a.s. if* $\sum_{j=1}^{\infty} j^{-3/2}\,\mathrm{var}(X_j) < \infty$

where $L_0 = \limsup_{j\to\infty}(X_j - \mu_j)$ *and* $L_1 = \liminf_{j\to\infty}(X_j - \mu_j)$.

Proof: Let $\sigma_j^2 = \mathrm{var}(X_j)$. Then by (4.4.18) we have

$$E\Big|\sum_{j=1}^{n}(X_j - \mu_j)\Big|^2 = \mathrm{var}\Big(\sum_{j=1}^{n} X_j\Big) = \sum_{j=1}^{n}\sigma_j^2$$

Thus, (4.11.2)–(4.11.5) follows from (4.11.1) and (4.10).

(4.11.6): Let $k = \sup_j E|X_j - \mu_j|$, and set $Y_j = (X_j - \mu_j)^+$. Since $Y_i \perp\!\!\!\perp Y_j$ for $i \neq j$, we have that $\{Y_n\}$ are uncorrelated and

$$\mathrm{var}\,(Y_j) \le E\,Y_j^2 \le E|X_j - \mu_j|^2 = \sigma_j^2$$

Moreover, if $\nu_j = E\,Y_j$, then $(Y_j - \nu_j) \ge -k$ for all $j \ge 1$. Hence, applying (4.11.5) with $\{X_j\}$ replaced by $\{Y_j\}$, we see that

$$\frac{1}{n}\sum_{j=1}^{n}(Y_j - \nu_j) \to 0 \qquad P\text{-a.s.}$$

Similarly, if $Z_j = (X_j - \mu_j)^-$ and $\lambda_j = EZ_j$, then

$$\frac{1}{n}\sum_{j=1}^{n}(Z_j - \lambda_j) \to 0 \qquad P\text{-a.s.}$$

and since $X_j - \mu_j = Y_j - Z_j$, we have $\nu_j - \lambda_j = E(X_j - \mu_j) = 0$, and so

$$\frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_j) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \nu_j) - \frac{1}{n}\sum_{j=1}^{n}(Z_j - \lambda_j)$$

Thus, (4.11.6) follows.

(4.11.7): Set $\tau_j = \kappa_j = \sigma_j^2$. Then by (4.4.18) and (iii) in Exc.4.46 we see that (4.10.1) and (4.10.9) hold with $a = c = 1$, $q = 2$, $r = 0$ and $p = 1$. Let $\lambda = \frac{1}{2}$. Then (4.10.7) holds and $r - q - \lambda = \beta(1-\lambda) - q = -\frac{3}{2}$. Thus, (4.11.7) follows from (4.10.10). □

4.12 The Law of Large Numbers III

Let $\{X_n\}$ be a sequence of pairwise independent and identically distributed random variables; that is, X_i and X_j are independent and $X_i \sim X_j$ for all $i \neq j$. Suppose that $\mu = E X_1$ exists [but possibly $+\infty$ or $-\infty$]. Then we have

$$\frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \quad P\text{-a.s.} \tag{4.12.1}$$

Proof: Let $R(t) = P(|X_1| > t)$ for $t \geq 0$, and suppose that μ is finite. Then $E|X_1| < \infty$ and if we set

$$Y_j = \begin{cases} X_j & \text{if } |X_j| \leq j \\ 0 & \text{if } |X_j| > j \end{cases}$$

then $Y_j \perp\!\!\!\perp Y_i$ for $i \neq j$, $P(|Y_j| > t) \leq R(t)$ if $0 \leq t < j$ and $P(|Y_j| > t) = 0$ if $t \geq j$ because $X_i \sim X_1$. Hence, by (4.3.24) we have

$$\begin{aligned}
&\sum_{n=1}^{\infty} n^{-2}\,\mathrm{var}(Y_n) \leq \sum_{n=1}^{\infty} n^{-2}\, EY_n^2 \leq 2\sum_{n=1}^{\infty} n^{-2} \int_0^n t\,R(t)\,dt \\
&= 2\sum_{n=1}^{\infty}\sum_{j=1}^{n} n^{-2} \int_{j-1}^{j} t\,R(t)\,dt = 2\sum_{j=1}^{\infty}\Big(\sum_{n=j}^{\infty} n^{-2}\Big)\int_{j-1}^{j} t\,R(t)\,dt \\
&\leq 4\sum_{j=1}^{\infty} j^{-1}\int_{j-1}^{j} t\,R(t)\,dt \leq 4\int_0^{\infty} R(t)\,dt = 4E|X_1|
\end{aligned}$$

where I have used (4.8.16) with $\alpha = 2$ and the fact that $j^{-1}t \leq 1$ for all $t \in [j-1, j]$. Let $\mu_j = EY_j$. Since $X_j \sim X_1$, by (4.3.24) and (3.13.1) we have

$$\begin{aligned}
&E|Y_j - \mu_j| \leq E|Y_j| + |\mu_j| \leq 2E|Y_j| \leq 2\int_0^j R(t)\,dt \leq 2E|X_1| < \infty \\
&\mu_j = \int_{\{|X_j|\leq j\}} X_j\,dP = \int_{\{|X_1|\leq j\}} X_1\,dP \to \mu \qquad \text{as } j \to \infty
\end{aligned}$$

Thus, by (4.11.6) and (4.9.1) we have

$$\frac{1}{n}\sum_{j=1}^{n} Y_j = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \mu_j) + \frac{1}{n}\sum_{j=1}^{n}\mu_j \to \mu \qquad P\text{-a.s.}$$

Moreover, since

$$\begin{aligned}\sum_{j=1}^{\infty} P(X_j \neq Y_j) &= \sum_{j=1}^{\infty} R(j) \leq \sum_{j=1}^{\infty}\int_{j-1}^{j} R(t)\,dt \\ &= \int_0^{\infty} R(t)\,dt = E|X_1| < \infty\end{aligned}$$

then by the first Borel–Cantelli lemma (2.11.1) we have that $P(N) = 0$ where $N = \limsup\{X_j \neq Y_j\}$. If $\omega \in \Omega \setminus N$, then there exists $p \in \mathbf{N}$ so that $X_j(\omega) = Y_j(\omega)$ for $j \geq p$. Hence, $X_j - Y_j \to 0$ P-a.s. Thus, by (4.9.1) we find

$$\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n}\sum_{j=1}^{n} Y_j + \frac{1}{n}\sum_{j=1}^{n}(X_j - Y_j) \to \mu \qquad P\text{-a.s.}$$

and so (4.12.1) holds whenever μ is finite.

Suppose that $\mu = \infty$ and let $k \in \mathbf{N}$. Since $E X_1^- < \infty$, we have that $\mu_k = E(X_1 \wedge k)$ exists and is finite. Moreover, $\{X_n \wedge k \mid n = 1, 2, \cdots\}$ are pairwise independent and identically distributed; so by the argument above we have

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j \geq \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n}(X_j \wedge k) = \mu_k \qquad P\text{-a.s.}$$

But $\mu_k \uparrow \mu = \infty$ by the monotone convergence theorem (3.4), and so we see that (4.12.1) holds whenever $\mu = \infty$. In exactly the same way, one shows that (4.12.1) holds whenever $\mu = -\infty$. □

4.13 The Borel-Cantelli Lemma Revisited

Let $A_1, A_2, \ldots$ be a sequence of pairwise independent events. Then we have

$$\text{(4.13.1)} \qquad \sum_{n=1}^{\infty} P(A_n) = \infty \quad \Rightarrow \quad P(\limsup A_n) = 1$$

Remark: The second Borel-Cantelli lemma (2.11.2) gives the same result but under the much more restrictive assumption that the sequence $\{A_n\}$ is totally independent.

Proof: Let $S_n = \sum_{j=1}^n 1_{A_j}$ and $S = \sum_{j=1}^\infty 1_{A_j}$, then

$$ES_n = \sum_{j=1}^n p_j \ , \quad \mathrm{var}(S_n) = \sum_{j=1}^n p_j(1-p_j) \le \sum_{j=1}^n p_j$$

where $p_j = P(A_j)$. Set $\mu_n = p_1 + \cdots + p_n$. Since $S_n \le S$, by Chebyshev's inequality (4.4.4) we have

$$\begin{aligned} &P(S \le \tfrac{1}{2}\mu_n) \le P(S_n \le \tfrac{1}{2}\mu_n) = P(S_n - \mu_n \le -\tfrac{1}{2}\mu_n) \\ &\le P(|S_n - \mu_n| \ge \tfrac{1}{2}\mu_n) \le 4\mu_n^{-2}\mathrm{var}(S_n) \le 4\mu_n^{-1} \end{aligned}$$

By assumption, we know that $\mu_n \to \infty$ and so $\mu_n^{-1} \to 0$. Hence, letting $n \to \infty$, we find that $P(S < \infty) = 0$. Thus, $P(S = \infty) = 1$ and since $\{S = \infty\} = \limsup A_n$ we have proved (4.13.1). □

4.14 Transforms of Random Functions

Let X be a random function on the probability space $(\Omega, \mathcal{F}, P)$ taking values in the measurable space $(S, \mathcal{A})$. Then the properties of X as a function of $\omega \in \Omega$, such as continuity, differentiability, etc., are irrelevant for probability theory. What matters is the distribution P_X of X or possibly the joint distribution of X and other given random functions. However, as we saw in Chapter 1, measures are complicated objects. Usually distributions are given by specifying their distribution functions or probability mass functions or density function or by specifying some other transform of the distribution, as defined in (3.41).

Let $(S, \mathcal{A})$ and $(T, \mathcal{B})$ be two measurable spaces, and let Γ be a kernel on $S \times T$. Suppose that X is an S-valued random function, then we define *the* Γ*-transform* of X , denoted Γ_X , and its *domain*, denoted $\mathcal{D}(\Gamma_X)$, as in (3.41), that is,

(4.14.1) $\quad \mathcal{D}(\Gamma_X) := \{t \in T \mid E|\Gamma(X,t)| < \infty\}$

(4.14.2) $\quad \Gamma_X(t) := E\{\Gamma(X,t)\} \qquad \forall t \in \mathcal{D}(\Gamma_X)$

If $\Gamma^*(t,s) = \Gamma(s,t)$ is the transposed kernel, and Y is a T-valued random function, then we define *the* Γ^**-transform* and its *domain* similarly.

Let Γ be a measurable kernel on $S \times T$, and let X and Y be random functions taking values in S and T, respectively. Then by (3.41.9)-(3.41.11) we find

(4.14.3) $\mathcal{D}(\Gamma_X) \in \mathcal{B}$ and Γ_X is $\mathcal{B}$-measurable on $\mathcal{D}(\Gamma_X)$

(4.14.4) $\mathcal{D}(\Gamma_Y^*) \in \mathcal{A}$ and Γ_Y^* is $\mathcal{A}$-measurable on $\mathcal{D}(\Gamma_Y)$

(4.14.5) $E\{\Gamma_Y^*(X)\} = E\{\Gamma_X(Y)\}$ if $\Gamma(X,Y) \in L^1(P)$

Equation (4.14.5) is called *Parseval's equation.* In (4.16)–(4.21) you will find ten special cases of transforms which have proved to be useful in probability theory.

Suppose that Γ is a multiplicative kernel on $S \times T$ [see (3.41)] and that the functions $t \curvearrowright t^*$ and $(t_0,t_1) \curvearrowright t_0 \nabla t_1$ are measurable. Let X be an S-valued random function, and let Y and Z be T-valued random functions. Then by (3.32.7) and (4.5.5) we have

(4.14.6) $\Gamma_{Y^*}^*(s) = \overline{\Gamma_Y^*(s)} \qquad \forall s \in \mathcal{D}(\Gamma_Y)$

(4.14.7) $\Gamma_{Y\nabla Z}^*(s) = \Gamma_Y^*(s)\Gamma_Z^*(s) \qquad \forall s \in \mathcal{D}(\Gamma_Y) \cap \mathcal{D}(\Gamma_Z) \qquad$ if $Y \underline{\|} Z$

(4.14.8) $\Gamma_X(t^*) = \overline{\Gamma_X(t)} \qquad \forall t \in \mathcal{D}(\Gamma_X)$

Moreover, by (3.41) and (1.37) we get the following inversion theorem for multiplicative transforms:

4.15 Inversion of Multiplicative Transforms

Let Γ be a multiplicative kernel on $S \times T$ where $(S,\mathcal{A})$ and $(T,\mathcal{B})$ are measurable spaces. Let $t \curvearrowright t^*$ and $(t_0,t_1) \curvearrowright t_0 \nabla t_1$ satisfy (3.41.1+2), let X and Y be two S-valued random function, and suppose that $T_0 \subseteq T_1 \subseteq T$ satisfies

(4.15.1) $t_0^* \in T_0$ *and* $t_0 \nabla t_1 \in T_1 \qquad \forall t_0 \in T_0 \qquad \forall t_1 \in T_1$

(4.15.2) $\limsup_{n\to\infty} \left\{ \frac{1}{n!} E \mid \Gamma(X,t) \mid^n \right\}^{1/n} < \infty \qquad \forall t \in T_0$

(4.15.3) $T_1 \subseteq \mathcal{D}(\Gamma_X) \cap \mathcal{D}(\Gamma_Y)$ *and* $\Gamma_X(t) = \Gamma_Y(t) \qquad \forall t \in T_1$

Then $P(X \in H) = P(Y \in H)$ *for all* $H \in \mathcal{H}$ *where* $\mathcal{H} = \sigma(\Gamma(\cdot, t) \mid t \in T_0)$. *Moreover, if* S *is a polish space,* $\mathcal{A} = \mathcal{B}(S)$ *and*

(4.15.4) $\forall s_0, s_1 \in S$ *with* $s_0 \neq s_1$ $\exists H_0, H_1$ *disjoint open sets, with* $H_1, H_2 \in \mathcal{H}$, $s_1 \in H_1$ *and* $s_2 \in H_2$

then we have $X \sim Y$.

Remarks: (1): Let Γ be a kernel on $S \times T$ and let X be an S-valued random function. If $T_1 \subseteq T$, then we say that X is Γ-*determined* on T_1 if $T_1 \subseteq \mathcal{D}(\Gamma_X)$ and $X \sim Y$, whenever Y is an S-valued random function such that $T_1 \subseteq \mathcal{D}(\Gamma_Y)$ and $\Gamma_X(t) = \Gamma_Y(t)$ for all $t \in T_1$.

(2): By (3.41.13) and (3.41.14) we have that (3.41.2) is equivalent to any of the following two conditions:

$$\limsup_{n\to\infty} \left\{ \frac{1}{(2n)!} E|\Gamma(X,t)|^{2n} \right\}^{1/2n} < \infty \qquad \forall t \in T_0 \tag{4.15.5}$$

$$\limsup_{n\to\infty} \left\{ \tfrac{1}{n} (E|\Gamma(X,t)|^n)^{1/n} \right\} < \infty \qquad \forall t \in T_0 \tag{4.15.6}$$

It can be shown [see Remark (4) to (3.41)] that the above theorem remains valid if we replace (3.41.2) by

$$\limsup_{n\to\infty} \left\{ a_n (E|\Gamma(X,t)|^n)^{1/n} \right\} < \infty \qquad \forall t \in T_0 \tag{4.15.7}$$

where $\{a_n\}$ is any given sequence of non-negative numbers with $\sum a_n = \infty$.

(3): Note that if $\Gamma(\cdot, t)$ is bounded for all $t \in T_0$, then (4.15.2) holds by (3.39.7). Also observe that if Γ satisfies

(4.15.8) $\Gamma(\cdot, t)$ is continuous for all $t \in T$

(4.15.9) $\forall s_0, s_1 \in S$ with $s_0 \neq s_1$ $\exists t \in T_0$: $\Gamma(s_0, t) \neq \Gamma(s_1, t)$

then $\mathcal{H}$ satisfies (4.15.4). By some more refined methods from the theory of analytic sets, it can be shown that we have

(4.15.10) Suppose that S is an analytic space, and that Γ is a kernel on $S \times T$ where S is endowed with its Borel σ-algebra $\mathcal{B}(S)$. If there exist a countable set $D \subseteq T$ satisfying

$$\forall s_0 \neq s_1 \; \exists t \in D \; : \; \Gamma(s_0, t) \neq \Gamma(s_1, t)$$

then $\mathcal{B}(S) = \sigma(\Gamma(\cdot, t) \mid t \in D)$. Hence, if X and Y are S-valued random variables satisfying (4.5.1)–(4.5.3) for some sets T_0 and T_1 with $D \subseteq T_0 \subseteq T_1 \subseteq T$, then $X \sim Y$

4.16 Distribution Functions

Let $F : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ be defined by $F(s,t) = 1$ if $s \leq t$, and $F(s,t) = 0$ if $s > t$. Then F is a measurable kernel on $\mathbf{R} \times \mathbf{R}$ and the F-transform F_X of a random variable X is nothing but our old acquaintance, *the distribution function* of X :

(4.16.1) $\quad F_X(t) = P(X \leq t) \qquad \forall t \in \mathbf{R}$

The F^*-transform F_X^* is essentially the tail function of X [see (4.2)]:

(4.16.2) $\quad F_X^*(s) = P(X \geq s) = 1 - F_X(s-) \qquad \forall s \in \mathbf{R}$

Setting $t^* := t$ and $t_0 \nabla t_1 := t_0 \wedge t_1$, we see that F satisfies (3.41.1) and (3.41.2), and setting $s := s^*$ and $s_0 \nabla s_1 := s_0 \vee s_1$, we see that F^* satisfies (3.41.1) and (3.41.2). By (1.9) we have that $\mathcal{B}(\mathbf{R}) = \sigma(F(\cdot, t) \mid t \in D)$ whenever D is dense in $\mathbf{R}$; hence, by (4.15) we have

(4.16.3) $\quad$ If D is dense in $\mathbf{R}$ and $F_X(t) = F_Y(t) \quad \forall t \in D$, then $X \sim Y$

Thus, X is F-determined on D whenever D is a dense subset of $\mathbf{R}$. Moreover, if X and Y are random variables, then by (4.14.5) - (4.14.8) we have

(4.16.4) $\quad F_{X \vee Y}(t) = F_X(t)\, F_Y(t) \qquad \forall t \in \mathbf{R} \qquad$ if $X \underline{\|} Y$

(4.16.5) $\quad F_{X \wedge Y}^*(s) = F_X^*(s)\, F_Y^*(s) \qquad \forall s \in \mathbf{R} \qquad$ if $X \underline{\|} Y$

(4.16.6) $\quad E(F_X(Y)) = E(F_Y^*(X))$

4.17 Probability Mass Functions

Let $(T, \mathcal{B})$ be a measurable space such that $\mathcal{B}$ contains all singletons, and let $p : T \times T \to \mathbf{R}$ be given by $p(s,t) = 0$ if $s \neq t$ and $p(s,s) = 1$. Then p is a kernel on $T \times T$, and the p-transform p_X of X is nothing but our old acquaintance, *the probability mass function* of X :

(4.17.1) $\quad p_X(t) = P(X = t) \qquad \forall t \in T$

By (2.14) we have that every discrete random function X is p-determined on T. If there exists just two different diffuse probability measures on $(T, \mathcal{B})$ [this is for instance, the case if $T = \mathbf{R}$ and $\mathcal{B} = \mathcal{B}(\mathbf{R})$], then it is easily seen that X is p-determined on T if and only if X is discrete.

4.18 Characteristic Functions

Let $\varphi : \mathbf{R}^k \times \mathbf{R}^k \to \mathbf{C}$ be given by $\varphi(s,t) = e^{i(s \bullet t)}$ where $s \bullet t = \sum_1^k s_j t_j$ is the usual inner product in $\mathbf{R}^k$ [see (P.8)]. Then φ is a measurable kernel on $\mathbf{R}^k \times \mathbf{R}^k$ and $\varphi = \varphi^*$. The φ-transform φ_X of a k-dimensional random vector $X = (X_1, \ldots, X_k)$ is called *the characteristic function* of X. Since $|e^{iv}| = 1$ for all $v \in \mathbf{R}$, we have that $\mathcal{D}(\varphi_X) = \mathbf{R}^k$ and

(4.18.1) $$\varphi_X(t) = Ee^{i(X \bullet t)} = E\cos(X \bullet t) + iE\sin(X \bullet t) \qquad \forall t \in \mathbf{R}^k$$

Let X and Y be k-dimensional random vectors, let $a \in \mathbf{R}^n$, and let A be a $(k \times n)$-matrix with transposed matrix A^t. Since $(a + XA) \bullet v = a \bullet v + X \bullet (vA^t)$, then by (3.32.9), (4.14.5) and (3.31) we have

(4.18.2) $$|\varphi_X(t)| \le 1 \quad \forall t \in \mathbf{R}^k \quad , \quad \varphi_X(0) = 1$$

(4.18.3) $$\varphi_X \text{ is continuous on } \mathbf{R}^k$$

(4.18.4) $$E\varphi_X(Y) = E\varphi_Y(X)$$

(4.18.5) $$\varphi_{a+XA}(v) = e^{ia \bullet v}\varphi_X(vA^t) \qquad \forall v \in \mathbf{R}^n$$

Setting $t^* = -t$ and $t_0 \nabla t_1 = t_0 + t_1$, we see that (3.41.1) and (3.41.2) hold. Hence, by (4.14.6), (4.14.7) and (4.15) we find

(4.18.6) $$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)} \qquad \forall t \in \mathbf{R}^k$$

(4.18.7) $$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) \qquad \forall t \in \mathbf{R}^k \text{ if } X \underline{\|} Y$$

(4.18.8) $$X \sim Y \iff \varphi_X(t) = \varphi_Y(t) \qquad \forall t \in \mathbf{R}^k$$

whenever X and Y are k-dimensional random vectors. Thus, every k-dimensional random vector is φ-determined on $\mathbf{R}^k$.

Let $X = (X_1, \ldots, X_k)$ be a random vector, and let $\mathbf{e}_1, \ldots, \mathbf{e}_k$ be the standard unit vectors in $\mathbf{R}^k$. Since $e^0 = 1$, we see that

(4.18.9) $$\varphi_{X_j}(v) = \varphi_X(v\mathbf{e}_j) = \varphi_X(0, \ldots, 0, v, 0, \ldots, 0) \qquad \forall v \in \mathbf{R}$$

where v sits on the jth coordinate. Moreover, since

$$\exp(i(X \bullet t)) = \prod_{j=1}^{k} \exp(i(X_j \bullet t_j))$$

then by (4.5.9) we see that

$$\text{(4.18.10)} \quad X_1, \ldots, X_n \text{ independent} \Rightarrow \varphi_X(t) = \prod_{j=1}^{n} \varphi_{X_j}(t_j) \ \forall t = (t_1, \ldots, t_n)$$

Conversely, suppose that there exist functions $\psi_j : \mathbf{R} \to \mathbf{C}$ satisfying

$$\varphi_X(t_1, \ldots, t_k) = \prod_{j=1}^{k} \psi_j(t_j) \qquad \forall (t_1, \ldots, t_k) \in \mathbf{R}^k$$

and let $a_j = \psi_j(0)$. Then by (4.18.2) we have that $a_1 \cdots a_n = 1$. Hence, by (4.18.9) we have that $\varphi_{X_j} = a_j^{-1} \psi_j$ and by (4.18.10) and (4.18.8) we see that $X_1, \ldots, X_k$ are independent. Thus, we obtain the following important criterion for independence:

(4.18.11) If $\varphi_X(t_1, \ldots, t_k) = \prod_{j=1}^{k} \psi_j(t_j)$ for all $(t_1, \ldots, t_k)$, where $\psi_1, \ldots \psi_n$ are a functions from $\mathbf{R}$ into $\mathbf{C}$, then $X_1, \ldots, X_k$ are independent and $\varphi_{X_j}(t) = \psi_j(0)^{-1} \psi_j(t)$ for all $t \in \mathbf{R}$ and all $j = 1, \ldots, k$

Let us consider the one-dimensional case. Expanding e^{itx} in its Taylor series at a , we find

$$e^{itx} = \sum_{k=0}^{n} \frac{i^k (t-a)^k}{k!} x^k e^{iax} + r_n(t, x, a)$$

where the remainder term r_n is given by

$$r_n(t, x, a) = \frac{(ix)^{n+1}}{n!} \int_a^t (t-s)^n e^{isx} \, ds$$

A simple integration by parts shows that

$$\int_a^t (t-s)^n e^{isx} ds = \frac{(t-a)^n e^{iax}}{ix} + \frac{n}{ix} \int_a^t (t-s)^{n-1} e^{isx} \, ds$$

and since $|e^{isx}| = 1$ and $|v| \wedge 2 \leq 2 \arctan |v|$ for all $v \in \mathbf{R}$, then an easy computation shows that

$$\text{(4.18.12)} \quad |r_n(t, x, a)| \leq \frac{2|x(t-a)|^n}{n!} \arctan\left(\frac{|x(t-a)|}{n+1}\right)$$

Let X be a random variable such that $E|X|^n < \infty$. Setting $\lambda_k(a) = i^k E(X^k e^{iaX})$ and $R_n(t,a) = Er_n(t,X,a)$ we find

$$\varphi_X(t) = \sum_{k=0}^{n} \frac{(t-a)^k}{k!} \lambda_k(a) + R_n(t,a)$$

By Lebesgue's dominated convergence theorem we have that $\lambda_k(a)$ and $R_n(t,a)$ are continuous in a and (t,a) , respectively. Moreover, we see that $|t-a|^{-n} R_n(t,a) \to R_n(a,a) = 0$ as $|t-a| \to 0$. Hence, φ_X is n-times differentiable and $\varphi_X^{(k)}(a) = \lambda_k(a)$ is continuous for $k = 0,1,\ldots,n$.

Thus, if X is a real random variable with $E|X|^n < \infty$, then φ_X is n-times continuously differentiable and

$$\varphi_X^{(k)}(a) = i^k E(X^k e^{iaX}) \qquad \forall a \in \mathbf{R} \quad \forall k = 0,1,\ldots,n \tag{4.18.13}$$

$$\varphi_X(t) = \sum_{k=0}^{n} \frac{(t-a)^k}{k!} \varphi_X^{(k)}(a) + R_n(t,a) \qquad \forall t,a \in \mathbf{R} \tag{4.18.14}$$

where the remainder term R_n satisfies

$$|R_n(t,a)| \le \frac{2|t-a|^n}{n!} E\left\{ |X|^n \arctan\left(\frac{|X(t-a)|}{n+1}\right)\right\} \tag{4.18.15}$$

Recall that $\frac{1}{i} = -i$. Hence, setting $a = 0$ in (4.18.13) and (4.18.14) we find

$$EX^k = (-i)^k \varphi_X^{(k)}(0) \qquad \forall k = 0,1,\ldots,n \tag{4.18.16}$$

$$\varphi_X(t) = \sum_{k=0}^{n} \frac{i^k t^k}{k!} EX^k + R_n(t,0) \qquad \forall t \in \mathbf{R} \tag{4.18.17}$$

Let μ be a finite Borel measure on $\mathbf{R}^k$. Then the φ-transform of μ is called *the Fourier transform* of μ and it is usually denoted $\hat{\mu}$, that is,

$$\hat{\mu}(t) := \varphi_\mu(t) = \int_{\mathbf{R}^k} e^{i(s \bullet t)} \mu(ds) \qquad \forall t \in \mathbf{R}^k \tag{4.18.18}$$

Similarly, if $f : \mathbf{R}^k \to \mathbf{C}$ is Lebesgue integrable, then its *Fourier transform*, denoted $\hat{f}$, is defined by

$$\hat{f}(t) := \int_{\mathbf{R}^k} e^{i(s \bullet t)} f(s) ds \qquad \forall t \in \mathbf{R}^k \tag{4.18.19}$$

The Fourier transform of measures and functions plays an important role in analysis.

4.19 Laplace Transforms

Let $L : \mathbf{R}^k \times \mathbf{C}^k \to \mathbf{C}$ be given by $L(s,t) = e^{s \bullet t}$ where $s \bullet t = \sum_1^k s_j t_j$. Then L is a measurable kernel on $\mathbf{R}^k \times \mathbf{C}^k$ and if $X = (X_1, \ldots, X_k)$ is a random vector, then we define *the Laplace transform* X , denoted L_X , and *the cumulant transform* of X , denoted κ_X , as follows:

(4.19.1) $$L_X(t) := Ee^{X \bullet t} \ \ \forall t \in \mathcal{D}^*(L_X) \ , \ \ \kappa_X(t) := \log L_X(t) \ \ \forall t \in \mathbf{R}^k$$

with the convention: $\log \infty = \infty$, where $\mathcal{D}(L_X) := \{t \in \mathbf{C}^k \mid E\,|e^{X \bullet t}| < \infty\}$ is *the complex domain* of L_X and $\mathcal{D}^*(L_X) := \mathcal{D}(L_X) \cup \mathbf{R}^k$ is *the extended domain* of L_X . Note that $e^{X \bullet t} > 0$ whenever $t \in \mathbf{R}^k$, and so the expectation in (4.19.1) exists and is strictly positive (but possibly $= \infty$) whenever $t \in \mathbf{R}^k$. We define *the real domain*, denoted $\mathcal{R}(L_X)$, as follows:

(4.19.2) $$\mathcal{R}(L_X) = \mathcal{D}(L_X) \cap \mathbf{R}^k = \left\{ t \in \mathbf{R}^k \ \middle| \ L_X(t) < \infty \right\}$$

and we define the *interior domain*, denoted $\mathcal{D}^\circ(L_X)$, to be the interior of $\mathcal{D}(L_X)$ in $\mathbf{C}^k$, that is, $\mathcal{D}^\circ(L_X) := \operatorname{int}(\mathcal{D}(L_X)$. Let $u, v \in \mathbf{R}^k$, and let $0 < \lambda < 1$. Then by Hölder's inequality (3.11) with $p = \frac{1}{\lambda}$, $q = \frac{1}{1-\lambda}$ and $r = 1$, we get

$$\begin{aligned} L_X(\lambda\, u + (1-\lambda)\, v) &= E\left\{ e^{\lambda\,(X \bullet u)}\, e^{(1-\lambda)\,(X \bullet v)} \right\} \\ &\le \left\{ Ee^{X \bullet u} \right\}^\lambda \left\{ Ee^{X \bullet v} \right\}^{1-\lambda} = L_X(u)^\lambda\, L_X(v)^{1-\lambda} \end{aligned}$$

Thus, taking logarithms, we see that $\kappa_X(v)$ is convex on $\mathbf{R}^k$, and so $L_X(v) = \exp(\kappa_X(v))$ is convex on $\mathbf{R}^k$ because exp is increasing and convex on $\mathbf{R}$. Moreover, if we set $\Re t = (\Re t_1, \ldots, \Re t_k)$ whenever $t = (t_1, \ldots, t_k) \in \mathbf{C}^k$, then we have $|e^{X \bullet t}| = e^{\Re(X \bullet t)} = e^{X \bullet \Re t}$. Thus, we conclude that

(4.19.3) $\mathcal{D}(L_X) = \{t \in \mathbf{C}^k \mid \Re t \in \mathcal{R}(L_X)\}$ and $\mathcal{R}(L_X)$ is a convex subset of $\mathbf{R}^k$ containing 0

(4.19.4) $L_X : \mathbf{R}^k \to]0, \infty]$ and $\kappa_X : \mathbf{R}^k \to]-\infty, \infty]$ are convex functions with $L_X(0) = 1$ and $\kappa_X(0) = 0$

Let X and Y be k-dimensional random vectors, let $a \in \mathbf{R}^n$ be a vector, and let A be a $(k \times n)$-matrix with transposed matrix A^t . Then in analogy with (4.18.5)–(4.18.8), we have

(4.19.5) $$L_X(\bar{t}) = \overline{L_X(t)} \qquad \forall t \in \mathcal{D}(L_X)$$

(4.19.6) If X and Y are independent, then $\mathcal{D}(L_{X+Y}) = \mathcal{D}(L_X) \cap \mathcal{D}(L_Y)$ and $L_{X+Y}(t) = L_X(t)\, L_Y(t)$ for all $t \in \mathcal{D}^*(L_{X+Y})$

(4.19.7) $EL_X(Y) = EL_Y(X)$

(4.19.8) $L_X(it) = \varphi_X(t) \qquad \forall t \in \mathbf{R}^k$

(4.19.9) $w \in \mathcal{D}(L_{a+XA}) \iff wA^t \in \mathcal{D}(L_X)$

(4.19.10) $L_{a+XA}(w) = e^{a \bullet w} L_X(wA^t) \qquad \forall w \in \mathcal{D}^*(L_{a+XA})$

Let $a_1, \ldots, a_m$ be given vectors in $\mathbf{R}^k$. Then the *polytope* generated by $\{a_1, \ldots, a_m\}$ is the convex hull $K := \mathrm{co}(a_1, \ldots, a_m)$ of $\{a_1, \ldots, a_m\}$. That is, the set of all convex combinations: $r_1 a_1 + \cdots + r_m a_m$ where $r_j \geq 0$ and $r_1 + \cdots + r_m = 1$. Suppose that $a_1, \ldots, a_m \in \mathcal{R}(L_X)$, and let K be the polytope generated by $\{a_1, \ldots, a_m\}$. Then $K \subseteq \mathcal{R}(L_X)$ by (4.19.3) and $(s \bullet v) \leq \max(s \bullet a_j)$ for all $s \in \mathbf{R}^k$ and all $v \in K$. Hence, if $t \in \mathbf{C}^k$ such that $v = \Re t \in K$, then

$$|e^{s \bullet t}| = e^{s \bullet v} \leq \sum_{j=1}^{m} e^{s \bullet a_j} \qquad \forall s \in \mathbf{R}^k$$

Thus, by Lebesgue's dominated convergence theorem we have

(4.19.11) If $K \subseteq \mathcal{R}(L_X)$ is a polytope and $C = \{t \in \mathbf{C}^k \mid \Re t \in K\}$, then L_X is continuous at t_0 along C for all $t_0 \in C$

(4.19.12) L_X is continuous on $\mathcal{D}^\circ(L_X)$, and if $k = 1$ then L_X is continuous on $\mathcal{D}(L_X)$. However, if $k \geq 2$, then L_X need not be continuous on $\mathcal{D}(L_X)$ [see (4.27)]

[Recall that a function f is continuous at t_0 along B if $f(t_n) \to f(t_0)$ whenever $t_n \in B$ and $t_n \to t_0$.] Setting $t^* := \bar{t}$ and $t_0 \nabla t_1 = t_0 + t_1$, we see that (3.41.1) and (3.41.2) hold, and so (4.15) may be applied to obtain criteria for X to be L-determined on $T_1 \subseteq \mathbf{C}^k$. In particular, we have that every random vector X is L-determined on $T_1 = \{iv \mid v \in \mathbf{R}^k\}$.

Let us consider the one-dimensional case. So let X be a real random variable. Then by (4.19.3) we have that $\mathcal{R}(L_X)$ is an interval on $\mathbf{R}$ containing 0 . As in (3.39), we define

(4.19.13) $R(X) = \limsup_{n \to \infty} \left\{ \frac{1}{n!} E|X|^n \right\}^{1/n} \quad , \quad \rho(X) = R(X)^{-1}$

(4.19.14) $\rho_c(X) := \sup\{r \geq 0 \mid \Re c \pm r \in \mathcal{R}(L_X)\} \qquad \forall c \in \mathbf{C}$

Let $c \in \mathbf{C}$ and suppose that $\rho_c(X) > r \geq 0$. Then $a \pm r \in \mathcal{R}(L_X)$, where $a = \Re c$, and so by Beppo Levi's theorem (3.7) we have

$$\sum_{n=0}^{\infty} \frac{1}{n!} E|e^{cX} X^n (z-c)^n| = E\, e^{(a+|z-c|)X}$$
$$\leq E\, e^{(a+r)X} + E\, e^{(a-r)X} < \infty$$

for all $z \in \mathbf{C}$ with $|z-c| \leq r$. Hence by, (3.32.3) we find

$$(4.19.15) \qquad L_X(z) = \sum_{n=0}^{\infty} \frac{(z-c)^n}{n!} E(X^n e^{cX}) \qquad \forall\, |z-c| < \rho_c(X)$$

(4.19.16) L_X is analytic in $\mathcal{D}^\circ(L_X)$

$$(4.19.17) \qquad L_X^{(n)}(c) = E(X^n e^{cX}) \qquad \forall\, c \in \mathcal{D}^\circ(L_X)$$

[Recall that a function f from an open connected set $U \subseteq \mathbf{C}$ into $\mathbf{C}$ is *analytic* in U if f can be expanded in a power series of $(z-c)$ in a neighborhood of c for all $c \in U$.] Note that the two first derivatives of cumulant transform $\kappa_X(t) = \log L_X(t)$ are given by

$$\kappa'_X(t) = \frac{L'_X(t)}{L_X(t)} \quad , \quad \kappa''_X(t) = \frac{L''_X(t)}{L_X(t)} - \left(\frac{L'_X(t)}{L_X(t)}\right)^2$$

whenever t is an interior point of the real domain $\mathcal{R}(L_X)$. In particular, if 0 is an interior point of $\mathcal{D}(L_X)$, then by (4.19.17) and (4.4.3) we have

$$(4.19.18) \qquad L_X(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \mu_X(n) \qquad \forall\, |z| < \rho_0(X)$$

$$(4.19.19) \qquad EX = \kappa'_X(0) \quad , \quad \mathrm{var}(X) = \kappa''_X(0) \quad , \quad EX^n = L_X^{(n)}(0) \quad \forall\, n \in \mathbf{N}$$

Applying (3.39.5) and (3.39.8) on the measure $\mu(d\omega) = e^{X(\omega)\Re c} P(d\omega)$, it follows easily that we have

$$(4.19.20) \qquad \frac{1}{\rho_c(X)} = R\big(Xe^{X\Re c}\big) = \limsup_{n\to\infty} \Big\{ \tfrac{1}{(2n)!} E\big(X^{2n} e^{X\Re c}\big) \Big\}^{\frac{1}{2n}} \qquad \forall\, c \in \mathbf{C}$$

In particular, we see that $\rho(X) = \rho_0(X)$.

Let U be an open connected subset of $\mathbf{C}$ and let f and g be two analytic functions on U . One of the fundamental theorems of complex analysis states that if $f(t) = g(t)$ for all $t \in T$, where T is a subset of U having a condensation point belonging to U , then $f = g$. [Recall that t_0 is a *condensation point* of T

if every neighborhood of t_0 contains at least one point in $T \setminus \{t_0\}$.] Using this fact we obtain the following striking uniqueness theorem for the one-dimensional Laplace transform:

(4.19.21) Let X and Y be real random variables, and let T be a subset of $\mathcal{D}(L_X) \cap \mathcal{D}(L_Y)$ such that T has a condensation point in $\mathcal{D}^\circ(L_X) \cap \mathcal{D}^\circ(L_Y)$. Then $L_X(t) = L_Y(t)$ for all $t \in T$ if and only if $X \sim Y$

Let μ be a Borel measure on $\mathbf{R}^\mathbf{k}$. Then the L-transform L_μ of μ is called *the Laplace transform* of μ and usually denoted $\tilde{\mu}(t)$, that is,

$$\tilde{\mu}(t) := \int_{\mathbf{R}}^{k} e^{s \bullet t}\, \mu(ds) \qquad \forall t \in \mathcal{D}(\tilde{\mu}) \cup \mathbf{R}^k \tag{4.19.22}$$

where *the complex domain* $\mathcal{D}(\tilde{\mu})$ is defined as above. Similarly, if $f : \mathbf{R}^k \to \mathbf{C}$ is a measurable function, then we define *the Laplace transform* of f by

$$\tilde{f}(t) := \int_{\mathbf{R}^k} f(s) e^{s \bullet t}\, ds \qquad \forall t \in \mathcal{D}(\tilde{f}) \tag{4.19.23}$$

where *the complex domain* $\mathcal{D}(\tilde{f})$ is defined as above. The Laplace transform of measures and functions plays an important role in analysis, and it is pretty clear that all the propositions above carries over to Laplace transforms of measures and function, except that 0 need not belong to the domain, and even when it does the Laplace transform evaluated at 0 need not be 1 .

4.20 Moment Transforms

Let $\mu : \mathbf{R} \times \mathbf{N}_0 \to \mathbf{R}$ be given by $\mu(s,n) = s^n$ with the convention: $0^0 = 1$. Then μ is a kernel on $\mathbf{R} \times \mathbf{N}_0$, and the μ-transform μ_X of X is nothing but *the moment transform* as defined in (4.3). Let $M : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}$ be given by $M(s,t) = s^t$ with the convention: $0^0 = 1$. Then M is a kernel on $\mathbf{R}_+ \times \mathbf{R}_+$, and if X is a random variable, then the M-transform $M_{|X|}$ of $|X|$ is nothing but *the absolute moment transform* of X as defined in (4.3). Substituting $t = e^s$ in (4.3.24), we find

$$m_X(q) = q \int_{-\infty}^{\infty} e^{qs}\, P(|X| > e^s)\, ds$$

Thus, $q^{-1}m_X(q)$ is the Laplace transform of the function $f(s) = P(|X| > e^s)$. Hence, by (4.19.21) we obtain the following uniqueness theorem for the absolute moment transform:

(4.20.1) If $m_X(q) = m_Y(q) < \infty$ for all $q \in T$ and T has a condensation point in the interior of $\mathcal{D}(m_X) \cap \mathcal{D}(m_Y)$, then $|X| \sim |Y|$

Setting $n^* := n$ and $n_0 \nabla n_1 = n_0 + n_1$, we see that $\mu(s,n)$ satisfies (3.41.1) and (3.41.2). Hence, letting $T_0 = \{1\}$ and $T_1 = \mathbf{N}$, we find from (4.15)

(4.20.2) If $R(X) < \infty$ and $\mu_X(n) = \mu_Y(n)$ for all $n \in \mathbf{N}$, then $X \sim Y$

where $R(X)$ is defined by (4.19.13). By (4.15.6), we have that

$$R(X) < \infty \Rightarrow \sum_{n=1}^{\infty} (E|X|^n)^{-1/n} = \infty \tag{4.20.3}$$

By the remark to (4.15) [see (4.15.7)], we have that (4.20.2) remains valid if we replace the condition $R(X) < \infty$ by the slightly weaker condition

$$\sum_{n=1}^{\infty} (E|X|^n)^{-1/n} = \infty \tag{4.20.4}$$

However, there do exist decent random variables with moments of all orders which are not μ-determined on $\mathbf{N}$. For instance, the log-normal distribution $\log N(\mu, \sigma^2)$ and the hyper-exponential distribution $H\exp(\alpha, \beta, \gamma)$ with $0 < \gamma < 1$ have moments of all orders, but they are not μ-determined on $\mathbf{N}$ [see (4.25) and (4.26)].

In study of integer-valued random variables, it is often useful to consider the so-called factorial moment transform. If $n \in \mathbf{N}_0$ we define *the* n^{th} *factorial* of x, denoted $x^{(n)}$, as follows:

$$x^{(0)} = 1 \quad , \quad x^{(n)} = x(x-1)\cdots(x-n+1) \quad \text{if} \quad n \geq 1 \tag{4.20.5}$$

Then $x^{(n)}$ is a polynomial of degree n and $\nu(s,n) = s^{(n)}$ is a kernel on $\mathbf{R} \times \mathbf{N}_0$. The ν-transform ν_X of a random variable X is called *the factorial moment transform* of X:

$$\nu_X(n) = EX^{(n)} \qquad \forall n \in \mathcal{D}(\nu_X) \tag{4.20.6}$$

Since $x^{(n)}$ is a polynomial of exact degree n with leading coefficient 1, we know that there exists coefficients $A_{ij} \in \mathbf{R}$ such that $A_{nn} = 1$ and

$$x^{(n)} = \sum_{j=0}^{n} A_{nj}\, x^j \quad , \quad x^n = \sum_{j=0}^{n} A_{jn}\, x^{(j)}$$

and so $\mathcal{D}(\nu_X) = \mathcal{D}(\mu_X)$ and

$$\mu_X(n) = \sum_{j=0}^{n} A_{jn}\, \nu_X(j) \ , \ \nu_X(n) = \sum_{j=0}^{n} A_{nj}\, \mu_X(j) \quad \forall n \in \mathcal{D}(\mu_X) \tag{4.20.7}$$

In Exc.4.32 you will find formulas for computing the coefficients A_{ij} .

4.21 Generating Functions

Let $r : \mathbf{N}_0 \times \mathbf{R} \to \mathbf{R}$ be given by $r(n,t) = t^n$, that is, $r = \mu^*$ where μ is the kernel defined in (4.20). Then r is a kernel on $\mathbf{N}_0 \times \mathbf{R}$, and the r-transform r_X of an $\mathbf{N}_0$-valued random variable X is called *the real generating function* of X :

$$r_X(t) = Et^X \qquad \forall t \in \mathcal{D}^*(r_X) \tag{4.21.1}$$

where $\mathcal{D}^*(r_X)$ is the set of all $t \in \mathbf{R}$ for which the expectation exists, but possibly equal to $\pm\infty$. The set $\mathcal{D}^*(r_X)$ is called *the extended domain* of r_X , and as usual we let *the domain* $\mathcal{D}(r_X)$ be the set of all $t \in \mathbf{R}$ for which the expectation exists and is finite. Note that if X is an $\mathbf{N}_0$-valued random variable, then by (4.10.10) we have

$$r_X(t) = \sum_{j=1}^{\infty} t^n\, P(X = n) \qquad \forall t \in \mathcal{D}(r_X) \tag{4.21.2}$$

If ρ is the radius of convergence of the power series in (4.21.2), then evidently $\rho \geq 1$ and so we have

$$]-\rho,\rho[\, \subseteq \mathcal{D}(r_X) \subseteq [-\rho,\rho] \quad , \quad 1 \in \mathcal{D}(r_X) \tag{4.21.3}$$

$$r_X(0) = P(X = 0) \quad , \quad r_X(1) = 1 \tag{4.21.4}$$

$$P(X = n) = \frac{r_X^{(n)}(0)}{n!} \tag{4.21.5}$$

Setting $t^* := t$ and $t_0 \nabla t_1 = t_0 t_1$, we see that (3.41.1) and (3.41.2) hold. Since $|t^n| \leq 1$ for $0 < t < 1$, we find that every integer-valued random variable X is r-determined on $]0,1[$. Also note that by (4.14.5) - (4.14.8) we have

(4.21.6) $\quad r_{X+Y}(t) = r_X(t)\, r_Y(t) \qquad \forall t \in \mathcal{D}(r_X) \cap \mathcal{D}(r_Y) \quad \text{if } X \underline{\|} Y$

(4.21.7) $\quad \mu_{UV}(n) = \mu_U(n)\, \mu_V(n) \qquad \forall n \in \mathcal{D}(\mu_U) \cap \mathcal{D}(\mu_V) \quad \text{if } U \underline{\|} V$

(4.21.8) $\quad E r_X(W) = E \mu_W(X) \quad \text{if } W^X \in L(P)$

whenever X and Y are $\mathbf{N}_0$-valued random variables, and U, V and W are real random variables.

Let $q : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}$ be given by $q(s,t) = t^s$; that is, $q = M^*$ where M is the kernel defined in (4.20). Then q is a kernel on $\mathbf{R}_+ \times \mathbf{R}_+$, and the q-transform q_X of an non-negative random variable X is called *the generating function* of X :

(4.21.9) $\quad q_X(t) = E t^X \quad \forall t \in \mathbf{R}_+ \quad , \quad \mathcal{D}(q_X) = \{\, t \geq 0 \mid q_X(t) < \infty \,\}$

Note that if X is $\mathbf{N}_0$-valued, then we have

(4.21.10) $\quad \mathcal{D}(q_X) = \mathbf{R}_+ \cap \mathcal{D}(r_X)$

(4.21.11) $\quad q_X(t) = r_X(t) \qquad \forall t \in \mathcal{D}(q_X)$

As above, we find

(4.21.12) $\quad [0,1] \subseteq \mathcal{D}(q_X) \quad , \quad q_X(0) = P(X = 0) \quad , \quad q_X(1) = 1$

(4.21.13) $\quad q_X$ is continous on $\mathcal{D}(q_X)$

(4.21.14) $\quad q_X(t) = q_Y(t) \quad \forall 0 < t < 1 \iff X \sim Y$

(4.21.15) $\quad q_{X+Y}(t) = q_X(t)\, q_Y(t) \qquad \forall t \geq 0 \quad \text{if } X \underline{\|} Y$

(4.21.16) $\quad m_{XY}(s) = m_X(s)\, m_Y(s) \qquad \forall s \geq 0 \quad \text{if } X \underline{\|} Y$

(4.21.17) $\quad E q_X(Y) = E m_Y(X)$

whenever X and Y are non-negative random variables. Note that $q(s,\cdot)$ is infinitely often differentiable on $]0,\infty[$ for all $s \geq 0$ and

$$\frac{\partial^n}{\partial t^n} q(s,t) = s^{(n)}\, t^{s-n} \qquad \forall t > 0\ \forall s \geq 0$$

$$\sup_{s \geq 0} |s^{(n)}\, t^{s-n}| < \infty \qquad \forall 0 < t < 1$$

$$|s^{(n)}\, t^{s-n}| \leq 2^n |s^{(n)}| \qquad \forall \tfrac{1}{2} \leq t \leq 1\ \forall s \geq 0$$

where $s^{(n)}$ is the nth factorial of s [see (4.20.4)]. Hence, if ν_X is the factorial moment of X [see (4.20.5)], then by (3.31) we have

(4.21.18) q_X is infinitely often differentiable on $]0,1[$

(4.21.19) $q_X^{(n)}(t) = E\{X^{(n)}t^{X-n}\} \qquad \forall 0 < t < 1 \qquad \forall n \in \mathbf{N}_0$

(4.21.20) $q_X^{(n)}(1) = \nu_X(n) \qquad \forall n \in \mathcal{D}(\nu_X)$

4.22 The Normal Distribution

The normal distribution $N(\mu, \sigma^2)$, where $\mu \in \mathbf{R}$ and $\sigma^2 > 0$, is given by the density function

(4.22.1) $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad \forall x \in \mathbf{R}$$

Moreover, we set $N(\mu, 0) := \delta_\mu$ where δ_μ is the Dirac measure [see (1.15)]. Let $X \sim N(\mu, \sigma^2)$ if $\sigma^2 = 0$, then clearly $L_X(t) = e^{\mu t}$, and if $\sigma^2 > 0$, then

$$L_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(tx - \frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

Substituting $x = y + t\sigma^2 + \mu$ gives

$$\begin{aligned} L_X(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(ty + t^2\sigma^2 + \mu t - \frac{(y+t\sigma^2)^2}{2\sigma^2}\right) dy \\ &= \frac{\exp(t\mu + \frac{1}{2}t^2\sigma^2)}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2}\, dy \\ &= e^{t\mu + \frac{1}{2}t^2\sigma^2} \end{aligned}$$

Hence, the Laplace transform of $X \sim N(\mu, \sigma^2)$ is given by:

(4.22.2) $$L_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2} \qquad \forall t \in \mathbf{C}$$

Formally speaking, the substitution: $x = y + t\sigma^2 + \mu$ is only allowed when t is real. However, the substitution for complex numbers t may be justified by methods from complex analysis. In particular, we see that the characteristic function of $X \sim N(\mu, \sigma^2)$ is given by

(4.22.3) $$\varphi_X(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2} \qquad \forall t \in \mathbf{R}$$

Substituting $v = (x-\mu)^2/(2\sigma^2)$, we get

$$m_X^0(q) = E|X-\mu|^q = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-\mu|^q \, e^{-(x-\mu)^2/(2\sigma^2)} \, dx$$
$$= \sigma^q \, 2^{q/2} \, \pi^{-1/2} \int_0^{\infty} v^{(q-1)/2} \, e^{-v} dv$$
$$= \sigma^q \, 2^{q/2} \, \pi^{-1/2} \, \Gamma(\tfrac{q+1}{2})$$

Thus, the absolute central moment transform of $X \sim N(\mu, \sigma^2)$ is given by

$$m_X^0(q) = \sqrt{\frac{2^q \, \sigma^{2q}}{\pi}} \, \Gamma\Big(\frac{q+1}{2}\Big) \qquad \forall q \geq 0 \tag{4.22.4}$$

Hence, by (4.19.19) and (4.4.3) we find

$$EX = \mu \quad , \quad \mathrm{var}(X) = \sigma^2 \tag{4.22.5}$$

By antisymmetry we have $\mu_X^0(n) = 0$ if n is odd, and by symmetry we have $\mu_X^0(n) = m_X^0(n)$ if n is even. Hence, by (4.22.4) and the well-known formulas

$$\Gamma(x) = (x-1)\Gamma(x-1) \quad \forall x > 1 \quad , \quad \Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

we find the following expressions for the central moments of $X \sim N(\mu, \sigma^2)$:

$$\mu_X^0(2n-1) = 0 \qquad \forall n \in \mathbf{N} \tag{4.22.6}$$

$$\mu_X^0(2n) = \sigma^{2n} \cdot 1 \cdot 3 \cdot 5 \cdots (2n-1) \qquad \forall n \in \mathbf{N} \tag{4.22.7}$$

Let $X = (X_1, \ldots, X_k)$ be a k-dimensional random vector, let $\mu \in \mathbf{R}^k$ be a given vector, and let $\Sigma = \{\sigma_{ij}\}$ be a given symmetric non-negative definite $(k \times k)$-matrix. Then we say that X has a *k-dimensional normal distribution* $N_k(\mu, \Sigma)$ if $EX = \mu$, $VX = \Sigma$ and $a_1 X_1 + \cdots + a_k X_k$ has a normal distribution for all $a_1, \ldots, a_k \in \mathbf{R}$. Suppose that $X \sim N_k(\mu, \Sigma)$. Then by (4.4.16) we have

$$E(X \bullet a) = \mu \bullet a \quad , \quad \mathrm{var}\,(X \bullet a) = a \bullet (a\Sigma) = a \, \Sigma \, a^t$$

Hence, $(X \bullet a) \sim N(\mu \bullet a, a \, \Sigma \, a^t)$ and so we find the following formulas for the Laplace transform and the characteristic function of $X \sim N_k(\mu, \Sigma)$:

$$L_X(z) = \exp\Big\{(\mu \bullet z) + \tfrac{1}{2} z \, \Sigma \, z^t\Big\} \qquad \forall z \in \mathbf{C}^k \tag{4.22.8}$$

$$\varphi_X(u) = \exp\Big\{i\,(\mu \bullet u) - \tfrac{1}{2} u \, \Sigma \, u^t\Big\} \qquad \forall u \in \mathbf{R}^k \tag{4.22.9}$$

Let X and Y be two independent k-dimensional random vectors such that $X \sim N_k(\mu, \Sigma)$ and $Y \sim N_k(\lambda, \Upsilon)$. Then by (4.18.7) we have

$$\varphi_{X+Y}(u) = \varphi_X(u)\,\varphi_Y(u) = e^{i(\mu+\lambda)\bullet u - \frac{1}{2}u\,(\Sigma+\Upsilon)\,u^t}$$

Thus, by (4.18.8) we see that we have

(4.22.10) If $X \sim N_k(\mu, \Sigma)$, $Y \sim N_k(\lambda, \Upsilon)$ and X and Y are independent, , then $X + Y \sim N_k(\mu + \lambda, \Sigma + \Upsilon)$

Let $X = (X_1, \ldots, X_k) \sim N_k(\mu, \Sigma)$ where $\mu = (\mu_1, \ldots, \mu_k)$ and $\Sigma = \{\sigma_{ij}\}$. Then by definition we have

(4.22.11) $EX_j = \mu_j \qquad \forall j = 1, \ldots, k$

(4.22.12) $\operatorname{cov}(X_i, X_j) = \sigma_{ij} \qquad \forall i, j = 1, \ldots, k$

(4.22.13) $X_j \sim N(\mu_j, \sigma_{jj}) \qquad \forall j = 1. \ldots, k$

Suppose that $X_1, \ldots, X_k$ are uncorrelated. Then $\sigma_{ij} = 0$ for all $i \neq j$. Hence, by (4.22.9) we have

$$\varphi_X(t_1, \ldots, t_k) = \prod_{j=1}^{k} e^{i\mu_j t_j - \frac{1}{2} t_j^2 \sigma_{jj}}$$

So by (4.18.11) and (4.4.19) we have

(4.22.14) If $X = (X_1, \ldots, X_k) \sim N_k(\mu, \Sigma)$, then $X_1, \ldots, X_k$ are independent if and only if $X_1, \ldots, X_k$ are uncorrelated

Similarly, we find

(4.22.15) If X_j is a k_j-dimensional random vector for $j = 1, \ldots, q$ such that $X = (X_1, \ldots, X_q)$ has a k-dimensional normal distribution $N_k(\mu, \Sigma)$, where $k = k_1 + \cdots + k_q$, then $X_1, \ldots, X_k$ are independent if and only if $\operatorname{Cov}(X_i, X_j) = 0$ for all $i \neq j$

Let $X \sim N_k(\mu, \Sigma)$, let $\xi \in \mathbf{R}^q$, and let A be a $(k \times q)$-matrix. Inserting the characteristic function φ_X from (4.22.9) in formula (4.18.5) gives

(4.22.16) If $X \sim N_k(\mu, \Sigma)$, then $\xi + XA \sim N_q(\xi + \mu A, A^t \Sigma A)$

Suppose that Σ is invertible. Since Σ is positive definite, then by linear algebra there exists an invertible $(k \times k)$-matrix A with $A^t \Sigma A = I$, where I is the identity matrix. By (4.22.16) we have that $U = (X - \mu)A \sim N_k(0, I)$. Hence, if $U = (U_1, \ldots, U_k)$, then $U_1, \ldots, U_k$ are independent $N(0,1)$-distributed random variables. Thus, by (4.5.11) we have that U has density:

$$g(s) = \prod_{j=1}^{k} (2\pi)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2} s_j^2\right) = (2\pi)^{-\frac{k}{2}} \exp\left(-\tfrac{1}{2} \|s\|^2\right)$$

and since $X = \mu + UA^{-1}$, then by the usual transformation theorem [see (8.2.2) in Volume II], we see that X has density

$$f(u) = \frac{g((u-\mu)A)}{|\det A^{-1}|}$$

Moreover, since $A^t \Sigma A = I$ we have $\Sigma^{-1} = AA^t$, $|\det A^{-1}| = \sqrt{\det \Sigma}$ and

$$||(u-\mu)A||^2 = (u-\mu)\, AA^t\, (u-\mu)^t = (u-\mu)\, \Sigma^{-1}\, (u-\mu)^t$$

Hence, if Σ is invertible and $X \sim N_k(\mu, \Sigma)$, then X is absolute continuous with density given by

$$f(u) = \frac{1}{\sqrt{(2\pi)^k \det \Sigma}} \exp\left\{-\tfrac{1}{2}(u-\mu)\Sigma^{-1}(u-\mu)^t\right\} \quad \forall u \in \mathbf{R}^k \tag{4.22.1}$$

However, if Σ is **not** invertible, then X is **not** absolutely continuous. In Section (8.24) of Volume II, I shall show that X has density with respect to the q-dimensional Hausdorff volume λ_{kq} [see (8.3) in Volume II] where $q = \operatorname{rank}(\Sigma)$.

The normal distribution is unquestionably the most important distribution of probability theory and statistics. It was first found by Abraham de Moivre (1667–1754), but in a rather indirect way. In his book *Doctrine of Chances* (second edition, 1733) de Moivre shows that if $\Lambda_n(s) = \{j \mid \frac{n}{2} \le j \le \frac{n}{2} + s\sqrt{n}\}$, then

$$\lim_{n\to\infty} \sum_{j \in \Lambda_n(s)} \binom{n}{j} 2^{-n} =$$

$$\frac{2}{B}\left(s - \frac{2s^3}{3} + \frac{4s^5}{5 \cdot 2} - \frac{8s^7}{7 \cdot 6} + \frac{16s^9}{9 \cdot 24} - \frac{32s^{11}}{11 \cdot 120} + \&c\right)$$

In the third edition of "Doctrine of Chances" (1756), de Moivre writes

> *When I first began that inquiry, I contended myself to determine at large the value of B ... I desisted from proceeding farther till my worthy and learned friend Mr. James Stirling, who applied himself after me to that inquiry, found that the quantity B did denote the square-root of the circumference of a circle whose radius is unity*

That is, Stirling showed that $B = \sqrt{2\pi}$. Actually, Stirling did that by first proving his famous formula [see Exc.1.7(xxiii)]. The series that de Moivre derived is actually equal to

$$\sqrt{\tfrac{2}{\pi}} \int_0^s e^{-2x^2}\, dx = P(0 \le U \le s)$$

where $U \sim N(0, \frac{1}{4})$. However, de Moivre never found the integral expression and he considered the series as a useful approximation of a binomial probability, but he did not realize that he had derived what was to become the most important distribution of probability theory and statistics.

In 1808, the American surveyor Robert Adrain (1775–1843) published a remarkable paper in the Philadelphia journal *The Analyst or Mathematical Museum* bearing the title "Research concerning the probabilities of errors, which happen in making observations." In his paper Adrain suggested to use the normal distribution as a description of errors in measurements. One year later, Karl Friederich Gauss (1777-1855) did the same thing. At that time the United States was far from the mathematical circles that mattered, and Adrain's paper passed unnoticed until a few years ago. And since Gauss was the leading mathematician of his times, the normal distribution is also called *the Gauss distribution.*

Adrain gives six justifications for the normal distribution, many of which are very insufficient indeed. But one of these is interesting because it determines the normal distribution uniquely. I shall now describe the reasoning of Adrain, not in his own language, which is rather fogged, but in a modern language.

We want to measure the distance from the origin $O \in \mathbf{R}^2$ to a point $A \in \mathbf{R}^2$. To do this we place one leg of a compasses at A and try to hit O . But, of course, we don't hit O exactly; but hit a random point (X, Y) . Adrain then maintains that it is clear that X and Y are independent with the same density function f , and that points in the same distance from O have the same probability of being hit. This means that the simultaneous density $g(x, y)$ of (X, Y) is a function, say h , of $\sqrt{x^2 + y^2}$; that is,

$$g(x, y) = f(x)f(y) = h\Big(\sqrt{x^2 + y^2}\Big) \tag{4.22.18}$$

Let $k(x) = \log f(\sqrt{x}) - \log f(0)$. Then

$$\begin{aligned} k(x) + k(y) &= \log\Big(f\big(\sqrt{x}\big)\, f\big(\sqrt{y}\big)\Big) - 2\log f(0) \\ &= \log h\big(\sqrt{x+y}\big) - 2\log f(0) \end{aligned}$$

Setting $y = 0$ in (4.22.18) gives $h(x) = f(x)\, f(0)$ and so

$$k(x) + k(y) = \log f\big(\sqrt{x+y}\big) - \log f(0) = k(x+y)$$

But it is well-known that the equation $k(x+y) = k(x) + k(y)$ is only satisfied by linear functions (provided k is measurable). Hence, $k(x) = -\frac{x}{2\tau}$, where $\tau = -\frac{1}{2}k(1)^{-1}$, and since

$$f(x) = f(0)\, e^{k(x^2)} = f(0)\, e^{-x^2/2\tau} \quad , \quad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

we see that $\tau > 0$ and $f(0) = \frac{1}{\sqrt{2\pi\tau}}$. Hence, f is the density of the $N(0,\tau)$-distribution. Thus, the only density $g(x,y)$ satisfying (4.22.18) is a two–dimensional normal distribution $N_2(0, \tau I_2)$ where I_2 is the (2×2)-identity matrix.

4.23 The Gamma Distribution

The Gamma distribution $\Gamma(\alpha,\beta)$, where $\alpha,\beta > 0$, is given by the density function

$$(4.23.1) \qquad f(x) = \begin{cases} \dfrac{x^{\alpha-1}}{\beta^\alpha\,\Gamma(\alpha)}\, e^{-x/\beta} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

Suppose that $X \sim \Gamma(\alpha,\beta)$; then

$$\begin{aligned} m_X(q) &= \beta^{-a}\,\Gamma(\alpha)^{-1} \int_0^\infty x^{q+\alpha-1}\, e^{-x/\beta}\, dx \\ &= \beta^q\,\Gamma(\alpha)^{-1} \int_0^\infty y^{q+\alpha-1}\, e^{-y}\, dy \\ &= \beta^q\,\Gamma(q+\alpha)\,\Gamma(\alpha)^{-1} \end{aligned}$$

Hence, the absolute moment transform of $X \sim \Gamma(\alpha,\beta)$ is given by

$$(4.23.2) \qquad m_X(q) = \beta^q\, \frac{\Gamma(q+\alpha)}{\Gamma(\alpha)} \qquad \forall q \ge 0$$

Using (3.37.5) we find

$$(4.23.3) \qquad EX = \alpha\beta \ , \ \mathrm{var}(X) = \alpha\beta^2 \ , \ EX^n = \beta^n(\alpha+n-1)^{(n)} \quad \forall n \in \mathbf{N}_0$$

By (3.37.7), we see that the Laplace transform and the characteristic function of $X \sim \Gamma(\alpha, b)$ are given by

(4.23.4) $\mathcal{D}(L_X) = \{z \in \mathbf{C} \mid \Re z < \frac{1}{\beta}\}$

(4.23.5) $L_X(z) = (1 - z\beta)^{-\alpha}$ if $\Re z < \frac{1}{\beta}$

(4.23.6) $\varphi_X(t) = (1 - it\beta)^{-\alpha} \quad \forall t \in \mathbf{R}$

Let $X \sim \Gamma(\alpha, \beta)$ and $Y \sim \Gamma(\gamma, \beta)$, and suppose that X and Y are independent. Then by (4.18.7) we have that

$$\begin{aligned}\varphi_{X+Y}(t) &= \varphi_X(t)\,\varphi_X(t) = (1 - it\beta)^{-\alpha}(1 - it\beta)^{-\gamma} \\ &= (1 - it\beta)^{-(\alpha+\gamma)}\end{aligned}$$

Thus, by (4.18.8) we have

(4.23.7) If X and Y are independent, $X \sim \Gamma(\alpha, \beta)$ and $Y \sim \Gamma(\gamma, \beta)$, then $X + Y \sim \Gamma(\alpha + \gamma, \beta)$

Note that in formulas (4.23.5+6), we raise a complex number to the power $-\alpha$. If $c \in \mathbf{C}$ and $q \in \mathbf{R}$, then in general there exists several q^{th} roots of c , and the formulas are only correct if we use the main root, that is, [see (3.37.8)],

(4.23.8) $c^q := |c|^q\, e^{iq\theta}$ where $-\pi < \theta \le \pi$ and $c = |c|\, e^{i\theta}$

Suppose that $\beta > 0$, $t \in \mathbf{R}$ and $\theta = -\arctan(t\beta)$, then $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ and $1 - it\beta = |1 - it\beta|\, e^{i\theta}$. Hence, the characteristic function of $X \sim \Gamma(\alpha, b)$ is given by

$$\begin{aligned}\varphi_X(t) &= \left(1 + t^2\beta^2\right)^{-\alpha/2} e^{i\,\alpha\,\arctan(t\beta)} \\ &= \left(1 + t^2\beta^2\right)^{-\alpha/2} \left\{\cos\left(\alpha \arctan(t\beta)\right) + i \sin\left(\alpha \arctan(t\beta)\right)\right\}\end{aligned}$$

Note that $\Im\varphi_X(\xi) = E\sin(\xi X)$, $\Re\varphi_X(\xi) = E\cos(\xi X)$ and that $\sin(m\pi) = \cos(\frac{\pi}{2}k) = 0$ whenever m is an integer and k is an odd integer. Setting $\beta = \frac{1}{\lambda}$ and $\alpha = \frac{\pi}{v} m$ [respectively $\alpha = \frac{\pi}{2v}(2m - 1)$], we find

(4.23.9) If $\lambda > 0$, $0 < v < \frac{\pi}{2}$ and $\xi = \tan v$, then we have

$$\int_0^\infty x^{(\pi m/v)-1}\, e^{-\lambda x} \sin(\lambda\xi x)\, dx = 0 \quad \forall m \in \mathbf{N}$$

$$\int_0^\infty x^{(\pi k/2v)-1}\, e^{-\lambda x} \cos(\lambda\xi x)\, dx = 0 \quad \forall k \in \mathbf{N}\ ,\ k \text{ odd}$$

4.24 The Poisson Distribution

The Poisson distribution $Po(\lambda)$, where $\lambda > 0$, is given by the probability mass function

$$(4.24.1) \qquad p(n) = \begin{cases} \dfrac{\lambda^n e^{-\lambda}}{n!} & \text{if } n = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Let $X \sim Po(\lambda)$, then

$$L_X(t) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{nt} e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = \exp(\lambda e^t - \lambda)$$

$$r_X(t) = \sum_{n=0}^{\infty} t^n \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda + \lambda t}$$

Hence, if $X \sim Po(\lambda)$, then by (4.21.11) and (4.21.20) we find

(4.24.2) $\quad L_X(t) = \exp(\lambda e^t - \lambda) \quad \forall t \in \mathbf{C}$

(4.24.3) $\quad \varphi_X(t) = \exp(\lambda e^{it} - \lambda) \quad \forall t \in \mathbf{R}$

(4.24.4) $\quad r_X(t) = e^{-\lambda + \lambda t} \quad \forall t \in \mathbf{R}$

(4.24.5) $\quad \nu_X(n) = \lambda^n \quad \forall n \in \mathbf{N}_0 \quad , \quad EX = \operatorname{var} X = \lambda$

(4.24.6) If X and Y are independent, $X \sim Po(\lambda)$ and $Y \sim Po(\mu)$ then $X + Y \sim Po(\lambda + \mu)$

4.25 The log-Normal Distribution

The log-*normal distribution* $\log N(\mu, \sigma^2)$, where $\mu \in \mathbf{R}$ and $\sigma^2 > 0$, is given by the density function

$$(4.25.1) \qquad f(x) = \begin{cases} \dfrac{1}{x\,\sigma\sqrt{2\pi}} \exp\left\{ -\dfrac{(\mu - \log x)^2}{2\sigma^2} \right\} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

From the usual transformation theorem [see (4.2.6)], we find

(4.25.2) $\quad X \sim \log N(\mu, \sigma^2) \iff \log X \sim N(\mu, \sigma^2)$

(4.25.3) $\quad Y \sim N(\mu, \sigma^2) \iff e^Y \sim \log N(\mu, \sigma^2)$

Let $X \sim \log\ N(\mu, \sigma^2)$, then $Y = \log X \sim N(\mu, \sigma^2)$ and $X = e^Y$. Hence, by (4.22.2) we have

$$m_X(q) = EX^q = Ee^{qY} = L_Y(q) = e^{q\mu + \frac{1}{2}q^2\sigma^2}$$

Thus, the moment transform of $X \sim \log N(\mu, \sigma^2)$ is given by

$$EX^q = e^{q\mu + \frac{1}{2}q^2\sigma^2} \qquad \forall q \geq 0 \tag{4.25.4}$$

$$EX = e^{\mu + \frac{1}{2}\sigma^2} \quad , \quad \mathrm{var}(X) = \left(e^{\sigma^2} - 1\right)(EX)^2 \tag{4.25.5}$$

Let X have a $\log N(\mu, \sigma^2)$ distribution and let $f(x)$ be density of X given by (4.25.1). Set $U = (\log X) - \mu$, then $U \sim N(0, \sigma^2)$ and $X^n = e^{n(\mu + U)}$. Let $n \in \mathbf{Z}$. Then we have

$$\begin{aligned}
&E\{X^n \sin(\pi\sigma^{-2}(\log X - \mu))\} = E\{e^{n(\mu+U)} \sin(\pi\sigma^{-2}U)\} \\
&= e^{n\mu} E\Im \exp((n + i\pi\sigma^{-2})U)\} = e^{n\mu}\Im L_U(n + i\pi\sigma^{-2}) \\
&= e^{n\mu}\Im \exp(\tfrac{1}{2}\sigma^2(n + i\pi\sigma^{-2})^2) \\
&= \sin(n\pi)\exp(n\mu + \tfrac{1}{2}\sigma^2(n^2 - \pi^2\sigma^{-4})) = 0
\end{aligned}$$

Hence, we see that

$$\int_0^\infty x^n f(x) \sin(\pi\sigma^{-2}(\log x - \mu))dx = 0 \qquad \forall n \in \mathbf{Z}$$

Let $-1 \leq \alpha \leq 1$, and define

$$f_\alpha(x) = \begin{cases} f(x)\left\{1 - \alpha \sin\left[\pi\,\sigma^{-2}(\mu - \log x)\right]\right\} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Then $f_\alpha \geq 0$ and by the computations above we have

$$\int_0^\infty x^n f_\alpha(x)\,dx = EX^n \qquad \forall n \in \mathbf{Z} \qquad \forall \alpha \in [-1, 1] \tag{4.25.6}$$

Setting $n = 0$ we see that f_α is a density for all $\alpha \in [-1, 1]$. If $Z \sim f_\alpha$, then $EZ^n = EX^n$ for all $n \in \mathbf{Z}$. Hence, Z and X have the same moments; but if $\alpha \neq 0$ then Z and X have different distributions.

Thus, the log-normal distribution $\log N(\mu, \sigma^2)$ have moments of all orders, but it is **not** uniquely determined by its moments.

4.26 The Hyper-Exponential Distribution

The one-sided hyper-exponential distribution $H\exp^+(\alpha,\beta,\gamma)$, where α , β and γ are positive numbers, is given by the density function

$$(4.26.1)\qquad f(x) = \begin{cases} \gamma\,\beta^{-\alpha/\gamma}\,\Gamma\left(\frac{\alpha}{\gamma}\right)^{-1} x^{\alpha-1}\exp\left(-\frac{x^\gamma}{\beta}\right) & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

By the usual transformation theorem [see (4.2.6)], we have

$$(4.26.2)\qquad X \sim H\exp^+(\alpha,\beta,\gamma) \iff X^\lambda \sim H\exp^+(\tfrac{\alpha}{\lambda},\beta,\tfrac{\gamma}{\lambda})$$

for all $\lambda > 0$. The gamma distribution, $\Gamma(\alpha,\beta)$, and the exponential distribution, $E(\lambda)$, are both special cases of the hyper-exponential distributions:

$$(4.26.3)\qquad \Gamma(\alpha,\beta) = H\exp^+\left(\alpha,\beta,1\right)$$

$$(4.26.4)\qquad E(\lambda) = \Gamma\left(1,\tfrac{1}{\lambda}\right) = H\exp^+\left(1,\tfrac{1}{\lambda},1\right)$$

If $X \sim H\exp^+(\alpha,\beta,\gamma)$, then by (4.26.2+3) we have $Y = X^\gamma \sim \Gamma(\frac{\alpha}{\gamma},\beta)$, and since $X^q = Y^{q/\gamma}$ we see that the absolute moment transform of $X \sim H\exp^+(\alpha,\beta,\gamma)$ is given by

$$(4.26.5)\qquad m_X(q) = \frac{\beta^{q/\gamma}\,\Gamma\left(\frac{q+\alpha}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \qquad \forall q \ge 0$$

The two-sided hyper-exponential distribution $H\exp(\mu,\alpha,\beta,\gamma)$, where $\mu \in \mathbf{R}$ and $\alpha,\beta,\gamma > 0$, is given by the density function

$$(4.26.6)\qquad f(x) = \tfrac{1}{2}\gamma\,\beta^{-\alpha/\gamma}\,\Gamma(\tfrac{\alpha}{\gamma})^{-1}\,|x-\mu|^{\alpha-1}\,e^{-|x-\mu|^\gamma/\beta} \qquad \forall x \in \mathbf{R}$$

If $x \in \mathbf{R}$ and $\lambda > 0$, we define $x^{[\lambda]} = |x|^\lambda\,(\operatorname{sign} x)$ where $\operatorname{sign} x = -1$ if $x < 0$, $\operatorname{sign} x = 1$ if $x > 0$, and $\operatorname{sign} 0 = 0$. Then by symmetry of f around μ and the usual transformation formula [see (4.2.6)], we find

$$(4.26.7)\qquad X \sim H\exp(\mu,\alpha,\beta,\gamma) \;\Rightarrow\; EX = \mu$$

$$(4.26.8)\qquad X \sim H\exp(\mu,\alpha,\beta,\gamma) \;\Rightarrow\; (X+c)^{[\lambda]} \sim H\exp\left(\mu+c,\,\tfrac{\alpha}{\gamma},\,\beta,\,\tfrac{\gamma}{\lambda}\right)$$

$$(4.26.9)\qquad X \sim H\exp(\mu,\alpha,\beta,\gamma) \;\Rightarrow\; |X-\mu| \sim H\exp^+(\alpha,\beta,\gamma)$$

whenever $\mu, c \in \mathbf{R}$ and $\alpha, \beta, \gamma, \lambda > 0$. Thus, the central moment transforms of $X \sim H\exp(\mu, \alpha, \beta, \gamma)$ are given by

$$(4.26.10) \qquad m_X^0(q) = \frac{\beta^{q/\beta}\,\Gamma\left(\frac{q+\alpha}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \qquad \forall q \geq 0$$

$$(4.26.11) \qquad \mu_X^0(2n+1) = 0 \qquad \forall n \in \mathbf{N}_0$$

$$(4.26.12) \qquad \mu_X^0(2n) = \frac{\beta^{2n/\beta}\,\Gamma\left(\frac{2n+\alpha}{\gamma}\right)}{\Gamma\left(\frac{\alpha}{\gamma}\right)} \qquad \forall n \in \mathbf{N}_0$$

Note that the normal distribution, $N(\mu, \sigma^2)$, and *the Laplace distribution*, $Lap(\mu, \beta)$, are special case of the two-sided hyper-exponential distribution:

$$(4.26.13) \qquad N(\mu, \sigma^2) = H\exp(\mu, 1, 2\sigma^2, 2)$$

$$(4.26.14) \qquad Lap(\mu, \beta) = H\exp(\mu, 1, \beta, 1)$$

Let $X \sim H\exp^+(\alpha, \beta, \gamma)$, where $\alpha, \beta > 0$ and $0 < \gamma < \frac{1}{2}$, and let f be the density of X [see (4.26.1)]. Then we may choose $\rho > 0$ such that $k := \alpha + \rho$ is an integer, and we set $r = \frac{\rho}{\gamma}$, $\lambda = \frac{1}{\beta} + r$, $\xi = \tan(\gamma\pi)$ and

$$\psi(y) = y^\rho\, e^{-r\,y^\gamma}\, \sin(\lambda\xi\, y^\gamma) \qquad \forall y > 0$$

Since $e^{-x} \leq r^r\, x^{-r}$ for all $x > 0$, we see that $|\psi(y)| \leq 1$ for all $y > 0$. Let $n \in \mathbf{N}_0$ be a given non-negative integer. Then $m = n + k \in \mathbf{N}$ and $v = \pi\gamma \in]0, \frac{\pi}{2}[$. Hence, substituting $x = y^\gamma$ and applying (4.23.9) we get

$$\gamma \int_0^\infty y^{n+\alpha-1}\, \psi(y)\, e^{-y^\gamma/\beta} dy = \int_0^\infty x^{(m\pi/v)-1}\, e^{-\lambda x}\, \sin(\lambda\xi x)\, dx = 0$$

Thus, we find

$$\int_0^\infty y^n\,(1 + \delta\,\psi(y))\, f(y)\, dy = \int_0^\infty y^n\, f(y)\, dy \qquad \forall n \in \mathbf{N}_0\, \forall \delta \in \mathbf{R}$$

and since $|\psi(y)| \leq 1$, we conclude that for every $-1 \leq \delta \leq 1$, we have that

$$f_\delta(x) := \begin{cases} f(x)\,(1 + \delta\,\psi(x)) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

is a density function having the same moments as $H\,exp(\alpha, \beta, \gamma)$.

Thus, the one-sided hyper-exponential distribution $H\exp^+(\alpha,\beta,\gamma)$ has moments of all orders, but if $0<\gamma<\frac{1}{2}$, then it is **not** uniquely determined by its moments. If $\gamma\geq 1$, then the domain of the Laplace transform of $H\exp^+(\alpha,\beta,\gamma)$ contains a neighborhood of 0. Thus, it is uniquely determined by its moments. If $\frac{1}{2}\leq\gamma<1$, then the situation is much more tricky, but using an idea of Christian Berg (1988) it can be shown that $H\exp^+(\alpha,\beta,\gamma)$ is uniquely determined by its moments.

Summing up, we have shown that the one-sided hyper-exponential distribution $H\exp^+(\alpha,\beta,\gamma)$ is uniquely determined by its moments if $\gamma\geq\frac{1}{2}$, but not if $0<\gamma<\frac{1}{2}$.

Let us consider the two-sided hyper-exponential distribution $H\exp(\mu,\alpha,\beta,\gamma)$. Suppose that $X\sim H\exp(\mu,\alpha,\beta,\gamma)$ where $\mu\in\mathbf{R}$, $\alpha>0$, $\beta>0$ and $0<\gamma<1$. Then by (4.26.9) and (4.26.2) we have that $U=\sqrt{|X-\mu|}\sim H\exp(\frac{\alpha}{2},\beta,\frac{\gamma}{2})$ so by the argument above there exists a random variable V such that $V\geq 0, EV^n=EU^n$ for all $n\in\mathbf{N}_0$, but V and U have different distributions. Let ε be a random variable independent of V such that $P(\varepsilon=1)=P(\varepsilon=-1)=\frac{1}{2}$ and set $W=\varepsilon V^2$. Then W is symmetric and so

$$EW^n=E(X-\mu)^n=0 \qquad \text{if } n \text{ is odd}$$
$$EW^n=EV^{2n}=EU^{2n}=E(X-\mu)^n \qquad \text{if } n \text{ is even}$$

Hence, W and $(X-\mu)$ have the same moments, and by (4.3.23) we see that $W+\mu$ and X have the same moments but different distributions. If $\gamma\geq 1$, then the domain of the Laplace transform of $H\exp(\mu,\alpha,\beta,\gamma)$ contains a neighborhood of 0; thus, it is uniquely determined by its moments.

Summing up we have shown that the two-sided hyper-exponential distribution $H\exp(\mu,\alpha,\beta,\gamma)$ is uniquely determined by its moments if $\gamma\geq 1$, but not if $0<\gamma<1$.

Let X have a normal distribution $N(0,1)$, and then by (4.26.13), (4.29.9), (4.26.2) and (4.26.8) we have

$$|X|^\lambda\sim H\exp^+(\tfrac{1}{\lambda},2,\tfrac{2}{\lambda}) \qquad \text{if } \lambda>0$$
$$X^n\sim H\exp(0,\tfrac{1}{n},2,\tfrac{2}{n}) \qquad \text{if } n\in\mathbf{N} \text{ and } n \text{ is odd}$$

Hence, the distribution of $|X|^\lambda$ is uniquely determined by its moments if $\lambda\leq 4$, but not if $\lambda>4$. Similarly, the distribution of X^n is uniquely determined by its moments if $n=1,2$ or 4, but not for any other $n\in\mathbf{N}$.

4.27 A Discontinuous Laplace Transform

By (4.19.12) we know that a Laplace transform is continuous in the interior of its domain, and that a one–dimensional Laplace transform is continuous on all of

their domain. However, there exist fairly simple examples of two–dimensional Laplace transforms which are discontinuous at the boundary. Let U and V be two independent $N(0,1)$-distributed random variables and set $W = V\sqrt{U^2+1-2\log(1+U^2)}$. If $X = (U,W)$, then by (4.5.2) and (4.22.2) we get

$$L_X(s,t) = Ee^{sU+tW} = \sqrt{\tfrac{e}{2\pi}}\int_{-\infty}^{\infty}\frac{\exp\left(sx-\frac{1}{2}(1-t^2)x^2\right)}{1+x^2}\,dx \qquad \forall(s,t)\in\mathbf{R}^2$$

An elementary but lengthy computation shows that

$$\mathcal{R}(L_X) = (\mathbf{R}\times\,]-1,1[\,)\cup\{\,(0,1),\,(0,-1)\,\}\quad,\quad L_X(0,1)=\sqrt{\tfrac{e\pi}{2}}$$

$$\lim_{s\downarrow 0} L_X\left(s,\sqrt{1-2s^3}\right)=\infty$$

Thus, L_X is not continuous at $(0,1)$.

4.28 The Mellin Transform and Moments

Let U be a non–negative random variable and let $g:\mathbf{R}_+\to\mathbf{R}_+$ be a non-negative Borel function. Set $G(t)=Eg(tU)$ for $t\geq 0$, and let M_g and M_G denote the Mellin of g and G [see (3.36)]. Then by Tonelli's theorem (3.21.1) and (3.36.3), we have

$$\begin{aligned}M_G(p) &= \int_0^\infty t^{p-1}\,G(t)\,dt = \int_0^\infty dt\int_\Omega t^{p-1}\,g(tU(\omega))P(d\omega)\\ &= \int_\Omega P(d\omega)\int_0^\infty t^{p-1}\,g(tU(\omega))\,dt = M_g(p)\,EU^{-p}\end{aligned}$$

if either $P(U=0)=0$, or $g(0)=0$ and $p<0$, or $g(0)>0$ and $p>0$. Thus, we obtain the following formula for EU^{-p} :

(4.28.1) If $G(t)=Eg(tU)$ and $p\in\mathcal{E}(g)$, where $g:[0,\infty[\to[0,\infty]$ is Borel measurable and U is a non-negative random variable, then we have:

$$EU^{-p} = \frac{M_G(p)}{M_g(p)}$$

in any of the following three cases: (a): $p<0$ and $g(0)=0$. (b): $p>0$ and $g(0)>0$. (c): $P(U=0)=0$.

Using (4.28.1) and Table 3.1 we obtain the following formulas for the moments of a non–negative random variable U :

$$EU^{-p} = \Gamma(p)^{-1} \int_0^\infty t^{p-1} L_U(-t)\, dt \qquad \forall p > 0 \tag{4.28.2}$$

$$EU^r = \frac{r}{\Gamma(1-r)} \int_0^\infty t^{-r-1} (1 - L_U(-t))\, dt \qquad \forall 0 < r < 1 \tag{4.28.3}$$

$$EU^r = C(r)^{-1} \int_0^\infty t^{-r-1} (1 - \Re\varphi_U(t))\, dt \qquad \forall 0 < r < 2 \tag{4.28.4}$$

$$EU^r = (r-\beta) \int_0^\infty t^{\beta-r-1} R_\beta\left(\tfrac{1}{t}\right) dt \qquad \forall r > \beta > 0 \tag{4.28.5}$$

$$EU^{-p} = \frac{\Gamma(\alpha+p)}{\Gamma(p)\Gamma(\alpha)} \int_0^\infty t^{-p-\alpha} G_\alpha(t)\, dt \qquad \forall \alpha, p > 0 \tag{4.28.6}$$

$$EU^r = (r-\beta) \int_0^\infty t^{r-\beta-1} R_\beta(t)\, dt \qquad \forall r > \beta > 0 \tag{4.28.7}$$

$$EU^{-p} = \Gamma(p)^{-1} \int_0^1 \tfrac{1}{t} \left(\log \tfrac{1}{t}\right)^{p-1} q_U(t)\, dt \qquad \forall p > 0 \tag{4.28.8}$$

$$EU^r = \frac{r}{\Gamma(1-r)} \int_0^1 \tfrac{1}{t} \left(\log \tfrac{1}{t}\right)^{-r-1} (1 - q_U(t))\, dt \qquad \forall 0 < r < 1 \tag{4.28.9}$$

where L_U is the Laplace transform, φ_U is the characteristic function, q_U is the generating function of U , $C(r)$ is defined by (3.37.14), and R_β and G_α are given by

$$R_\beta(t) := \int_{\{U>t\}} U(\omega)^\beta P(d\omega) \quad , \quad G_\alpha(t) := \int_{\{U<t\}} (t - U(\omega))^{\alpha-1} P(d\omega)$$

4.29 Subgaussian Random Variables

If U has a normal distribution with mean μ and variance v , then by (4.22.2) we know that its cumulant transform is given by $\kappa_U(s) = \mu s + \frac{1}{2}s^2 v$ for all $s \in \mathbf{R}$. Let X be a random variable, then we say that X is *subgaussian* if the cumulant transform of X is dominated on $\mathbf{R}$ by the cumulant transform of some normally distributed random variable. Hence, X is subgaussian if and only if there exists $\mu \in \mathbf{R}$ and $v \geq 0$ satisfying

$$\kappa_X(s) = \log E(e^{sX}) \leq \mu s + \tfrac{1}{2}s^2 v \qquad \forall s \in \mathbf{R} \tag{4.29.1}$$

I claim:

(4.29.2) If X satisfies (4.29.1), then X has moments of all orders and $\mu = EX$ and $\operatorname{var}(X) \le v$

To see this, note that (4.29.1) implies that $\mathcal{D}(L_X) = \mathbf{C}$; hence, by (4.19.19) we see that X has moments of all orders. Moreover, by (4.29.1) we have

$$\frac{\kappa_X(s) - \mu s}{s^2} \le \tfrac{1}{2} v \quad \text{and} \quad \frac{\kappa_X(s)}{s} \begin{cases} \le \mu + \frac{1}{2} s v & \text{if } s > 0 \\ \ge \mu + \frac{1}{2} s v & \text{if } s < 0 \end{cases}$$

Letting $s \downarrow 0$ and $s \uparrow 0$ in the last inequalities, we see that $\kappa_X'(0) = \mu$. Hence, by (4.19.19) we have that $\mu = EX$. Letting $s \to 0$ in the first inequality and applying (4.19.19) and L'Hospital's rule, we see that $\operatorname{var}(X) = \kappa_X''(0) \le v$.

Let X be an integrable random variable with mean $\mu = EX$. Then the smallest number $v \ge 0$ satisfying (4.29.1) is called *the Gaussian subvariance* of X and is denoted $\gamma(X)$; that is,

(4.29.3) $$\gamma(X) := \sup_{s \ne 0} \left\{ \frac{2(\kappa_X(s) - \mu s)}{s^2} \right\}$$

Thus, an integrable random variable X is subgaussian if and only if $\gamma(X) < \infty$.

4.30 The Subvariance Inequalities

Let $X, X_1, \ldots, X_n$ be independent integrable random variables. Then we have

(4.30.1) $$\gamma\Big(\sum_{j=1}^n X_j\Big) \le \sum_{j=1}^n \gamma(X_j)$$

(4.30.2) $$\gamma(X + a) = \gamma(X) \qquad \forall a \in \mathbf{R}$$

(4.30.3) $$\gamma(aX) = a^2 \gamma(X) \qquad \forall a \in \mathbf{R}$$

(4.30.4) $$E|X - EX|^\alpha \le 2^{\alpha/2} \alpha \Gamma\big(\tfrac{\alpha}{2}\big) \gamma(X)^{\alpha/2}$$

(4.30.5) $$\operatorname{var}(X) \le \gamma(X) \le \|X - EX\|_\infty^2$$

where $\|Y\|_\infty$ is the L^∞-norm, that is, the smallest number a such that $|Y| \le a$ a.s. Moreover, if $b = (b_1, \ldots, b_n) \in \mathbf{R}^n$ is a given vector, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent Bernoulli random variables [see the remark below], then we have

(4.30.6) $$\gamma\Big(\sum_{j=1}^n b_j \varepsilon_j\Big) = \sum_{j=1}^n b_j^2 = \|b\|^2$$

$$(4.30.7) \qquad E\left|\sum_{j=1}^{n} b_j\,\varepsilon_j\right|^{\alpha} \le C_\alpha\left\{\sum_{j=1}^{n} b_j^2\right\}^{\alpha/2} = C_\alpha\,\|b\|^{\alpha}$$

where C_α *is given by*

$$(4.30.8) \qquad C_\alpha = \begin{cases} 1 & \text{if } \alpha \le 2 \\ 2^{\alpha/2}\,\alpha\,\Gamma\left(\frac{\alpha}{2}\right) & \text{if } \alpha > 2 \end{cases}$$

Remarks: (1): A random variable ε is called a *Bernoulli variable* if ε takes the values ± 1 with probability $\frac{1}{2}$, that is, $P(\varepsilon = 1) = P(\varepsilon = -1) = \frac{1}{2}$.

(2): Inequality (4.30.7) is called *Khinchine's inequality*. The constant C_α appearing in (4.30.7) is very close to the best possible constant. U. Haagerup (1979) has shown that the best possible constant in (4.30.7) is 1 for $\alpha \le 2$ (this is easy!) and equal to $\pi^{-1/2}\,2^{\alpha/2}\,\Gamma(\frac{\alpha+1}{2})$ for $\alpha > 2$ (this is very difficult!).

Proof: (4.30.1)–(4.30.3) follows easily from (4.19.6) and (4.19.10).

(4.30.4): Let $Y = X - EX$, $v = \gamma(X)$ and $s = \frac{t}{v}$. Then by Markov's inequality (3.9) with $\varphi(x) = e^{s\,x}$ we find

$$\begin{aligned} P(|Y| > t) &\le e^{-st}\,Ee^{s|Y|} \le e^{-st}\left(Ee^{sY} + Ee^{-sY}\right) \\ &= e^{-st}\left(e^{-s\mu}\,L_X(s) + e^{s\mu}\,L_X(-s)\right) \\ &\le 2\,\exp\left(-st + \tfrac{1}{2}\,s^2\,v\right) = 2\,e^{-t^2/(2v)} \end{aligned}$$

for all $t \ge 0$. Hence, by (4.3.24) and (4.22.4) we get

$$\begin{aligned} E|Y|^{\alpha} &= \alpha\int_0^{\infty} t^{\alpha-1}\,P(|Y| > t)\,dt \le 2\alpha\int_0^{\infty} t^{\alpha-1}\,e^{-t^2/(2v)}\,dt \\ &= \frac{\alpha\sqrt{2\pi v}}{\sqrt{2\pi v}}\int_{-\infty}^{\infty} |t|^{\alpha-1}\,e^{-t^2/(2v)}\,dt = \alpha\,\Gamma\left(\tfrac{\alpha}{2}\right)(2v)^{\alpha/2} \end{aligned}$$

Thus, (4.30.4) is proved.

(4.30.5): The first inequality follows from (4.29.2). Set $Y = X - EX$. If $\|Y\|_\infty = \infty$, there is nothing to prove. So suppose that $a := \|Y\|_\infty < \infty$ and let $s \in \mathbf{R}$ and $u \in [-a, a]$ be given. If $\lambda = \frac{a+u}{2a}$, then $1 - \lambda = \frac{a-u}{2a}$, $0 \le \lambda \le 1$ and $su = \lambda(sa) + (1-\lambda)(-sa)$. Hence, by convexity of exp we see that

$$e^{su} \le \tfrac{a+u}{2a}\,e^{sa} + \tfrac{a-u}{2a}\,e^{-sa} = \cosh(sa) + \tfrac{u}{a}\,\sinh(sa)$$

and since $|Y| \le a$ a.s. and $EY = 0$, we conclude that $L_Y(s) \le \cosh(sa)$. Moreover, since $\cosh(x) \le \exp(\frac{1}{2}x^2)$, we have $L_Y(s) \le \exp(\frac{1}{2}s^2a^2)$ and so $\gamma(X) \le a^2$.

(4.30.6): Since $E\varepsilon_j = 0$ and $|\varepsilon_j| = 1$, then by (4.30.3) and (4.30.5) we see that $\gamma(b_j\,\varepsilon_j) \le b_j^2$. Hence, if $Y = b_1\,\varepsilon_1 + \cdots + b_n\,\varepsilon_n$, then by (4.30.1) we have that $\gamma(Y) \le ||b||^2$, and since $\mathrm{var}(Y) = ||b||^2$, then by (4.30.5) we conclude, that $\gamma(Y) = ||b||^2$.

(4.30.7): If $\alpha > 2$, then (4.30.7) simply follows from (4.30.4+6). If $\alpha \le 2$ then (4.30.7) follows from (4.3.9) and (4.4.18) because $b_1\varepsilon_1, \ldots b_n\varepsilon_n$ are independent with means 0 and variances $b_1^2, \ldots, b_n^2$. □

4.31 Sign-Symmetric Random Vectors

Khinchine's inequality (4.30.7) provides us with a very powerful tool to prove moment inequalities of the type (4.10.1) which are needed in the law of large numbers. The method to extend (4.30.7) to arbitrary independent random variables depends on a certain "symmetrization" of random vectors. Recall that a random variable X is *symmetric* if $X \sim (-X)$. A random vector $X = (X_1, \ldots, X_k)$ is called *sign-symmetric* if for every choice of signs, $\beta_1 = \pm 1, \ldots, \beta_k = \pm 1$ we have that $(X_1, \ldots, X_k)$ has the same distribution as $(\beta_1 X_1, \ldots, \beta_k X_k)$. I claim that we have:

(4.31.1) Let $X = (X_1, \ldots, X_k)$ be a sign-symmetric random vector, and let $\eta = (\eta_1, \ldots, \eta_k)$ be a random vector such that η is independent of X and $\eta_1, \ldots, \eta_k$ only assumes the values ± 1 . Then $(X_1, \ldots, X_k) \sim (\eta_1 X_1, \ldots, \eta_k X_k)$

So let $B \in \mathcal{B}(\mathbf{R}^k)$, and let $Y = (\eta_1 X_1, \ldots, \eta_k X_k)$. Then

$$P(Y \in B) = \sum_{\beta \in \Delta} P(\eta_1 = \beta_1, \ldots, \eta_k = \beta_k \ , \ (\beta_1 X_1, \ldots, \beta_k X_k) \in B)$$

where $\Delta = \{(\beta_1, \ldots, \beta_k) \mid \beta_1 = \pm 1, \ldots, \beta_k = \pm 1\}$. Since $\eta \,\underline{||}\, X$ and X is sign-symmetric, we find

$$P(Y \in B) = \sum_{\beta \in \Delta} P(\eta_1 = \beta_1, \ldots, \eta_k = \beta_k)\, P(X \in B) = P(X \in B)$$

and so $X \sim Y$. In the same manner we get

(4.31.2) Let $Y = (Y_1, \ldots Y_k)$ and $X = (X_1, \ldots, X_k)$ be independent random vectors. If Y is sign-symmetric, then $(Y_1 X_1, \ldots, Y_k X_k)$ is sign-symmetric. If $X \sim Y$ and $X_1, \ldots, X_k$ are independent, then $X - Y = (X_1 - Y_1, \ldots, X_k - Y_k)$ is sign-symmetric

With these results at hand, we can prove the general *Khinchine's inequality.*

4.32 Khinchine's Inequality

Let $X_1, \ldots, X_n$ be independent random variables with finite means $\mu_1, \ldots, \mu_n$. Then for every $\alpha > 0$, we have

$$(4.32.1) \qquad E\left| \sum_{j=1}^{n} (X_j - \mu_j) \right|^{\alpha} \le 2^{\alpha} C_{\alpha} n^{\lambda} \sum_{j=1}^{n} E|X_j - \mu_j|^{\alpha}$$

where $\lambda = (\frac{\alpha}{2} - 1)^+$ and C_α is given by

$$(4.32.2) \qquad C_{\alpha} = \begin{cases} 1 & \text{if } \alpha \le 2 \\ 2^{\alpha/2}\, \alpha\, \Gamma\left(\frac{\alpha}{2}\right) & \text{if } \alpha > 2 \end{cases}$$

Moreover, if $V = (V_1, \ldots, V_n)$ is a sign-symmetric random vector, then we have

$$(4.32.3) \qquad E\left| \sum_{j=1}^{n} V_j \right|^{\alpha} \le C_{\alpha} E\|V\|^{\alpha} \le C_{\alpha} n^{\lambda} \sum_{j=1}^{n} E|V_j|^{\alpha}$$

Proof: (4.32.3): Let $\varepsilon_1, \ldots, \varepsilon_k$ be independent Bernoulli variables such that $(\varepsilon_1, \ldots, \varepsilon_k)$ is independent of V. Then by (4.30.7) we have

$$h(u) := E\left| \sum_{j=1}^{n} \varepsilon_j u_j \right|^{\alpha} \le C_{\alpha} \|u\|^{\alpha} \qquad \forall\, u = (u_1, \ldots, u_n) \in \mathbf{R}^n$$

By (4.31.1) and sign-symmetry of V, we have $(V_1, \ldots, V_k) \sim (\varepsilon_1 V_1 \ldots, \varepsilon_k V_k)$. Moreover, by (4.8.9) with $q = \frac{\alpha}{2}$ we have $\|u\|^{\alpha} \le n^{\lambda}(|u_1|^{\alpha} + \cdots |u_n|^{\alpha})$. Hence, by (4.5.2) we get

$$E\left| \sum_{j=1}^{n} V_j \right|^{\alpha} = E\left| \sum_{j=1}^{n} \varepsilon_j V_j \right|^{\alpha} = Eh(V)$$

$$\le C_{\alpha} E\|V\|^{\alpha} \le C_{\alpha} n^{\lambda} \sum_{j=1}^{n} E|V_j|^{\alpha}$$

(4.32.1): If $0 < \alpha \le 1$, then (4.32.1) follows from Minkowski's inequality (3.12.3). So suppose that $1 < \alpha < \infty$. Let $Y_j = X_j - \mu_j$ and $S = Y_1 + \cdots + Y_n$, and let $Z = (Z_1, \ldots, Z_n)$ be a random vector such that Z and $Y = (Y_1, \ldots, Y_n)$ are independent and $Z \sim Y$. Set $T = Z_1 + \cdots + Z_n$, then $ET = 0$ and S and T are independent. Moreover, by (4.31.2) we have that $Y - Z = (Y_1 - Z_1, \ldots, Y_n - Z_n)$ is sign-symmetric. Since $\alpha > 1$ we have

that $(u+v)^\alpha \le 2^{\alpha-1}(u^\alpha + v^\alpha)$ for all $u, v \ge 0$. Hence, setting $q = \frac{\alpha}{2}$, then by (4.6.5), (4.32.3) and (4.8.9) we find

$$\begin{aligned} E|S|^\alpha &\le E|S-T|^\alpha \le C_\alpha\, E\|Y-Z\|^\alpha \\ &\le 2^{\alpha-1} C_\alpha\, (E\|Y\|^\alpha + E\|Z\|^\alpha) = 2^\alpha C_\alpha\, E\|Y\|^\alpha \\ &= 2^\alpha C_\alpha\, E\Big|\sum_{j=1}^n Y_j^2\Big|^q \le 2^\alpha C_\alpha\, n^\lambda\, E\Big(\sum_{j=1}^n |Y_j|^{2q}\Big) \\ &= 2^\alpha C_\alpha\, n^\lambda \sum_{j=1}^n E|Y_j|^\alpha \end{aligned}$$

Thus, (4.32.1) is proved. □

4.33 The Law of Large Numbers IV

Let $X_1, X_2, \cdots$ *be a sequence of independent random variables with finite means* $\mu_n = E\,X_n$ *and suppose that there exists* $\alpha > 0$ *such that*

$$\sum_{n=1}^\infty n^{-\beta}\, E|X_n - \mu_n|^\alpha < \infty \tag{4.33.1}$$

where $\beta = \alpha$ *if* $0 < \alpha \le 2$ *and* $\beta = \frac{1}{2}(\alpha+2)$ *if* $\alpha > 2$. *Then we have*

$$\frac{1}{n}\sum_{j=1}^n (X_j - \mu_j) \to 0 \quad \textit{P-a.s and in } \alpha\textit{-mean} \tag{4.33.2}$$

Proof: By (4.32.1) we see that (4.10.1) holds with $\tau_j = E|X_j - \mu_j|^\alpha$, $q = \alpha$, $c = 2^\alpha C_\alpha$ and $r = \lambda$, where C_α and λ are defined as in (4.32). Since $\beta = \alpha - \lambda$, we see that (4.10.2) follows from (4.33.1). Thus, (4.33.2) follows from (4.10.3) and (4.10.5). □

4.34 A Maximal Moment Inequality

Let $X_1, X_2, \ldots, X_n$ *be independent random variables with finite means* $\mu_1, \ldots, \mu_n$. *Then for all* $\alpha > 0$, *we have*

$$E\Big\{\max_{1\le k\le n}\Big|\sum_{j=1}^k (X_j - \mu_j)\Big|^\alpha\Big\} \le 6\cdot 4^\alpha C_\alpha\, n^\lambda \sum_{j=1}^n E|X_j - \mu_j|^\alpha \tag{4.34.1}$$

where λ *and* C_α *are defined as in (4.32).*

Proof: Let $S_k = (X_1 - \mu_1) + \cdots + (X_k - \mu_k)$ for $k = 1, \ldots, n$ and $M_n = \max\{|S_1|, \ldots, |S_n|\}$. By (2.18.7), we have

$$P(M_n > x) \min_{1 \le j \le n} P(|S_n - S_j| \le y) \le P(|S_n| > x - y)$$

for all $x, y \in \mathbf{R}$. Let $y > 0$ be given. By Markov's inequality (3.9) and Khinchine's inequality (4.32.1), we have

$$\begin{aligned} P(|S_n - S_j| \le y) &= 1 - P(|S_n - S_j|) > y) \\ &\ge 1 - y^{-\alpha} E|S_n - S_j|^\alpha \ge 1 - y^{-\alpha} 2^\alpha C_\alpha (n-j)^\lambda \sum_{i=j+1}^{n} E|X_i - \mu_i|^\alpha \\ &\ge 1 - y^{-\alpha} 2^\alpha C_\alpha n^\lambda \sum_{i=1}^{n} E|X_i - \mu_i|^\alpha \end{aligned}$$

Hence, if we set

$$\tau = \left\{ 2^{1+\alpha} C_\alpha n^\lambda \sum_{i=1}^{n} E|X_i - \mu_i|^\alpha \right\}^{1/\alpha}$$

then $P(|S_n - S_j| \le \tau) \ge \frac{1}{2}$ for all $j = 1, \ldots, n$ and so we have

$$P(M_n > x) \le 2P(|S_n| > x - \tau) \qquad \forall x \in \mathbf{R}$$

Moreover, since $|x + y|^\alpha \le 2^\alpha(|x|^\alpha + |y|^\alpha)$, then by (4.3.24) we have

$$\begin{aligned} EM_n^\alpha &= \alpha \int_0^\infty t^{\alpha-1} P(M_n > t)\, dt \le 2\alpha \int_0^\infty t^{-\alpha} P(|S_n| + \tau > t)\, dt \\ &= 2E(\tau + |S_n|)^\alpha \le 2^{\alpha+1}(\tau^\alpha + E|S_n|^\alpha) \end{aligned}$$

By Khinchine's inequality (4.32.1) we have $E|S_n|^\alpha \le \frac{1}{2}\tau^\alpha$; so we find

$$EM_n^\alpha \le 3 \cdot 2^\alpha \tau^\alpha = 6 \cdot 4^\alpha C_\alpha n^\lambda \sum_{j=1}^{n} E|X_j - \mu_j|^\alpha$$

Thus, (4.34.1) is proved. □

4.35 A Tail Inequality

Let X be a non-negative random variable, and let $\Gamma : S \times [0,\infty[\to [0,\infty[$ be a measurable kernel where $(S,\mathcal{B})$ is given measurable space. Suppose that Y is an S-valued random function satisfying

$$P(X \geq t) \leq E(\Gamma(Y,t)) \qquad \forall t > 0 \tag{4.35.1}$$

Let $\varphi : [0,\infty[\to [0,\infty[$ be an increasing right continuous function. Then we have

$$E\varphi(X) \leq \varphi(a) + E(\hat{\Gamma}_a(Y)) \qquad \forall a \geq 0 \tag{4.35.2}$$

where $\hat{\Gamma}_a(s)$ is given by

$$\hat{\Gamma}_a(s) = \int_a^\infty \Gamma(s,t)\,\varphi(dt) \qquad \forall s \in S\ \forall a \geq 0 \tag{4.35.3}$$

Proof: By (3.30.4), (4.35.1) and Tonelli's theorem (3.21.1) we find

$$\begin{aligned}
E\varphi(X) &= \varphi(0) + \int_0^\infty P(X \geq t)\,\varphi(dt) \\
&\leq \varphi(0) + \int_0^a 1\,\varphi(dt) + \int_a^\infty P(X \geq t)\,\varphi(dt) \\
&\leq \varphi(a) + \int_a^\infty \varphi(dt) \int_\Omega \Gamma(Y(\omega),t)\,P(d\omega) \\
&= \varphi(a) + \int_\Omega P(d\omega) \int_a^\infty \Gamma(Y(\omega),t)\,\varphi(dt) \\
&= \varphi(a) + E(\hat{\Gamma}_a(Y))
\end{aligned}$$

Thus, (4.35.3) is proved. □

4.36 Convergence of Partial Sums Revisited

(1): Let $X_1, X_2, \ldots$ be a sequence of independent random variables with finite means $\mu_1, \mu_2, \ldots$ and let $0 < q \leq 2$ be a given number satisfuing

$$\sum_{j=1}^\infty E|X_j - \mu_j|^q < \infty \tag{4.36.1}$$

Then $\sum_1^\infty (X_j - \mu_j)$ *is summable a.s. and in q-mean.*

(2): Suppose that $Y_1, Y_2, \ldots$ *are identically distributed and that* $E|Y_1|^q < \infty$ *for some* $q \in]0, \infty[$. *Then for every* $r > q$ *we have*

$$\sum_{n=1}^{\infty} n^{-r/q} |Y_n|^r < \infty \quad \text{a.s.} \tag{4.36.2}$$

$$n^{-r/q} \sum_{j=1}^{n} |Y_j|^r \to 0 \quad \text{a.s. and in } \tfrac{q}{r}\text{-mean} \tag{4.36.3}$$

(3): Suppose that $Y_1, Y_2, \ldots$ *are independent and identically distributed such that* $EY_1 = 0$ *and* $Y_1 \in L^q(P) < \infty$ *for some* $q \geq 1$. *Then we have*

$$1 < q < 2 \quad \Rightarrow \quad \sum_{j=1}^{\infty} j^{-1/q} Y_j \quad \text{is summable a.s.} \tag{4.36.4}$$

$$1 \leq q < 2 \quad \Rightarrow \quad n^{-1/q} \sum_{j=1}^{n} Y_j \to 0 \quad \text{a.s. and in } q\text{-mean} \tag{4.36.5}$$

Proof: (1): Let $Y_j = X_j - \mu_j$ and $S_n = Y_1 + \cdots + Y_n$. Since $q \leq 2$, then by Khinchine's inequality (4.32.1) we get

$$\begin{aligned} E|S_{n+k} - S_k|^q &= E\Big| \sum_{j=k+1}^{k+n} (X_j - \mu_j)\Big|^q \\ &\leq 2^q C_q \sum_{j=k+1}^{\infty} E|X_j - \mu_j|^q \end{aligned}$$

Hence, by (4.36.1) we see that $\{S_n\}$ is a Cauchy sequence in $L^q(P)$. But then $\{S_n\}$ converges in q-mean (and thus, in probability) to some random variable $S \in L^q(P)$ [see (3.25.5) and (3.25.10)]. Hence, (1) follows from (4.7).

(2): Let $V = |Y_1|$, $A_n = \{|Y_n| \leq n^{1/q}\}$ and $U_n = Y_n \, 1_{A_n}$. Then $|U_n| \sim V \, 1_{\{V^q \leq n\}}$. Hence, by (4.8.16) and Beppo Levi's theorem (3.7) we get

$$\begin{aligned} \sum_{n=1}^{\infty} n^{-r/q} E|U_n|^r &= \sum_{n=1}^{\infty} n^{-r/q} \int_\Omega V^r \, 1_{\{V^q \leq n\}} \, dP \\ &= \int_\Omega V^r \Big\{ \sum_{n \geq V^q} n^{-r/q} \Big\} dP \\ &\leq \frac{r}{r-q} \int_\Omega V^r \, (V^q)^{1-\frac{r}{q}} \, dP = \frac{r \, EV^q}{r-q} < \infty \end{aligned}$$

Thus, by Beppo Levi's theorem (3.7) we have $\sum_1^\infty n^{-r/q}|U_n|^r < \infty$ a.s. Moreover, by (4.2.7) we find

$$\sum_{n=1}^{\infty} P(U_n \neq Y_n) = \sum_{n=1}^{\infty} P(V^q > n) \leq \int_0^\infty P(V^q > t)\,dt = EV^q < \infty$$

Hence, by the first Borel-Cantelli lemma (2.11.1) we have

$$\sum_{n=1}^{\infty} n^{-1/q}|U_n - Y_n|^r < \infty \text{ a.s.}$$

Using the inequality: $|x+y|^r \leq 2^r(|x|^r + |y|^r)$, we see that (4.36.2) holds. Let $W_n = n^{-1}\left(|Y_1|^r + \cdots + |Y_n|^r\right)^{q/r}$. Then

$$n^{-r/q}\sum_{j=1}^{n}|Y_j|^r = (W_n)^{r/q}$$

By Kronecker's lemma (4.9.2), we see that the a.s. convergence in (4.36.3) holds, and so $W_n \to 0$ a.s. Note that $W_n \leq (|Y_1|^q + \cdots + |Y_n|^q)/n$. Hence, by Exc.4.35 and Exc.4.39 we have that $\{W_n\}$ is uniformly integrable, and so the convergence in $\frac{q}{r}$-mean follows from Dunford-Pettis theorem (3.23).

(3): Suppose that $1 < q < 2$, and let V and U_n be defined as above. Since $EY_1 = 0$ and $\sum_{1\leq j<x} j^{-\alpha} \leq x^{1-\alpha}/(1-\alpha)$ whenever $x \geq 0$ and $0 < \alpha < 1$, then, as above, we find

(i) $$\sum_{j=1}^{\infty} j^{-2/q}\,E|U_j|^2 \leq \frac{2\,EV^q}{2-q} < \infty$$

(ii) $$|EU_j| = \left|\int_{\{V^q\leq j\}} Y_1\,dP\right| = \left|\int_{\{V^q>j\}} Y_1\,dP\right| \leq \int_{\{V^q>j\}} V\,dP$$

(iii) $$\sum_{j=1}^{\infty} j^{-1/q}|EU_j| \leq \sum_{j=1}^{\infty} j^{-1/q}\int_{\{V^q>j\}} V\,dP \leq \frac{q\,EV^q}{q-1} < \infty$$

Hence, by (i), (iii) and (1) we see that $\sum j^{-1/q}\,U_j$ summable a.s., but then (4.36.4) follows exactly as above. Moreover, note that the a.s. convergence in (4.36.5) follows from (4.36.4) and Kronecker's lemma (4.9.1) whenever $1 < q < 2$. If $q = 1$ and $T_n = (Y_1 + \cdots + Y_n)/n$, then by the law of large numbers (4.12) we have that $T_n \to 0$ a.s. Moreover, by Exc.4.35 and Exc.4.39 we have that $\{T_n\}$ is uniformly integrable. Thus, $T_n \to 0$ in 1-mean by Dunford Pettis' theorem (3.23). So suppose that $1 < q < 2$, and let $Z_j = Y_j - U_j$ and $\mu_j = EU_j$. Note that $\mu_j \to 0$ by (ii) and $|Z_j| \sim |Z_1| = V\,1_{\{V^q>j\}}$. Hence, we see that

$E|Z_j + \mu_j|^q \to 0$ and since $EY_j = 0$, then $EZ_j = -\mu_j$. So by Khinchine's inequality (4.32.1) and Cesaro lemma (4.9.1), we get

$$\limsup_{n\to\infty} E \left| n^{-1/q} \sum_{j=1}^{n} (Z_j + \mu_j) \right|^q \le \lim_{n\to\infty} \frac{c}{n} \sum_{j=1}^{n} E|Z_j + \mu_j|^q = 0$$

where $c = 2^q C_q$. Moreover, by (4.4.18) we have

$$E \left| n^{-1/q} \sum_{j=1}^{n} (U_j - \mu_j) \right|^2 = n^{-2/q} \sum_{j=1}^{n} E(U_j - \mu_j)^2$$

$$\le n^{-2/q} \sum_{j=1}^{n} j^{2/q} \left(j^{-2/q} E U_j^2 \right)$$

Recall that $q < 2$, hence by (4.3.14), (i) and Kronecker's lemma (4.9.2) we find

$$n^{-1/q} \sum_{j=1}^{n} Y_j = n^{-1/q} \sum_{j=1}^{n} (Z_j + \mu_j) + n^{-1/q} \sum_{j=1}^{n} (U_j - \mu_j) \to 0 \quad \text{in } q\text{-mean}$$

Thus, (4.36.5) is proved. □

4.37 Alternative Expectations

I have defined the expectation EX in (4.1.1) as the P-integral of X . But what is the expected value of X really? Intuitively speaking, EX should be that real number which in some sense is "nearest" possible to the many values that X may assume. We shall now try to understand what "nearest" really means, or rather what it possibly could mean. Suppose that $X \in L^2(P)$. Then the quadratic polynomial:

$$G(s) = E(X - s)^2 = s^2 - 2sEX + EX^2$$

attains its minimums at $s = EX$ and $G(EX) = \text{var}X$. Hence, the expectation EX is that real number "nearest" to X , when we use $||\cdot||_2$ as our measure of "nearness." Moreover, the standard deviation is exactly the distance from EX to X in $(L^2(P), ||\cdot||_2)$. Thus, $\sigma(X)$ is a measure of the error we make when we use the expectation EX in place of X .

But there are many other ways of measuring the distance between random variables. For instance, we could use $||\cdot||_q$ for $0 \le q < \infty$ [see (4.3)], which means that we could define an alternative expectation of X as a minimum point of the function $G(s) = E|X - s|^q$, or even more liberally as a minimum point of the function $G(s) = E\,g(s, X)$ where g is some function. You may

think of $g(s,t)$ as the cost you have to pay if you believe that X equals s when X in reality equals t. I shall now show how to the minima of G when g is nice enough.

Let X be a random variable taking values in the open interval $I =]a,b[$, and let g be a function from $I \times I$ into $\mathbf{R}$ satisfying

(4.37.1) $\quad g(s,s) = 0 \quad \forall s \in I \quad$ and $\quad g(s,t) \geq 0 \quad \forall (s,t) \in I \times I$

(4.37.2) $\quad g(a,t) := \lim_{s \downarrow a} g(s,t) = \infty \quad , \quad g(b,t) := \lim_{s \uparrow b} g(s,t) = \infty$

(4.37.3) $\quad g(\cdot,t)$ is convex on I for all $t \in I$

(4.37.4) $\quad G(s) = Eg(s,X) < \infty \qquad \forall s \in I$

Note that (4.37.1) says that a correct guess is free of charge, and (4.37.2) says that wild guesses are expensive. Let

(4.37.5) $\quad v = \inf_{s \in I} G(s) \quad , \quad M = \{s \in I \mid G(s) = v\}$

By (4.37.2) and Fatou's lemma (3.5), we have that $G(a) = G(b) = \infty$ and by (4.37.3), we have that G is convex. Hence, we conclude that

(4.37.6) $\quad M$ is a closed, bounded non-empty interval $\subseteq I$

Consider the right and left difference quotients:

$$\Delta_+(h) = h^{-1}(g(s+h,t) - g(s,t)) \qquad h > 0$$
$$\Delta_-(h) = h^{-1}(g(s,t) - g(s-h,t)) \qquad h > 0$$

Since $g(\cdot,t)$ is convex we have that $\Delta_+(h)$ is increasing in h and $\Delta_-(h)$ is decreasing in h [see Exc.3.33]. Hence, by the monotone convergence theorem (3.4) we have

(4.37.7) $\quad D_+G(s) = E(D_+g(s,X)) \qquad D_-G(s) = E(D_-g(s,X))$

where D_+f and D_-f denotes the right and left derivatives of f; that is,

$$D_+f(s) = \lim_{h \downarrow 0} \frac{f(s+h) - f(s)}{h} \quad , \quad D_-f(s) = \lim_{h \downarrow 0} \frac{f(s) - f(s-h)}{h}$$

and since G is convex, we have

(4.37.8) $\quad M = \{s \in I \mid D_+G(s) \geq 0 \geq D_-G(s)\}$

In particular, we see that if $g(\cdot, t)$ is differentiable on I, then so is G and

(4.37.9) $$G'(s) = E\left\{\frac{\partial g}{\partial s}(s, X)\right\}$$

(4.37.10) $$M = \{s \in I \mid G'(s) = 0\}$$

Let us consider some special examples. Suppose that $I = \mathbf{R}$ and $g(s,t) = |s - t|$. Then we have

$$D_+g(s,t) = \begin{cases} 1 & \text{if } s \geq t \\ -1 & \text{if } s < t \end{cases}, \quad D_-g(s,t) = \begin{cases} 1 & \text{if } s > t \\ -1 & \text{if } s \leq t \end{cases}$$

and so we find that

$$D_+G(s) = P(X \leq s) - P(X > s) = 2P(X \leq s) - 1$$
$$D_-G(s) = P(X(s) - P(X \geq s) = 2P(X < s) - 1$$

Hence we get:

(4.37.11) $G(s) = E|X - s|$ assumes its minimum at s_0 if and only if $P(X < s_0) \leq \frac{1}{2} \leq P(X \leq s_0)$

A number s_0 satisfying (4.37.11) is called a *median* for X. Setting $g(s,t) = |s - t|^2$ we are back in our original case, and we find $G'(s) = 2(s - EX)$. Hence the usual expectation is the only minimum of $G(s) = E(s - X)^2$. Setting $g(s,t) = |s - t|^4$ we find

(4.37.12) $$G'(s) = 4(s^3 - 3s^2EX + 3s^2EX^2 - EX^3)$$

which is a cubic polynomial. Using Cardano's formula for the roots of a cubic equation, we find that $G'(s) = 0$ has a unique real root s_0, called *the biquadratic mean* of X, given by

(4.37.13) $$s_0 = \mu + \sqrt[3]{\tfrac{1}{2}\kappa + \sqrt{\tfrac{1}{4}\kappa^2 + \sigma^6}} + \sqrt[3]{\tfrac{1}{2}\kappa - \sqrt{\tfrac{1}{4}\kappa^2 + \sigma^6}}$$

where $\mu = EX$, $\sigma^2 = \operatorname{var} X$ and $\kappa = E(X - \mu)^3$. Thus, alternative expectations may look rather nasty. Setting $I =]0, \infty[$ and $g(s,t) = (\log \frac{s}{t})^2$, we find

$$G'(s) = 2s^{-1}(\log s - E \log X)$$

Hence, the unique solution to $G'(s) = 0$ is given by

(4.37.14) $$s_0 = e^{E \log X}$$

which is called *the geometric mean* of X [the usual expectation EX is also called *the arithmetic mean* of X].

Let me finish with another alternative expectation which goes back to Euler and Lambert (c.1750). If X is a random variable, we define

$$\text{(4.37.15)} \quad \operatorname{ess\,sup} X = \inf\{a \in \mathbf{R} \mid P(X \leq a) = 1\}$$

$$\text{(4.37.16)} \quad \operatorname{ess\,inf} X = \sup\{a \in \mathbf{R} \mid P(X \leq a) = 0\}$$

with the conventions: $\inf \emptyset = \infty$, $\sup \emptyset = -\infty$. Note that if $a = \operatorname{ess\,inf} X$ and $b = \operatorname{ees\,sup} X$, then $[a, b]$ is the smallest closed interval containing almost all values of X . The interval $I = [a, b]$ is called the *range* of X . Note that

$$\text{(4.37.17)} \quad ||X||_\infty = \operatorname{ess\,sup}|X|$$

and a straightforward argument shows that $G(s) = ||s - X||_\infty$ assumes its minimum at s_0 , where

$$\text{(4.37.18)} \quad s_0 = \tfrac{1}{2}(\operatorname{ess\,sup} X + \operatorname{ess\,inf} X)$$

This number, which is the midpoint of the range of X , is called *the midrange* of X . Hence, if we measure distances by $||\cdot||_\infty$, then the midrange becomes the alternative expectation.

Are you confused? You need not be so, but you should realize that the mathematical expectation defined by (4.1.1) only reflects some, but not all, aspects of the intuitive concept of expectation. The expectation is often interpreted as the typical value of X . However, it may not be typical at all. For instance, let X denote the number eyes in one throw with a fair die. Then $EX = 3.5$, but I am sure that you have never got 3.5 eyes in one throw with a die. In this case the midrange equals 3.5, the set of medians equal the interval $[3, 4]$, the geometric mean equals 2.9938, and the biquadratic mean [see (4.37.12)] equals 3.5.

4.38 Averages

It has nowadays become so common to take averages of practically everything that nobody seems to question the reason for doing so. Yet 200 years ago the justification of taking averages was vigorously discussed among scientists, and many eminent scientists were found among the opponents of averages. Well, what is the problem? In its simplest form, we have a series of measurements, say $a_1, \ldots, a_n$ of the same quantity (for instance, the meridian altitude of a star). Today it is commonly accepted that *the average* $\frac{1}{n}(a_1 + \cdots + a_n)$ is the best approximation of "the true value" of the quantity based on the observations

$a_1, \ldots, a_n$. But why is that so? Let us go back in time and trace the history of averages.

The first real application of averages I have found was done by a Danish nobleman and astronomer Tycho Brahe (1546 – 1601). He seems to have been a very unpleasant person who was feared by his peasants because of his severity, and after a long quarrel with the Danish king Christian IV he was exiled from Denmark in 1597; he died in Prague in 1601. But he was an excellent astronomer, and he made a catalog of the meridian altitudes of stars, which was of an unseen precision. The telescope was invented a few years after the death of Tycho Brahe, but it took about 100 years before astronomers with telescopes could measure with the same accuracy as achieved by Tycho with the naked eye. Tycho usually made five observations and took the average of these observations as the "true value." Many astronomers before him made several observations of the same quantity, but they usually chose that observation they believed most in as "the true value". The basic question is whether the average is "better" than the "best" of the observations. Astronomers were arguing about this question for centuries.

The breakthrough came in 1801. On New Years Eve 1800-1801 the Italian astronomer Guiseppe Piazza (1746 – 1826) discovered a new planet which he named Ceres [actually, Ceres is the largest of the so-called planetoids orbiting in a belt between Earth and Mars]. Every night, Piazza made observations of his new planet until mid-February, when Ceres came so close to the Sun that observations were no longer possible. At the time when Ceres was expected to be visible again, Piazza and the other astronomers in Europe could not find it. All summer and fall went on without any trace of Ceres, but in November 1801 the eminent mathematician Karl Friederich Gauss (1777 – 1855) published a paper where computed the orbit of Ceres based on the observations of Piazza. That night the astronomers in Europe turned their telescope at the point on the sky where Gauss had predicted that Ceres should be, and there it was! Gauss had used "the method of least squares," which he had invented a few years before, to compute the orbit.

Let us look at little on the problem of finding the orbit of Ceres. Piazza's observations gave a series of positions $x_1, \ldots, x_n$ at times $t_1, \ldots, t_n$. We know from the laws of Kepler that the orbit is an ellipse with the Sun in one of the two focuses. So let us parametrize the set of all ellipses having the Sun as one focus in some way say $\gamma(t, \theta)$, where θ is a 5-dimensional vector [2 coordinates to describe the half axes, and 3 coordinates to describe the orbit plane and the direction of the major axis]. The problem is to find that value of θ , usual called $\hat{\theta}$, for which the two vectors $(x_1, \ldots, x_n)$ and $(\gamma(t_1, \theta), \ldots, \gamma(t_n, \theta))$ are as "near" as possible. However, as in (4.3.4) there is no canonical measure of "nearness" and different choices lead to different answers.

Setting the problem in a general framework, we have a given *observation vector* $x = (x_1, \ldots, x_n)$, a *theoretical position vector* $\gamma(\theta) = (\gamma_1(\theta), \ldots, \gamma_n(\theta))$ parametrized by the parameter θ , and *a distance measure* $g : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}_+$. Knowing the observation vector x , we seek that value of θ , usually denoted $\hat{\theta}$, which satisfies

$$(4.38.1) \qquad g(x, \gamma(\hat{\theta})) = \min_{\theta} g(x, \gamma(\theta))$$

That is, we seek the value of θ which minimizes "the distance" from the observation x to the theoretical value $\gamma(\theta)$. If the parameter set equals $\mathbf{R}$ and $\gamma_1(\theta) = \cdots = \gamma_n(\theta) = \theta$, then we are back to our original problem of finding "the true value" of $x_1, \ldots, x_n$.

The minimax method was first suggested by Euler and Lambert (≈ 1750). Euler and Lambert used the ℓ^∞-metric on $\mathbf{R}^n$, that is,

$$g(x, y) = ||x - y||_\infty = \max_{1 \le j \le n} |x_j - y_j|$$

and so (4.36.1) takes the form

$$(4.38.2) \qquad ||x - \gamma(\hat{\theta})||_\infty = \min_{\theta} \max_{j} |x_j - \gamma_j(\theta)|$$

If the parameter set equals $\mathbf{R}$ and $\gamma_1(\theta) = \cdots = \gamma_n(\theta) = \theta$, then the solution to (4.38.2) equals *the sample midrange* of $x_1, \ldots, x_n$; that is,

$$(4.38.3) \qquad \hat{\theta} = \tfrac{1}{2}(\max_j x_j + \min_j x_j)$$

In 1783, Laplace gave a practical method for solving (4.38.2) for certain types of functions $\gamma(\theta)$, and six years later he simplified the method considerably. Today, the minimax method is widely used in mathematical economics, and many election methods are founded on a minimax method.

The median method was first suggested by Boškovich in 1755. Boškovich used the ℓ^1-métric on $\mathbf{R}^n$; that is,

$$g(x, y) = ||x - y||_1 = \sum_{j=1}^{n} |x_j - y_j|$$

and so (4.38.1) takes the form

$$(4.38.4) \qquad ||x - \gamma(\hat{\theta})||_1 = \min_{\theta} \sum_{j=1}^{n} |x_j - \gamma_j(\theta)|$$

If the parameter set equals $\mathbf{R}$ and $\gamma_1(\theta) = \cdots \gamma_n(\theta) = \theta$, then $\hat{\theta}$ is a solution to (4.30.1) if and only if $\hat{\theta}$ is a *sample median* of $x_1, \ldots, x_n$, that is, if and only if

$$\#\{1 \le j \le n \mid x_j < \hat{\theta}\} \le \frac{n}{2} \le \#\{1 \le j \le n \mid x_j \le \hat{\theta}\} \tag{4.38.5}$$

The method has never been commonly used, but in recent times it has found applications in certain statistical problems.

The method of least squares was first found by Gauss ($\approx$ 1795), but he did not publish his results until 1809; he was thereby overtaken by the French mathematician Legendre (1808) and the American surveyor Adrain (1808). In this case we use the Euclidean metric in $\mathbf{R}^n$, that is,

$$g(x, y) = ||x - y||_2 = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2}$$

and so (4.38.1) takes the form

$$(||x - \gamma(\hat{\theta})||_2)^2 = \min_{\theta} \sum_j (x_j - \gamma_j(\theta))^2 \tag{4.38.6}$$

In this case, $\hat{\theta}$ can be found by solving the equation:

$$\sum_{j=1}^{n} (\gamma_j(\theta) - x_j) \frac{\partial \gamma_j}{\partial \theta_i}(\theta) = 0 \qquad \forall i = 1, \ldots, k \tag{4.38.7}$$

where k is the dimension of the parameter set. If the parameter set equals $\mathbf{R}$ and $\gamma_1(\theta) = \cdots = \gamma_n(\theta) = \theta$ it is easily seen that the solution to (4.38.6) is *the average*, also called *the sample mean*, of $x_1, \ldots, x_n$; that is,

$$\hat{\theta} = \frac{1}{n} \sum_{j=1}^{n} x_j \tag{4.38.8}$$

Today, this method is by far the most common method for two reasons. First, it was the method which Gauss applied so successfully to find the orbit of Ceres, and in this way he convinced his contemporaries about the power of the method. Second, contrary to the minimax method and the median method, it is easy to find the solution to (4.38.6) by solving (4.38.7). That is the history of averages, and that is, why the average is so common today.

The geometric method corresponds to the method of least squares after having taking logarithms of the observation. In this case we use the function

$$g(x, y) = \sum_{j=1}^{n} \left(\log \frac{y_j}{x_j}\right)^2 \qquad x_1, \ldots, x_n > 0 \ , \ y_1, \ldots, y_n > 0$$

and so (4.38.1) takes the form

$$(4.38.9) \qquad \sum_{j=1}^{n} \left(\log \frac{\gamma_j(\hat{\theta})}{x_j} \right)^2 = \min_{\theta} \left\{ \sum_{j=1}^{n} \left(\log \frac{\gamma_j(\theta)}{x_j} \right)^2 \right\}$$

The solution to (4.38.9) may be found by using the method of least squares on the numbers $y_j = \log x_j$ and the parameter functions $\eta_j(\theta) = \log \gamma_j(\theta)$. If the parameter set equals $]0, \infty[$ and $\gamma_1(\theta) = \cdots = \gamma_n(\theta) = \theta$, then the solution to (4.30.6) is *the geometric sample mean* of $x_1, \ldots, x_n$; that is,

$$(4.38.10) \qquad \hat{\theta} = \left\{ \prod_{j=1}^{n} x_j \right\}^{1/n}$$

Thus, there is nothing canonical about the average being the best approximation to "the true value" of a set of observations. It has proved its right of existence by the simplicity and the success of the method of least squares. But it is important to understand that there are many other methods of equal right, which in special cases may give a better description of the observations. We shall later see how the methods above relate to the so-called maximum likelihood method of statistics.

4.39 The Petersburg Paradox

In 1738, the probabilist Daniel Bernoulli (1700 – 1782) published a paper in *Commentarii Academiae Scientiae Imperialis Petropolitane* containing a problem which later was called *the Petersburg paradox*, and which led to a discussion of the justification of the mean of a random variable. The problem goes as follows.

Our good friends and gamblers Ann and Ben meet, and Ben suggests a bet. Ann pays a certain amount, say $\$a$ to Ben and a fair coin is thrown until head comes up for the first time. If this takes n throws, then Ben pays $\$2^{n-1}$ to Ann. The question is: What is the reasonable price a that Ann should pay Ben for joining the bet? If the bet should be fair, then according to the usual rule the price a should be equal to the expected winnings. Let X denote the number of throws until the first head comes up, then X takes the values $1, 2, \cdots$, and $P(X = n) = 2^{-n}$ for all $n \in \mathbb{N}$. Since Ann receives $\$2^{X-1}$, we see that the expected winnings are given by

$$E2^{X-1} = \sum_{n=1}^{\infty} 2^{n-1} P(X = n) = \sum_{n=1}^{\infty} \tfrac{1}{2} = \infty$$

Hence, according to the usual rules, Ann should pay an infinite sum for joining the bet. Are you ready to pay \$ 10.000 for joining that bet? The mathematician

d'Alembert was not, and the "paradox" led to a long discussion between Daniel Bernoulli and d'Alembert concerning the foundations of probability.

Personally I will pay at most \$ 2 for the bet based on the following reasoning: Let X be as above, then $EX = \sum_1^\infty n\,2^{-n} = 2$. Thus, on an average it takes two throws to make a head, and if so then I win \$ 2.

Daniel Bernoulli came to the same result but for a quite different reason. Daniel says that $\$x$ do not have the same value for a poor person as for a rich person. Then he goes on and makes a mathematical model for what he calls *the moral value* of *the physical value* $\$x$ when you fortune equals $\$a$. He says that if your fortune equals x , then the real (or moral) value of a small increment dx is proportional to dx and inversely proportional to the fortune. Hence, if dy is the increment in the moral value, then $dy = \frac{k\,dx}{x}$ where k is a proportionality factor. Thus, if $M(x)$ is the moral value of $\$x$, then

$$M(x) = k \int \frac{dx}{x} = k \log \frac{x}{A}$$

where k and A are (unknown) constants. Suppose that Ann has a fortune of $\$r$, and that she has chances $p_1, p_2, \ldots, p_n$ of winning $x_1, \ldots, x_n$. Then

$$y = \sum_{j=1}^{n} k\,p_j \log \frac{r + x_j}{A}$$

is the expected moral value of Ann's fortune after the bet. The physical value x associated with the moral value y is found by the formula $y = k \log \frac{x}{A} \iff x = A\,e^{y/k}$. Hence, the expected physical value associated to y is

$$x = Ae^{y/k} = (r + x_1)^{p_1}(r + x_2)^{p_2} \cdots (r + x_n)^{p_n} \tag{4.39.1}$$

The expected physical gain equals $x - a$. Note that if X is a random variable taking the values $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$ then

$$e^{E \log X} = (r + x_1)^{p_1}(r + x_2)^{p_2} \cdots (r + x_n)^{p_n}$$

Thus, in this way Daniel arrives at the conclusion that you should use the geometric mean [see (4.37.14)] rather that the usual mean.

Applying this to the Petersburg game we see that if the total fortune of Ann equals r , then the expected value of the winnings is given by

$$v(r) = \exp\left\{E \log(r + 2^{X-1})\right\} - r \tag{4.39.2}$$

where X is the number of throws it takes to make a head. If $r = 0$, we find

$$\begin{aligned} v(0) &= \exp(E \log 2^{X-1}) = \exp(\log 2\, E(X-1)) \\ &= \exp(\log 2) = 2 \end{aligned}$$

Thus, if Ann is very poor, then Daniel's advice to Ann is to bet at most \$ 2. Daniel computes the value of $v(r)$ for $r = 0, 10, 100$ and 1000 and he gives the following values:

$$v(0) = 2 \qquad v(10) \simeq 3 \qquad v(100) \simeq 4.333 \qquad v(1000) \simeq 6$$

In Table 4.1 you will find a table of the exact values of $v(r)$.

Table 4.1

The moral expectation $v(r)$ of the Petersburg game as a function of the capital r

r	0	10	100	500	1,000	10,000	1,000,000
$v(r)$	2	3.043	4.389	5.485	5.972	7.618	10.933

Daniel Bernoulli also used the geometric mean to solve the so-called *insurance paradox*. If you carry an insurance, say against burglary, then you make a bet with the insurance company: If you have a burglary in your house you win the indemnity; if not you lose the insurance premium. Usually, it is not advisable to make a bet on unfair odds, but the insurance company has to make a profit and so the odds must necessarily be unfavorable to you. But why should you insure? And if so, how unfair odds should you accept? Daniel solves this in the following way (in modern terminology):

Let r be your total fortune, and let X be the value of the items which are stolen from your house within one year. Then X is a non-negative random variable, and your fortune at the end of that year equals $r - X$. According to Daniel, the physical value of the expected moral loss in that year equals

$$\lambda(r) = r - \exp(E \log(r - X)) \tag{4.39.3}$$

which is also the fair premium. Since $\exp(x)$ is convex, then by Jensen's inequality (3.10.2), we have

$$\lambda(r) \geq r - E \exp(\log(r - X)) = EX \tag{4.39.4}$$

and $\lambda(r) - EX$ is the profit of the insurance company, which by (4.39.4) is positive. Letting $r \to \infty$, we get

$$\begin{aligned}\lim_{r\to\infty} \lambda(r) &= \lim_{r\to\infty} r\Big(1 - \exp\Big(1 - \frac{X}{r}\Big)\Big) \\ &= -\frac{d}{dt}\exp(\log(1 - tX))\Big|_{t=0} = EX\end{aligned}$$

Thus, the conclusion is that a rich person should not accept unfair odds, but a poor person has to do so.

4.40 The Anomal Distribution

A statistician once had great success on betting that a random number chosen in *The Danish Statistical Yearbook* has the first significant digit equal to $1, 2, 3$ or 4. At first glance you would expect that the nine digits $1, 2, \ldots, 9$ are equally probable; that is, the first significant digit of a random chosen number is uniformly distributed over $\{1, \ldots, 9\}$, and so the statistician has probability $\frac{4}{9} = 44\%$ of winning the bet. However, actual experiments shows that the so-called anomal distribution fits much better than the uniform distribution. The *anomal distribution* is the discrete distribution on $\{1, \ldots, 9\}$ having probability mass function given by

$$p_k = \log_{10}(1 + \tfrac{1}{k}) = \log_{10}(k+1) - \log_{10} k \qquad \text{for } k = 1, \ldots, 9 \tag{4.40.1}$$

where $\log_{10}$ is the logarithm to the base 10. Thus, the claim is that the probability of getting the digit k equals p_k as given by (4.40.1). So let us take a closer look at the problem.

Let X be a positive random variable, and let N denote the first significant digit of X. If we set $Y = \log_{10} X$ and $R = frac(Y)$, where $frac(x) = x - int(x)$ is the fractional part of x and $int(x)$ is the integer part of x, then we have $N = int(10^R)$. So we find

$$P(N = k) = P(\log_{10} k \le R < \log_{10}(k+1)) \quad \forall k = 1, 2 \ldots 9 \tag{4.40.2}$$

Note that R takes values in the unit interval $[0, 1[$. If R is approximately uniformly distributed over $[0, 1[$, then by (4.40.2) we see that N is approximately anomally distributed.

Table 4.2

100 random first digits from The Danish Statistical Year Book 1980									
Digits	1	2	3	4	5	6	7	8	9
Obs.	28	24	18	8	8	7	3	2	2
Uniform	11	11	11	11	11	11	11	11	11
Anomal	30	18	12	10	8	7	6	5	5

"Obs" are the observed numbers.
"Uniform" are the expected numbers if the digits are equally probable.
"Anomal" are the expected numbers if the digits are anomally distributed.

Thus, we have two models describing the distribution of the first significant digit of a randomly chosen number: Model (a) claims that the digit is uniformly distributed over $\{1, \ldots, 9\}$ and model (b) claims that the fractional part of the logarithm of the random number is uniformly distributed over $[0, 1]$. So which one should we choose? There is only one way to choose between two sensible probabilistic models: Make experiments and confront the outcome with the two models. I have performed the experiment: I chose a random number 100 times from *The Danish Statistical Yearbook (1980)*; the result is represented in Table 4.2. Clearly, the anomal distribution fits much better to the observed data than the uniform distribution; see also F. Benford, *Proc. Amer. Phil. Soc.* 78 (1938), pp.551-572, where you find an extensive empirical material.

Let us now see what the assumption that R is approximately uniformly distributed over $[0, 1]$ means in terms of the distribution or rather the density of X . Suppose that X is absolutely continuous with density f . Let $h(x) = xf(x)$ and let v be *the total variation* of h over $]0, \infty[$, that is,

$$v = \sup \sum_{j=1}^{n} |h(x_j) - h(x_{j-1})|$$

where we take the supremum over all $0 < x_0 < x_1 < \cdots < x_n$. Note that Y has density function $g(x) = c \cdot h(10^x)$ where $c = \log 10$ and that R has density

function $m(x) = \sum_{j=-\infty}^{\infty} g(x+j)$. Let $0 < x < 1$ be given. Then we have

$$|m(x) - 1| = \left| m(x) - \int_0^1 m(y)\,dy \right| \le \sum_{j=-\infty}^{\infty} \left| g(j+x) - \int_0^1 g(j+y)\,dy \right|$$
$$\le \sum_{j=-\infty}^{\infty} \int_0^1 |g(j+x) - g(j+y)|dy \le c\,v$$

and since $c = \log 10 = 2.3026$ and $c \cdot \log_{10} 5 = 1.6094$. Then by (4.40.2) we have

(4.40.3) $\quad |P(N = k) - p_k| \le 2.3026 \cdot v \cdot p_k \qquad \forall 1 \le k \le 9$

(4.40.4) $\quad |P(N \le 4) - \log_{10} 5| \le 1.6094 \cdot v$

Hence, if v is small, then N is approximately anomally distributed and $P(N \le 4) \simeq \log_{10} 5 = 0.6990 = 70\%$. The moral of this tale is: Never make a bet with a statistician

Table 4.3 presents the total variation v of $xf(x)$ for five common distributions on $\mathbf{R}_+$. The *Pareto distribution* $Par(\alpha, q)$ for $\alpha, q > 0$, occurring in Table 4.3, is the distribution given by the density function

$$f_{\alpha q}(x) = \begin{cases} \alpha\, q^{\alpha}(q+x)^{-\alpha-1} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

Table 4.3

The total variation v ***of*** $x\,f(x)$ ***for five common densities***					
$f(x)$	$E(\lambda)$	$\Gamma(\alpha,\beta)$	$Par(\alpha,q)$	$\log N(\mu,\sigma^2)$	$R(0,a)$
v	$\frac{2}{e\lambda}$	$\frac{2}{\Gamma(\alpha)}\left(\frac{\alpha}{e\beta^2}\right)^{\alpha}$	$2\left(\frac{1}{q}+\frac{1}{\alpha}\right)^{-\alpha-1}$	$\sqrt{\frac{2}{\sigma\pi}}$	2

4.41 Bad Luck

Our good friend Ben is depressed. He claims that he is the most unlucky person in the world and, moreover, he claims that probability theory proves it. His arguments goes as follows: He visits a gambling house and plays one game, then he lets his lucky neighbor Ann play the same game the number of times it takes before Ann is more unlucky than him. He claims that in average he has to wait infinitely long time before this happens.

Suppose that X_0 is Ben's profit (positive or negative) and let $X_1, X_2, \cdots$ denote Ann's profit in the first, second, etc. game. Then

$$N(\omega) = \min\{n \in \mathbf{N} \mid X_n(\omega) < X_0(\omega)\}$$

is the time Ben has to wait before Ann is more unlucky than himself. Suppose that $X_0, X_1, X_2, \cdots$ are independent random variables with the same distribution function F. Then by (4.5.3) we have

$$\begin{aligned} P(N > n) &= P(X_0 \le X_1, \ldots, X_0 \le X_n) \\ &= \int_{-\infty}^{\infty} P(t \le X_1, \ldots, t \le X_n) F(dt) \\ &= \int_{-\infty}^{\infty} (1 - F(t-))^n F(dt) \end{aligned}$$

where $F(t-) = \lim_{s \uparrow t} F(s)$. Since $g(x) = (1-x)^n$ is decreasing on $]0,1[$, by (3.29.7) we have

$$P(N > n) \ge \int_0^1 (1-x)^n \, dx = \frac{1}{n+1}$$

So by (4.2.9) we have

$$EN = \sum_{n=0}^{\infty} P(N > n) \ge \sum_{n=0}^{\infty} \frac{1}{n+1} = \infty$$

Thus, Ben is right, and he is the most unluckily person in the world.

4.42 References to Chapter 4

Lukacs' book [3] contains an exhaustive treatment of characteristic functions and their properties. Widder's book [6] is the standard handbook of Laplace transforms. The Laplace transform and the characteristic functions offers an

extension to semigroups, which you can find in [2] where you also find an extensive discussion of the so-called *moment problem*; that is, the problem of deciding when a distribution is uniquely determined by its moments. The law of large numbers is not a single theorem but a subject which has been extensively studied ever since James Bernoulli found the first version in 1695. In [5] you will a thorough treatment of the laws of large numbers for real-valued random variables.

[1] Berg, C. (1988): *The Cube of a Normal Distribution Is Indeterminate*, Ann. Prob. 16, pp. 910–913

[2] Berg, C., Christensen, J. P. R. and Ressel, P. (1984): *Harmonic Analysis on Semigroups*, Springer Verlag, Berlin and New York

[3] Lukacs, E. (1970): *Characteristic Functions*, Charles Griffin & Co. Ltd., London

[4] Haagerup, U. (1979): *Les Meilleures Constantes de l'Inéqualité de Khinchine*, C.R: Acad. Sci. Paris 286 Ser. A, pp. 259–262

[5] Révész, P. (1968): *The Laws of Large Numbers*, Academic Press, New York and London

[6] Widder, D. V. (1972): *The Laplace Transform*, (eighth ed.), Princeton University Press, Princeton

4.43 Exercises to Chapter 4

Exc. 4.1: Let $X_1, X_2, \ldots$ be a sequence of exponentially distributed random variables with parameters $\lambda_1, \lambda_2, \ldots$. Suppose that $\sum_1^\infty \frac{1}{\lambda_n} < \infty$. Show that $\sum_1^\infty X_n$ is summable a.s.

Exc. 4.2: Let X, Y and Z be independent random variables all having a gamma distribution $\Gamma(\alpha, \beta)$ where $\alpha, \beta > 0$ are given positive numbers. Find the correlation coefficient $\rho(X + Z, Y - Z)$.

Exc. 4.3: Let X, Y and Z be independent random variables all having an exponential distribution $E(\lambda)$ where $\lambda > 0$ is a given positive number. Find the correlation coefficient $\rho(U, V)$ where $U = X \wedge Y \wedge Z$ and $V = X \vee Y \vee Z$.

Exc. 4.4: A die is thrown until "6" has come up N times. Then we continue until an odd number of points has come up M times, where N and M are given integers. Find the expected number of throws.

Exc. 4.5: A particle is shot into a wall of infinite thickness. The penetration depth is measured and is called X . Below you will find four reasonable probabilistic models describing this experiment:

(1): Let us suppose that the particle is so small that it can pass between the molecules of the wall with being slowed down, but that it is absorbed if it hits one of the molecules of the wall. Moreover, let us suppose that the wall consists of thin shells of thickness Δ such that the conditional probability of passing the nth shell respectively being absorbed in the nth shell, given that it has passed the $(n-1)^{\text{th}}$ shell, equals $1-\rho\Delta$ respectively $\rho\Delta$, where $\rho > 0$ is a given parameter (depending of the mass and speed of the particle, and the structure of the wall). Show that

$$\text{(i)} \qquad P(X = n\Delta) = \rho\,\Delta\,(1-\rho\,\Delta)^{n-1} \qquad \forall n = 1,2,\ldots$$

(2): Consider the model described in (1). Let $\Delta \to 0$ in an appropriate way and obtain a new model in which X gets a density function. Find the density function of X .

(3): Consider the model described in (1), but let us now suppose that the particle is so large that its speed decreases as it penetrates the wall. More specific, suppose that if the particle has passed the $(n-1)^{\text{th}}$ shell and has speed v , the probability of absorption respectively passing the nth shell equals $\frac{\rho\Delta}{v}$ respectively $1-\frac{\rho\Delta}{v}$, and in the latter case the speed is decreased by $\rho\Delta$. Suppose that the initial speed of the particle equals $N\rho\Delta$ where $N \geq 1$ is a given integer. Show that

$$\text{(ii)} \qquad P(X > n\Delta) = \frac{N-n}{N} \qquad \forall n = 1,2,\ldots,N$$

(4): Consider the model described in (3). Let $\Delta \to 0$ in an appropriate way and obtain a new model in which X gets a density function. Find the density function of X .

Exc. 4.6: Let (X,Y) be a two-dimensional random vector with density function

$$f(x,y) = \begin{cases} x+y & \text{if } x,y \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Find the mean vector $\mu = (\mu_1,\mu_2)$ and the covariance matrix $\sigma = \{\sigma_{ij}\}$ of (X,Y) .

Exc. 4.7: Let X be random variable with finite mean μ and finite variance σ^2 . Show that X is degenerated if and only if $\sigma^2 = 0$; and if so, then $X = \mu$ a.s.

Exc. 4.8: A match of length r is placed at random on a line segment of length $s > r$ such that the midpoint of the of the match is uniformly distributed on the line segment. What is the probability that the two end points of the match lies inside the line segment?

Exc. 4.9: We are observing a machine. The time, say T, until the first breakdown is measured. It is supposed that the conditional probability that machine breaks down in the time interval $]x, x + \Delta]$, given that it functions at time x, approximately equals $\Delta\, g(x)$ for Δ small, where the proportionality factor $g(x)$ is a function of the time x; that is, we are assuming

$$P(x < T \leq x + \Delta \,|\, T > x) \approx \Delta\, g(x) \quad \text{as} \quad \Delta \to 0 \quad \forall x > 0$$

Show that T has density function f given by

$$\text{(i)} \qquad f(x) = g(x) \exp\left\{ -\int_0^x g(t)\, dt \right\} \qquad \forall x > 0$$

Note that if $g(x) \equiv \lambda$, then we obtain an exponential distribution $E(\lambda)$ [see (2.22)]. Usually g is increasing, meaning that the machine becomes more fragile with age.

Hint: Let $R(x) = P(T > x)$ and show that R satisfies the differential equation: $R'(x) + g(x)\, R(x) = 0$. Solve this differential equation observing that $R(0) = 1$.

Exc. 4.10: A random point (X, Y) is chosen according to a uniform distribution in the unit square $S = [0,1] \times [0,1]$. Are X and Y independent? Let the random point (U, V) be chosen according to a uniform distribution on the unit circle $C = \{(x, y) \,|\, x^2 + y^2 \leq 1\}$. Are U and V independent?

Exc. 4.11: A point (U, V) is chosen according to the uniform distribution on the unit circle $C = \{(x, y) \,|\, x^2 + y^2 \leq 1\}$. Find the marginal distributions of U and V. Let (R, Θ) be the polar coordinates of (U, V), that is, $0 \leq R \leq 1$, $-\pi < \Theta \leq \pi$, $U = R \sin\Theta$ and $V = R \cos\Theta$. Show that R and Θ are independent and find their distributions.

Exc. 4.12: We are shooting particles at a circular disk of radius 1. It is observed that all particles hit the disk and that the intensity is inversely proportional to the distance from the center to the hitting point. Make a reasonable probabilistic model for this experiment; that is, find a density for the hitting point which fits the observed behavior of the intensity.

Exc. 4.13: Let $\{\varepsilon_n\}$ be a sequence of independent, identically distributed random variables with $P(\varepsilon_n = 0) = P(\varepsilon_n = 3) = \frac{1}{2}$ for all $n \geq 1$. Let

$$X = \sum_{n=1}^{\infty} 4^{-n} \varepsilon_n \ , \ Y = \sum_{n=1}^{\infty} 2^{-n} \varepsilon_n \ , \ U = \sum_{n=1}^{\infty} 4^{-n} \varepsilon_{2n} \ , \ V = \sum_{n=1}^{\infty} 4^{-n} \varepsilon_{2n-1}$$

Show that X, Y, U and V are random variables such that $0 \leq X \leq 1$, $0 \leq Y \leq 3$, $0 \leq U \leq 1$ and $0 \leq V \leq 1$. Let $F(t) = P(X \leq t)$ be the distribution function of X , and show the following:

(i) X , U and V have the same distribution $F$and U and V are independent

(ii) Y is uniformly distributed on $[0, 3]$

(iii) $Y = U + 2V$ and U and V are independent

(iv) F is continuous on $\mathbf{R}$ and F is differentiable at t with $F'(t) = 0$ for Lebesgue a.a. $t \in \mathbf{R}$

(v) Y is absolutely continuous, but X , U and V are **not** absolutely continuous

Exc. 4.14: Let $X_1, \ldots, X_n$ be random variables and let $0 < q < \infty$. Show that

(i) $$E \left| \sum_{j=1}^{n} X_j \right|^q \leq n^{(q-1)^+} \sum_{j=1}^{n} E|X_j|^q$$

Hint: Apply Minkovski's inequality (3.12.3) if $0 < q \leq 1$, and apply convexity of $x \curvearrowright |x|^q$ if $q > 1$.

Exc. 4.15: Let X be a random variable. Show that

(i) $P(|X| \geq a) \leq a^{-p} E\,|X|^p \qquad \forall a, p > 0$

(ii) $P(|X| \geq a) \leq e^{-at} E\, e^{t|X|} \qquad \forall a, t > 0$

Exc. 4.16: (*The one-sided Markov inequality I*) Let X be a random variable and let $\varphi : \mathbf{R} \to [0, \infty[$ be a Borel function. Let $a \in \mathbf{R}$ such that φ is increasing on $[a, \infty[$ and $\varphi(a) > 0$. Show that

(i) $$P(X \geq a) \leq \frac{1}{\varphi(a)} \int_{\{X \geq a\}} \varphi(X)\, dP \leq \frac{E\,\varphi(X)}{\varphi(a)}$$

In particular, show that

(ii) $\quad P(X \geq a) \leq e^{-\lambda a} E\, e^{\lambda X} \qquad \forall \lambda > 0\ \forall a \in \mathbf{R}$

Exc. 4.17: Let X be a random variable with finite mean μ and finite variance σ^2 . Let m be a *median* for X , that is, $P(X < m) \leq \frac{1}{2} \leq P(X \leq m)$. Show that $|m - \mu| \leq \sigma\sqrt{2}$.

Exc. 4.18: *(The one-sided Markov inequality II)* Let X be a random variable, and let $g : \mathbf{R} \to \mathbf{R}$ be a Borel function such that $g(x) \geq b$ for all $x \geq a$, where $a, b \in \mathbf{R}$ are given numbers with $b > 0$. Show that

(i) $\quad P(X \geq a) \leq \dfrac{1}{b} \displaystyle\int_{\{X \geq a\}} g(X)\, dP$

Exc. 4.19: Let X be a random variable, and let $g : \mathbf{R} \to \mathbf{R}$ be a Borel function such that $g(x) \leq K < \infty$ for all $x \in \mathbf{R}$. Let $a \in \mathbf{R}$ such that $g(x) \leq C$ for all $x \leq a$ where $C \leq K$ is a given real number. Show that

(i) $\quad E\big(g(X) - C\big)^+ \leq (K - C)\, P(X > a)$

Exc. 4.20: *(Bienaymé's inequality)* Let X be a random variable with $\mu \geq 0$ and $0 < \tau < \infty$ where $\mu = EX$ and $\tau = EX^2$. Show that

(i) $\quad P(X > t\mu) \geq \dfrac{(1-t)^2\, \mu^2}{\tau} \qquad \forall 0 \leq t \leq 1$

Hint: Set $A = \{X > t\mu\}$ and $B = A^c = \{X \leq t\mu\}$. Show that $X \leq t\mu\, 1_B + X\, 1_A$ and that $\mu \leq t\mu + \sqrt{\tau\, P(A)}$ [apply Hölder's inequality (3.11)]..

Exc. 4.21: Let X be a random variable such that $P(a \leq X \leq b) = 1$ where $a, b \in \mathbf{R}$ are given numbers with $a < b$. Let μ and σ^2 be the mean and variance of X . Show that $\sigma^2 \leq (\mu - a)(b - \mu)$, and that we have equality if and only if $P(X = a) + P(X = b) = 1$.

Hint: Show that $E\,(X - a)(b - X) = (\mu - a)(b - \mu) - \sigma^2$

Exc. 4.22: Let X be a random variable with $P(X \in I) = 1$ where $I \subseteq \mathbf{R}$ is an interval, and let $\varphi : I \to \mathbf{R}$ be a convex function. Suppose that the mean EX exists and is finite. Show that $EX \in I$ and $\varphi(EX) \leq E\varphi(X)$. Moreover, show that we have $\varphi(EX) = E\varphi(X)$ if and only if $P(X \in L_\mu \cap R_\mu) = 1$, where $\mu = E\,X$ and L_m and R_m are defined as in Exc.3.34. [Apply Exc.3.34 with a suitable choice of μ , f and g .]

Exc. 4.23: Let $X \in L^1(P)$ be a random variable. Show that

(i) $E\{\frac{1}{X}\} \geq \frac{1}{E\,X}$ if $X \geq 0$ a.s.

(ii) $E\,e^X \geq e^{E\,X}$

(iii) $E \log X \leq \log E\,X$ if $X \geq 0$ a.s.

(iv) $E(X \log X) \geq (E\,X) \cdot \log E\,X$ if $X \geq 0$ a.s.

where we in (iv) use the convention: $0 \cdot \log 0 = 0$.

Exc. 4.24: Let $X_1, \ldots, X_n$ be *exchangeable* random variables, that is, $(X_1, \ldots, X_n)$ and $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ have the same distribution for every permutation σ of $1, 2, \ldots, n$. Suppose that $X_j \geq 0$ a.s. for all $j = 1, \ldots, n$, and show that

(i) $$E\left(\frac{X_1}{X_1 + \cdots + X_n}\right) = E\left(\frac{X_j}{X_1 + \cdots + X_n}\right) \qquad \forall j = 1, \ldots, n$$

(ii) $$E\left(\frac{X_1 + \cdots + X_k}{X_1 + \cdots + X_n}\right) = \frac{k}{n} \qquad \forall k = 1, \ldots, n$$

where we use the convention: $\frac{0}{0} = 1$.

Exc. 4.25: Let X be a random variable, and let $k \in \mathbf{N}$ be an integer. Show that

(i) $$E(|X|^{2k+1}) \leq \sqrt{E(X^{2k}) \cdot E(X^{2k+2})}$$

Exc. 4.26: Let X be a random variable such that $P(|X| \leq a) = 1$, where $a > 0$ is a given number. Show that

(i) $$P(|X| > t) \geq \frac{E\,X^2 - t^2}{a^2} \qquad \forall t \geq 0$$

Exc. 4.27: Let $0 < q < \infty$ be a given number and let X and Y be independent random variables with $E|X + Y|^q < \infty$. Show that $E\,|X|^q < \infty$ and $E\,|Y|^q < \infty$.

Hint: Let $h(s) = E|X + s|^q$ for all $s \in \mathbf{R}$. Apply (4.5.2) to show that there exists $a \in \mathbf{R}$ with $h(a) < \infty$, and observe that $|X|^q \leq 2^q(|a|^q + |X + a|^q)$.

Exc. 4.28: Let X and Y be independent non-negative random variables. Show that

(i) $$E\left(\frac{X}{Y + a}\right) = (E\,X) \cdot E\left(\frac{1}{Y + a}\right) \qquad \forall a > 0$$

Exc. 4.29: *(The Weierstrass-Bernstein theorem)* Let $f : [0,1] \to \mathbf{R}$ be a function. Then we define *the global oscillation* of f, denoted $W_f(s)$, and *the nth Bernstein polynomial*, denoted $B_n(f,t)$, as follows:

$$(*) \qquad W_f(s) = \sup\left\{ |f(u) - f(v)| \;\middle|\; u,v \in [0,1]\,,\ |u-v| \le s \right\} \qquad \forall s > 0$$

$$(**) \qquad B_n(f,t) = \sum_{j=0}^{n} \binom{n}{j} f\left(\tfrac{j}{n}\right) t^j (1-t)^{n-j} \qquad \forall t \in \mathbf{R}$$

Let $n \in \mathbf{N}$ and $t \in [0,1]$ be given and let X be a random variable having a binomial distribution $Bi(n,t)$. Show the following:

(i) $\quad B_n(f,t) = E\, f\left(\frac{X}{n}\right)$

(ii) $\quad |f(t) - B_n(f,t)| \le W_f(s) + \dfrac{W_f(1)}{4\,n\,s^2} \qquad \forall s > 0\ \forall 0 \le t \le 1\ \forall n \ge 1$

(iii) $\quad |f(t) - B_n(f,t)| \le W_f\left(\dfrac{1}{\sqrt[3]{n}}\right) + \dfrac{W_f(1)}{4\sqrt[3]{n}} \qquad \forall 0 \le t \le 1\ \forall n \ge 1$

(iv) $\quad$ If f is continuous on $[0,1]$, then $B_n(f,t) \rightrightarrows f(t)$ uniformly on $[0,1]$

Remark: Note that (iv) shows that every continuous function on a compact interval can be approximated uniformly by polynomials. In this form the result is due to Karl Theodor Wilhelm Weierstrass (1815–1897); later Sergey Natanovich Bernstein (1880–1968) provided the specific sequence of approximating polynomials in (**). The Bernstein polynomials converge rather slowly even for highly regular functions; for instance, if $f(t) = t^2$, then $B_n(f,t) = t^2 + \frac{t(1-t)}{n}$, and if f is many times differentiable, then there exists much more efficient polynomial approximations to f. However, the Bernstein polynomials have other virtues. For instance, the function: $f \curvearrowright B_n(f,\cdot)$ is a positive linear function from the set of all continuous real-valued functions on $[0,1]$ into the set of all polynomials of degree $\le n$, and if f is convex (respectively concave), then $\{B_n(f,t)\}$ is decreasing (respectively increasing) in n for all $t \in [0,1]$. Moreover, if f is k-times continuously differentiable in $[0,1]$, then the k first derivatives of $B_n(f,t)$ converges uniformly to the k first derivatives of f.

Hint: (ii): Let $A_s = \{|nt - X| > ns\}$. Then by (i) we have

$$\begin{aligned} |f(t) - B_n(f,t)| &\le E\left|\, f(t) - f\left(\tfrac{X}{n}\right) \right| \\ &= \int_{A_s} \left|\, f(t) - f\left(\tfrac{X}{n}\right) \right| dP + \int_{\Omega \setminus A_s} \left|\, f(t) - f\left(\tfrac{X}{n}\right) \right| dP \end{aligned}$$

and by definition of the global oscillation we have

$$\left| f(t) - f\left(\frac{X(\omega)}{n}\right) \right| \leq \begin{cases} W_f(1) & \text{if } \omega \in A_s \\ W_f(s) & \text{if } \omega \in \Omega \setminus A_s \end{cases}$$

Apply Chebyshev's inequality (4.4.4) to show that

$$P(A_s) \leq \frac{n\,t(1-t)}{(ns)^2} \leq \frac{1}{4\,n\,s^2}$$

(iii): Set $s = n^{-1/3}$ in (ii). (iv): Note that by compactness of $[0,1]$, we have that f is uniformly continuous, which implies that $W_f(s) \to 0$ as $s \to 0$.

Exc. 4.30: Let X be a random variable having finite moment of all orders, and suppose that $E\,X = 0$ and

(*) $\quad E\,|X|^k \leq v\,\rho^{k-2}\,k! \qquad \forall k = 2, 3, \ldots$

where $v, \rho \geq 0$ are given non-negative constants. Show that the Laplace transform $L_X(\lambda)$ exists for all $\lambda \in \mathbf{C}$ with $|\Re\lambda| < \frac{1}{\rho}$ and that

(i) $\quad |L_X(\lambda)| \leq \exp\left(\dfrac{v|\lambda|^2}{1 - |\lambda|\,\rho}\right) \qquad \forall |\lambda| < \frac{1}{\rho}$

Hint: Let $|\Re\lambda| < \frac{1}{\rho}$ and set $\mu = |\Re\lambda|$. Since $|e^z| = e^{\Re z} \leq e^{|\Re z|}$, then by Beppo Levi's theorem (3.32.3) we have

$$E\,|e^{\lambda X}| \leq E\,e^{\mu|X|} = \sum_{k=0}^{\infty} \frac{\mu^k\,E\,|X|^k}{k!}$$

which is finite by (*). Use Beppo Levi again and that $EX = 0$ to show

$$|E\,e^{\lambda X}| = \left| 1 + \sum_{k=2}^{\infty} \frac{\lambda^k\,E\,X^k}{k!} \right| \leq 1 + \Lambda(\lambda) \leq e^{\Lambda(\lambda)}$$

for all $|\lambda| < \frac{1}{\rho}$, where

$$\Lambda(\lambda) = \sum_{k=2}^{\infty} \frac{|\lambda|^k\,E\,|X|^k}{k!} \leq \frac{v|\lambda|^2}{1 - |\lambda|\,\rho}$$

Exc. 4.31: (*Bernstein's inequality*) Let $X_1, \ldots, X_n$ be independent random variables with $EX_j = 0$ for all $j = 1, \ldots, n$, and suppose that there exist positive numbers $0 < v_j < \infty$ and $0 < \rho < \infty$ satisfying

$$(*) \qquad E\,|X_j|^k \le v_j\,\rho^{k-2}\,k! \qquad \forall j = 1, \ldots, n\ \forall k = 2, 3, \ldots$$

Set $S = X_1 + \cdots + X_n$ and $v = v_1 + \cdots + v_n$. Show that the Laplace transform $L_S(\lambda)$ exists for all $\lambda \in \mathbf{C}$ with $|\Re\lambda| < \frac{1}{\rho}$ and that

(i) $$|L_S(\lambda)| \le \exp\left(\frac{v|\lambda|^2}{1 - |\lambda|\,\rho}\right) \qquad \forall |\lambda| < \tfrac{1}{\rho}$$

(ii) $$P(S \ge t) \le \exp\left(-\frac{(\gamma - 1)\,t^2}{\gamma(\gamma\,v + \rho\,t)}\right) \qquad \forall t > 0\ \forall \gamma > 1$$

(iii) If c is a positive number such that $|X_j| \le c$ for all $j = 1, \ldots, n$, then (*) holds with $\rho = \frac{c}{3}$ and $v_j = \frac{1}{2}\,\mathrm{var}(X_j)$, and

$$P(S \ge t) \le \exp\left(-\frac{(\gamma - 1)\,t^2}{\gamma\left(\frac{\gamma}{2}\,\sigma^2 + \frac{c}{3}\,t\right)}\right) \qquad \forall t > 0\ \forall \gamma > 1$$

where $\sigma^2 = \mathrm{var}(S)$

Remark: Note that the constant γ in inequalities (ii) and (iii) may be chosen freely subject to the condition $\gamma > 1$. Hence, the best choice of γ is the value which minimizes the right-hand side of (ii) or (iii); it is customary to choose $\gamma = 2$, but this is not, in general, the best possible choice.

Hint: (i): Apply Exc.4.30. (ii): Set $\lambda = t/(\gamma v + \rho t)$ and apply (ii) in Exc.4.16.

Exc. 4.32: If $x \in \mathbf{R}$ and $n \in \mathbf{N}_0$, then *the nth factorial*, denoted $x^{(n)}$, is defined by [see (4.20)]

$$x^{(0)} := 1 \quad , \quad x^{(n)} := x(x-1)\cdots(x-n+1) \quad \text{if} \quad n \ge 1$$

Show that there exist non-negative integers A_{ij} satisfying

(i) $$x^{(n)} = \sum_{j=0}^{n} (-1)^{n-j}\,A_{nj}\,x^j \quad , \quad x^n = \sum_{i=0}^{n} A_{in}\,x^{(i)}$$

(ii) $$A_{0j} = A_{j0} = 0 \quad \forall j \ge 1 \quad , \quad A_{ii} = 1 \quad \forall i \ge 0$$

(iii) $$A_{ij} = \begin{cases} A_{i-1,j-1} + (i-1)\,A_{i-1,j} & \text{if } 1 \le j \le i-1 \\ A_{i-1,j-1} + i\,A_{i,j-1} & \text{if } 1 \le i \le j-1 \end{cases}$$

Show that A_{ij} can be computed recursively from (ii) and (iii), and verify the following table for $\{A_{ij}\}$.

A_{ij}	j=0	j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8
i=0	1	0	0	0	0	0	0	0	0
i=1	0	1	1	1	1	1	1	1	1
i=2	0	1	1	3	7	15	31	63	127
i=3	0	2	3	1	6	25	90	301	966
i=4	0	6	11	6	1	10	65	350	1701
i=5	0	24	50	35	10	1	15	140	1050
i=6	0	120	274	225	85	15	1	21	266
i=7	0	720	1764	1624	735	175	21	1	28
i=8	0	5040	13068	13132	6769	1960	322	28	1

Suppose that X is random variable and that $n \in \mathbf{N}$ is an integer such that $E|X|^n < \infty$, then the jth moment, central moment and factorial moment exists for all $j = 0, 1, \ldots, n$, and are given by [see (4.20)]

$$\mu_X(j) = E\,X^j \quad , \quad \mu_X^0(j) = E(X-m)^j \quad , \quad \nu_X(j) = E\,X^{(j)}$$

where $m = E\,X$ is the mean of X . Show that

(iv) $$\mu_X(n) = \sum_{j=0}^{n} \binom{n}{j} m^{n-j}\,\mu_X^0(j) \quad , \quad \mu_X^0(n) = \sum_{j=0}^{n} \binom{n}{j} (-m)^{n-j}\,\mu_X(j)$$

(v) $\mu_X(n) = \sum_{j=0}^{n} A_{jn}\, \nu_X(j)$, $\nu_X(n) = \sum_{j=0}^{n} (-1)^{n-j} A_{nj}\, \mu_X(j)$

Exc. 4.33: Let $\{X_n\}$ be a sequence of random variables whose distributions are given by $P(X_n = a_n) = \frac{1}{n}$ and $P(X_n = 0) = 1 - \frac{1}{n}$ for all $n \geq 1$, where $\{a_n\}$ is a given sequence of non-negative numbers. Show that $\{X_n \mid n \geq 1\}$ is uniformly integrable if and only if $\frac{a_n}{n} \to 0$ as $n \to \infty$.

Exc. 4.34: Let $\{X_n\}$ be a uniformly integrable sequence of random variables and let $\{Y_n\}$ and $\{Z_n\}$ be two sequences of random variables. Let $f : [0, \infty[\to [0, \infty]$ be a Borel function and let $\alpha, \beta, \gamma, \delta, c > 0$ be positive finite constants satisfying $\int_\alpha^\infty f(s)\, ds < \infty$ and

$$P(|Y_n| > t) \leq f(t) + \beta\, P(|X_n| > \gamma t - \delta) \quad \forall t > c\, \forall n \geq 1$$

Show the following:

(i) $\{Y_n\}$ is uniformly integrable

(ii) If $X_n \sim Z_n$ for all $n \geq 1$, then $\{Z_n\}$ is uniformly integrable

(iii) If $P(|Z_n| > t) \leq f(t)$ for all $t > c$ and all $n \geq 1$, then $\{Z_n\}$ is uniformly integrable

(iv) If $|Z_n| \leq a|X_n| + b$ a.s. for all $n \geq 1$, then $\{Z_n\}$ is uniformly integrable

(v) If $\{Y_n\}$ and $\{Z_n\}$ are uniformly integrable, then so is $\{Y_n + Z_n\}$

Hint: (i): Apply (4.2.7) to show

$$\int_{\{|Y_n|>a\}} |Y_n|\, dP = aP(|Y_n| > a) + \int_a^\infty P(|Y_n| > t)\, dt$$

$$\leq af(a) + \int_a^\infty f(t)dt + \beta a P(|X_n| > \gamma a - \delta) + \beta \int_a^\infty P(|X_n| > \gamma t - \delta)\, dt$$

$$= af(a) + \int_a^\infty f(t)dt + \beta a P(|X_n| > \gamma a - \delta) + \tfrac{\beta}{\gamma} \int_{\gamma a - \delta}^\infty P(|X_n| > u)\, du$$

$$\leq af(a) + \int_a^\infty f(t)dt + \tfrac{2\beta}{\gamma} \int_{\{|X_n|>\gamma a - \delta\}} |X_n|\, dP$$

for all $a > c + \frac{2\delta}{\gamma}$. Show that $\liminf_{a\to\infty} af(a) = 0$ and apply (3.24.4).

Exc. 4.35: Let $\{X_n\}$ be a sequence of identically distributed random variables such that $E|X_1| < \infty$. Show that $\{X_n\}$ is uniformly integrable. [Apply (3.24.4).]

Exc. 4.36: Let $\{X_n\}$ be a sequence of random variables such that $E|X_n|^q \le C$ for all $n \ge 1$, where $q > 1$ and $0 \le C < \infty$. Show that $\{X_n\}$ is uniformly integrable. [Apply (3.24.5).]

Exc. 4.37: Let $\{X_n\}$ be a sequence of random variables with finite means and variances. Suppose that $|EX_n| \le C$ and $\mathrm{var}(X_n) \le C$ for all $n \ge 1$, where $0 \le C < \infty$. Show that $\{X_n\}$ is uniformly integrable.

Hint: Show that $E|X_n|^2 \le C + C^2$ and apply Exc.4.36

Exc. 4.38: Let $\{X_n\}$ be a sequence of random variables with finite means and variances. Suppose that $X_n \to X$ in pr and that $EX_n^2 < C$ for all $n \ge 1$, where $0 \le C < \infty$. Show that $\{X_n\}$ is uniformly integrable and that $X_n \to X$ in 1-mean. [Apply (3.25.6) and Exc.4.36.]

Exc. 4.39: Let $\{X_n\}$ be a uniformly integrable sequence of random variables, and let $Y_n = (X_1 + \cdots + X_n)/n$. Show that $\{Y_n\}$ is uniformly integrable.

Hint: Let

$$\nu(F) = \sup_{n\ge1}\left\{ \int_F |X_n|\, dP \right\} \quad , \quad \kappa(F) = \sup_{n\ge1}\left\{ \int_F |Y_n|\, dP \right\}$$

for all $F \in \mathcal{F}$. Show that $\kappa(F) \le \nu(F)$ for all $F \in \mathcal{F}$ and apply (3.22.34).

Exc. 4.40: Let $X_1, \ldots, X_n$ be independent random variables with $E|X_j|^4 < \infty$ and $E\,X_j = 0$ for all $j = 1, \ldots, n$. Let $S = X_1 + \cdots + X_n$ and show that

(i) $$E\,S^i = \sum_{j=1}^{n} E\,X_j^{\,i} \qquad \forall i = 1,2,3$$

(ii) $$E\,S^4 = \sum_{j=1}^{n} E\,X_j^{\,4} + 3\left\{ \sum_{j=1}^{n} E\,X_j^{\,2} \right\}^2 - 3\sum_{j=1}^{n} \left\{ E\,X_j^{\,2} \right\}^2$$

Exc. 4.41: Let $X_1, X_2, \ldots$ be independent identically distributed random variables such that $P(X_n = -1) = 1 - p$ and $P(X_n = 1) = p$, where $0 < p < 1$ is a given number. Let $S_n = X_1 + \cdots + X_n$ and find the following quantities:

(i) $$\mu_n = E\,S_n \ , \ \sigma_n^2 = \mathrm{var}(S_n) \ , \ \kappa_n = E(S_n - \mu_n)^3 \ , \ \tau = \lim_{n\to\infty} \frac{\sigma_n^3}{\kappa_n}$$

Exc. 4.42: Let X and Y be two Poisson-distributed random variables with parameters λ and μ, and suppose that X and Y are independent. Show that

(i) $$P(X-Y=k) = e^{-\lambda-\mu}\left(\tfrac{\lambda}{\mu}\right)^{k/2} I_{|k|}\left(2\sqrt{\mu\lambda}\right) \qquad \forall k = 0, \pm 1, \pm 2, \ldots$$

where I_j is *the jth modified Bessel function of the first kind*:

$$I_j(x) = \sum_{n=0}^{\infty} \frac{1}{n!\,(n+j)!}\left(\frac{x}{2}\right)^{2n+j} \qquad \forall x \in \mathbf{R}\ \forall j = 0, 1, 2, \ldots$$

Exc. 4.43: Let $v : \mathbf{N} \to \mathbf{N}$ be a rapidly increasing function. Then by definition (4.8.13) there exists and integer $m \geq 1$ and a number $c > 1$ such that $v(n+1) \geq c\,v(n)$ for all $n \geq m$. Let $c_\alpha := m + \frac{1}{c^\alpha - 1}$ and show that

(i) $$\sum_{j=k}^{\infty} v(j)^{-\alpha} \leq c_\alpha\, v(k)^{-\alpha} \qquad \forall \alpha > 0\ \forall k = 1, 2, \ldots$$

Hint: Note that $v(j) \leq c^{k-j}\, v(k)$ for all $j \geq k \geq m$, and so

$$\sum_{j=k}^{\infty} v(j)^{-\alpha} \leq \sum_{j=k}^{\infty} \left(c^{-\alpha}\right)^{j-k} v(k)^{-\alpha} = \frac{c^\alpha\, v(k)^{-\alpha}}{c^\alpha - 1}$$

for all $k \geq m$. Argue that this shows (i).

Exc. 4.44: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of random variables. Let $\mu_n = E\,X_n$, $\sigma_n^2 = \mathrm{var}(X_n)$ and $\rho_{ij} = \rho(X_i, X_j)$ be the means, variances and correlations coefficients, respectively. Suppose that there exists a function $R : \mathbf{N}_0 \to [0, \infty[$ satisfying

(a) $$R(0) = 1 \quad \text{and} \quad \rho_{ij} \leq R(|i-j|) \qquad \forall i, j \geq 1$$

(b) $$\lim_{n\to\infty} \left\{ n^{-2}\left(\sum_{j=0}^{n-1} R(j)\right) \cdot \left(\sum_{j=1}^{n} \sigma_i^2\right) \right\} = 0$$

Show that

(i) $$\frac{1}{n}\sum_{j=1}^{n} (X_j - \mu_j) \to 0 \quad \text{in 2-mean and in } pr$$

Hint: Let $\sigma_{ij} = \operatorname{cov}(X_i, X_j)$. Then $\sigma_{ij} = \rho_{ij}\,\sigma_i\,\sigma_j$. Show that (a) implies

$$0 \le \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \le \left\{ \sum_{i=1}^{n} \sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} \sigma_i\,\sigma_{k+i}\,R(k) \right\}$$

$$= \left\{ \sum_{i=1}^{n} \sigma_i^2 + 2 \sum_{k=1}^{n-1} R(k) \sum_{i=1}^{n-k} \sigma_i\,\sigma_{k+i} \right\}$$

$$\le \left\{ \sum_{i=1}^{n} \sigma_i^2 + 2 \sum_{k=1}^{n-1} R(k) \sqrt{\sum_{i=1}^{n-k} \sigma_i^2} \sqrt{\sum_{i=1}^{n-k} \sigma_{i+k}^2} \right\}$$

$$\le 2 \left\{ \sum_{i=1}^{n} \sigma_i^2 \right\} \left\{ \sum_{k=0}^{n-1} R(k) \right\}$$

Then show that (i) follows from (b), (4.8.11) and the inequality above.

Exc. 4.45: Let $1 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables. Suppose that there exists numbers $\mu_j \in \mathbf{R}$, $\tau_j \ge 0$, $0 < c < \infty$ and $0 \le r < q - 1$ satisfying

(a) $$E \left| \sum_{j=1}^{n} (X_j - \mu_j) \right|^q \le c n^r \sum_{j=1}^{n} \tau_j \qquad \forall n \ge 1$$

(b) $$\sum_{j=1}^{\infty} j^{r-q+1} \tau_j < \infty$$

Show that

(i) $$\frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j) \to 0 \qquad \text{a.s. and in } q\text{-mean}$$

Hint: Let $\hat{X}_n = \frac{1}{n} \sum_1^n (X_j - \mu_j)$ and show that (a) implies

$$\sum_{n=1}^{\infty} E|\hat{X}_n|^q \le \sum_{n=1}^{\infty} c n^{r-q} \sum_{j=1}^{n} \tau_j = c \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} n^{r-q} \tau_j$$

Then show that (i) follows from (4.8.16), (4.8.6) and (b).

Exc. 4.46: Let $0 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables. Let $\{\mu_j\} \subseteq \mathbf{R}$, and set $S_0 = 0$ and

$$S_n = \sum_{j=1}^{n} (X_j - \mu_j) \quad , \quad M_{mk} = \max_{m \le j < k} |S_j - S_{m-1}|$$

for all $n \geq 1$ and all $k > m \geq 1$. Let $\gamma = (q-1)^+$ and show that [apply Exc.4.14]

(i) $$E|S_n|^q \leq n^\gamma \sum_{j=1}^{n} E|X_j - \mu_j|^q \quad \forall n \geq 1$$

(ii) $$E|S_n - S_m|^q \leq (n-m)^\gamma \sum_{j=m+1}^{n} E|X_j - \mu_j|^q \quad \forall 0 \leq m < n$$

(iii) $$E|M_{mn}|^q \leq (n-m)^\gamma \sum_{j=m}^{n-1} E|X_j - \mu_j|^q \quad \forall 1 \leq m < n$$

(iv) $$E|M_{mn}|^q \leq \sum_{j=m}^{n-1} E|S_j - S_{m-1}|^q \quad \forall 1 \leq m < n$$

(v) $$E|M_{mn}|^q \leq (n-m) \max_{m \leq j < n} E|S_j - S_{m-1}|^q \quad \forall 1 \leq m < n$$

Exc. 4.47: Let $0 \leq \lambda \leq 1$ and $d > \lambda$. Set

$$\xi_\lambda(k) = \begin{cases} int\big(k^{1/\lambda}\big) & \text{if } \lambda > 0 \text{ and } k \in \mathbf{N} \\ 2^{k-1} & \text{if } \lambda = 0 \text{ and } k \in \mathbf{N} \end{cases}$$

Show that $\xi_\lambda(1) = 1$ and that $\xi_\lambda : \mathbf{N} \to \mathbf{N}$ is strictly increasing. Moreover, show that there exists constants $0 < C_{\lambda d} < \infty$ satisfying

(i) $$\sum_{k=m}^{\infty} \xi_\lambda(k)^{-d} \leq C_{\lambda d}\, \xi_\lambda(m)^{\lambda - d} \quad \forall m \geq 1$$

(ii) $$\xi_\lambda(m+1) \leq C_{\lambda d}\, \xi_\lambda(m) \quad \forall m \geq 1$$

(iii) $$\xi_\lambda(m+1) - \xi_\lambda(m) \leq C_{\lambda d}\, \xi_\lambda(m)^{1-\lambda} \quad \forall m \geq 1$$

Exc. 4.48: Let $0 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables. Suppose that there exists numbers $\mu_j \in \mathbf{R}$, $\tau_j \geq 0$ and $p, r, c \geq 0$ satisfying

(a) $$E|S_n|^q \leq c\, n^r \sum_{j=1}^{n} \tau_j \quad \forall n \geq 1$$

(b) $$E|M_{mn}|^q \leq c\,(n-m)^p \sum_{j=m}^{n-1} \tau_j \quad \forall 1 \leq m < n$$

(c) $\sum_{j=1}^{\infty} j^{-q+\alpha} \tau_j < \infty$

where S_n and M_{mn} are defined as in Exc.4.46 and where α is given by

$$\alpha = \begin{cases} \frac{p(1+r\wedge p)}{1+p} & \text{if } q \geq 1+p \\ \frac{(q-1)(1+r\wedge p)}{q} & \text{if } 1+r\wedge p \leq q \leq 1+p \\ q-1 & \text{if } 1 \leq q \leq 1+r\wedge p \\ 0 & \text{if } 0 < q \leq 1 \end{cases}$$

for all $n > m \geq 0$. Show that

(i) $\frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j) \to 0$ a.s. and in q-mean

Hint: Note that by (b), we have $E|X_j - \mu_j|^q \leq c\,\tau_j$. Conclude from Exc.4.46 that (a) holds with r replaced by $\rho = r \wedge p \wedge (q-1)^+$. Set $\beta = p \wedge (q-1)^+$ and $\lambda = \frac{\beta-\rho}{1+\beta}$; then $0 \leq \lambda \leq 1$ and $\lambda \leq \beta - \rho < q - \rho$. Observe that $\rho + \lambda = \beta(1-\lambda) = \alpha$, and apply (4.10.7)–(4.10.10) with r replaced by ρ .

Exc. 4.49: Let $0 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables. Suppose that there exists numbers $\mu_j \in \mathbf{R}$ satisfying

(a) $\sum_{j=1}^{\infty} j^{-\beta} E|X_j - \mu_j|^q < \infty$

where $\beta = q \wedge 1$. Show that

(i) $\frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j) \to 0$ a.s. and in q-mean

Hint: Note that by Exc.4.46, we have that (a) and (b) in Exc.4.48 holds with $\tau_j = E|X_j - \mu_j|^q$, $c = 1$ and $r = p = (q-1)^+$, and note that if α is defined as in Exc.4.48, then $\beta = q - \alpha$.

Exc. 4.50: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of random variables satisfying

(a) $\sum_{j=1}^{\infty} j^{-1} \operatorname{var}(X_j) < \infty$

Show that [apply Exc.4.49]

(i) $\frac{1}{n}\sum_{j=1}^{n}(X_j - E\,X_j) \to 0$ a.s. and in 2-mean

Exc. 4.51: Let $0 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables. Suppose that there exists numbers $\mu_j \in \mathbf{R}$, $\tau_j \geq 0$ and $r, c \geq 0$ satisfying

(a) $E\left|\sum_{j=m}^{n-1}(X_j - \mu_j)\right|^q \leq c(n-m)^r \sum_{j=m}^{n-1}\tau_j \quad \forall 1 \leq m < n$

(b) $\sum_{j=1}^{\infty} j^{-q+\alpha}\tau_j < \infty$

where α is given by

$$\alpha = \begin{cases} \frac{(1+r)^2}{2+r} & \text{if } q \geq 2+r \\ \frac{(q-1)(1+r)}{q} & \text{if } 1+r \leq q \leq 2+r \\ q-1 & \text{if } 1 \leq q \leq 1+r \\ 0 & \text{if } 0 < q \leq 1 \end{cases}$$

Show that

(i) $\frac{1}{n}\sum_{j=1}^{n}(X_j - \mu_j) \to 0$ a.s. and in q-mean

Hint: Apply Exc.4.46 to show that (b) in Exc.4.48 holds with $p = r+1$.

Exc. 4.52: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of random variables, and let $0 \leq c < \infty$ be a number satisfying

(a) $\sum_{j=1}^{\infty} j^{-3/2}\sigma_{jj} < \infty$

(b) $\sum_{j=m}^{n-1}\sum_{i=j+1}^{n}\sigma_{ij} \leq c\sum_{j=1}^{n}\sigma_{jj} \quad \forall n \geq m+1\ \forall m \geq 1$

where $\sigma_{ij} = \text{cov}(X_i, X_j)$ are the covariances. Show that

(i) $\frac{1}{n} \sum_{j=1}^{n} (X_j - E X_j) \to 0$ a.s. and in 2-mean

[Apply Exc.4.51.] Note that if $\sigma_{ij} \leq 0$ for all $i \neq j$, then (b) holds with $c = 0$.

Exc. 4.53: Let $0 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables. Suppose that there exists numbers $\mu_j \in \mathbf{R}$ and $r, c \geq 0$ satisfying:

(a) $E \left| \sum_{j=1}^{n} (X_j - \mu_j) \right|^q \leq c n^r \sum_{j=1}^{n} E|X_j - \mu_j|^q$

(b) $\sum_{j=1}^{\infty} j^{-q+\alpha} E|X_j - \mu_j|^q < \infty$

where α is given by:

$$\alpha = \begin{cases} \frac{(q-1)(1+r)}{q} & \text{if } q \geq 1 + r \\ q - 1 & \text{if } 1 \leq q \leq 1 + r \\ 0 & \text{if } 0 < q \leq 1 \end{cases}$$

Show that

(i) $\frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j) \to 0$ a.s. and in q-mean

Hint: Apply Exc.4.46 and Exc.4.48 with $\tau_j = E|X_j - \mu_j|^q$ and $p = (q-1)^+$.

Exc. 4.54: Let $0 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables. Suppose that there exists numbers $\mu_j \in \mathbf{R}$, $0 \leq r < q$ and $c \geq 0$ satisfying

(i) $E \left| \sum_{j=1}^{n} (X_j - \mu_j) \right|^q \leq c n^r \quad \forall n \geq 1$

(ii) $\sum_{j=1}^{\infty} j^{-\beta} E|X_j - \mu_j|^q < \infty$

where $\beta = q$ if $0 < q \leq 1$, and $\beta < 1 + (q-1)(q-r)$ if $q > 1$. Show that

(i) $\frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j) \to 0$ a.s. and in q-mean

Hint: Apply (4.10.7)–(4.10.10) with $p = (q-1)^+$, $\tau_1 = 1$, $\tau_j = 0$ for $j \geq 2$ and $\kappa_j = E|X_j - \mu_j|^q$ for $j \geq 1$.

Exc. 4.55: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of random variables with means $\mu_j = E\,X_j$ and covariances $\sigma_{ij} = \mathrm{cov}(X_i, X_j)$. Suppose that there exists numbers $c \geq 0$, $0 \leq r < 2$ and $0 < \beta < 3 - r$ satisfying

(a) $$\sum_{j=1}^{n} \sum_{i=1}^{n} \sigma_{ij} \leq c\,n^r \qquad \forall n \geq 1 \quad , \quad \sum_{j=1}^{\infty} j^{-\beta}\,\sigma_{jj} < \infty$$

Show that [apply Exc.4.54 with $q = 2$]

(i) $$\frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j) \to 0 \qquad \text{a.s. and in } q\text{-mean}$$

Exc. 4.56: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of random variables with means $\mu_j = E\,X_j$, variances $\sigma_j^2 = \mathrm{var}(X_j)$ and correlation coefficients $\rho_{ij} = \rho(X_i, X_j)$. Suppose that there exists numbers $c \geq 0$, $0 \leq r < 2$, $0 \leq \beta < 3 - r$ and a function $R : \mathbf{N} \to [0, \infty[$ satisfying

(a) $$R(0) = 1 \qquad \text{and} \qquad \rho_{ij} \leq R(|i-j|) \qquad \forall i, j \geq 1$$

(b) $$\Big\{ \sum_{j=1}^{n} \sigma_j^2 \Big\} \cdot \Big\{ \sum_{k=0}^{n-1} R(k) \Big\} \leq c\,n^r \qquad \forall n \geq 1$$

(c) $$\sum_{j=1}^{\infty} j^{-\beta}\,\sigma_j^2 < \infty$$

Show that [apply Exc.4.55, and the hint to Exc.4.44]

(i) $$\frac{1}{n} \sum_{j=1}^{n} (X_j - \mu_j) \to 0 \qquad \text{a.s. and in 2-mean}$$

Exc. 4.57: Let $X \in L^1(P)$ be a random variable. Show that

(i) $$E|X - E\,X|^p \leq 2^p\,E|X|^p \qquad \forall 1 \leq p < \infty$$

Hint: Apply (4.3.9) and the inequality: $|x - y|^p \leq 2^{p-1}\,(|x|^p + |y|^p)$.

Exc. 4.58: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of independent random variables with the same means $\mu = E\,X_n$ and the same variances $\sigma^2 = \mathrm{var}(X_n)$. Set

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j \quad , \quad \mathfrak{s}_n = \frac{1}{n} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$$

Show that

(i) $E\overline{X}_n = \mu \quad , \quad E\,\mathfrak{s}_n = \frac{n-1}{n}\,\sigma^2 \quad , \quad \mathrm{var}(\overline{X}_n) = \frac{1}{n}\,\sigma^2$

(ii) $\overline{X}_n \to \mu$ a.s. and in 2-mean

Moreover, show that if just one of the following two conditions holds:

(a) $X_1, X_2, \ldots$ are identically distributed

(b) $\exists\, q \geq 2 \;:\; \sum_{j=1}^{\infty} j^{-\beta}\, E\,|X_j|^q < \infty$ where $\beta = \frac{1}{4}\,(q + 4 \wedge q)$

then we have

(iii) $\mathfrak{s}_n \to \sigma^2$ a.s.

Hint: Show that (ii) follows from (4.11.2) and (4.11.7). Note that

$$\mathfrak{s}_n = \left(\frac{1}{n} \sum_{j=1}^{n} X_j^2 \right) - (\overline{X}_n)^2$$

and show that (iii) follows from (ii), Exc.4.57, (4.11) and (4.33).

Exc. 4.59: Let λ_n be the n-dimensional Lebesgue measure on $\mathbf{R}^n$. Consider the spherical shell

$$S_{n\delta} = \left\{ x \in \mathbf{R}^n \;\middle|\; \tfrac{n}{3} - n\delta < \|x\|^2 < \tfrac{n}{3} + n\delta \right\} \qquad \forall n \in \mathbf{N}\; \forall \delta > 0$$

where $\|\cdot\|$ denotes the usual euclidean norm in $\mathbf{R}^n$. Let $I_n = [0,1]^n$ be the unit cube in $\mathbf{R}^n$ and show that

(i) $\lim_{n\to\infty} \lambda_n(I_n \cap S_{n\delta}) = 1 \qquad \forall \delta > 0$

Hint: Apply the law of large numbers (4.12) on a suitably chosen sequence of independent, identically distributed random variables.

Exc. 4.60: Let $\{X_n\}$ be a sequence of independent, identically distributed random variables with distribution function $F(t) = P(X_n \leq t)$. Let

$$F_n(\omega, t) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{X_j \leq t\}}(\omega) \qquad \forall n \geq 1\; \forall \omega \in \Omega\; \forall t \in \mathbf{R}$$

be the empirical distribution functions. Show that for every fixed $t \in \mathbf{R}$ we have $F_n(t) \to F(t)$ a.s. [apply (4.12)].

Exc. 4.61: Let $\{X_n\}$ be a sequence of independent, identically distributed random variables with a geometric distribution $Ge(p)$, where $0 < p < 1$. Let $k \geq 1$ be a given integer, and set

$$U_n = X_n \wedge k \quad , \quad V_n = 1_{\{X_n < k\}} \quad , \quad \mathsf{U}_n = \sum_{j=1}^{n} U_j$$

$$\mathsf{V}_n = \sum_{j=1}^{n} V_j \quad , \quad \mathsf{W}_n = \frac{\mathsf{U}_n}{\mathsf{V}_n + \mathsf{U}_n}$$

Show that

(i) $\quad E\,U_1 = \dfrac{p\,(1-p^k)}{1-p} \quad , \quad E\,V_1 = 1 - p^k$

(ii) $\quad \mathsf{W}_n \to p$ a.s. and in q-mean $\forall\, 0 \leq q < \infty$

Exc. 4.62: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of independent random variables with means $\mu_n = E\,X_n$ and variances $\sigma_n^2 = \mathrm{var}(X_n)$. Let $s_n = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$ and let $a_n > 0$ be positive numbers such that $\lim\limits_{n\to\infty} \frac{s_n}{a_n} = 0$. Show that

(i) $\quad \dfrac{1}{a_n} \displaystyle\sum_{j=1}^{n} (X_j - \mu_j) \to 0$ in pr and in 2-mean

Exc. 4.63: Let $\{X_n\}$ be a sequence of independent, identically distributed random variables with an exponential distribution $E(\alpha)$ where $\alpha > 0$. Let

$$S_n = \sum_{j=1}^{n} X_j \quad , \quad U(B) = \#\{\, n \in \mathbf{N} \mid S_n \in B \,\} = \sum_{n=1}^{\infty} 1_B(S_n)$$

for all $n \geq 1$ and all $B \in \mathcal{B}(\mathbf{R}_+)$. Let $Q_n(B) = P(S_n \in B)$ be the distribution law of S_n . Show the following:

(i) $\quad E\,U(B) = \displaystyle\sum_{n=1}^{\infty} Q_n(B) = \alpha\,\lambda(B) \qquad \forall\, B \in \mathcal{B}(\mathbf{R}_+)$

(ii) If $B \subseteq \mathbf{R}_+$ is a Borel set with $\lambda(B) < \infty$, then $U(B) < \infty$ a.s.

where λ as usual denotes the Lebesgue measure on $\mathbf{R}$.

Hint: Recall that S_n has a gamma distribution $\Gamma(n, \frac{1}{\alpha})$, and so

$$Q_n(B) = \frac{\alpha^n}{(n-1)!} \int_B x^{n-1}\, e^{-\alpha x}\, dx \qquad \forall\, B \in \mathcal{B}(\mathbf{R}_+)$$

Exc. 4.64: Let $\{X_n\}$ be a sequence of independent random variables such that X_n only assumes the values $\sqrt{n}$, $-\sqrt{n}$ and 0 with the following probabilities:

$$P(X_n = \sqrt{n}) = P(X_n = -\sqrt{n}) = \tfrac{1}{2} n^{-\beta} \quad , \quad P(X_n = 0) = 1 - n^{-\beta}$$

for all $n \geq 1$ where $\beta > 0$ is a given number. Show that

(i) $\quad E X_n = 0 \quad , \quad E|X_n|^q = n^{(q-2\beta)/2} \quad \forall n \geq 1 \; \forall q > 0$

(ii) $\quad \frac{1}{n} \sum_{j=1}^{n} X_j \to 0$ a.s. and in q-mean for all $0 \leq q < \infty$

Exc. 4.65: Let $\{X_n\} \subseteq L^2(P)$ be a sequence of independent random variables with means $E X_n = \mu_n$ and variances $\operatorname{var}(X_n) = 1$ for all $n \geq 1$. Let $a_n \in \mathbf{R}$ be real numbers, and set

$$T_n = \tfrac{1}{n} \sum_{i=1}^{n} a_i (X_i - \mu_i)$$

Show that

(i) $\quad \sum_{j=1}^{\infty} j^{-2} a_j^2 < \infty \quad \Rightarrow \quad T_n \to 0$ a.s.

(ii) $\quad \lim_{n\to\infty} \left\{ n^{-2} \sum_{j=1}^{n} a_j^2 \right\} = 0 \quad \Rightarrow \quad T_n \to 0$ in 2-mean

Exc. 4.66: Let $\{X_n\} \subseteq L^4(P)$ be a sequence of independent random variables with means $E X_n = 0$ and variances $\operatorname{var}(X_n) = \sigma^2$ for all $n \geq 1$. Suppose that there exists a constant $0 < c < \infty$ such that $E|X_n|^4 \leq c\sqrt{n}$ for all $n \geq 1$. Show that

(i) $\quad \frac{1}{n} \sum_{j=1}^{n} X_j \to 0$ a.s. and in 4-mean

(ii) $\quad \frac{1}{n} \sum_{j=1}^{n} X_j^2 \to \sigma^2$ a.s. and in 2-mean

Exc. 4.67: Let $\{X_n\}$ be a sequence of independent random variables such that X_n is a uniformly distributed on the interval $[-n^\beta, n^\beta]$ for all $n \geq 1$ where $\beta > 0$ is a given positive number. Show that

(i) $\quad E X_n = 0 \quad , \quad E|X_n|^q = \dfrac{n^{\beta q}}{1+q} \quad \forall n \geq 1 \; \forall q > 0$

(ii) $\quad 0 < \beta < \frac{1}{2} \;\Rightarrow\; \frac{1}{n} \sum_{j=1}^{n} X_j \to 0$ a.s. and in q-mean $\quad \forall 0 \leq q < \infty$

Exc. 4.68: Let $\{X_n\}$ be independent, identically distributed random variables with density function given by

$$f(x) = \begin{cases} \frac{a}{2}\,|x|^{a-1} & \text{if } |x| \le 1 \\ 0 & \text{if } |x| > 1 \end{cases}$$

where $a > 0$ is a given positive number. Let $b_n > 0$ be positive numbers. Show that

(i) $\quad E\,X_n = 0 \quad , \quad E|X_n|^q = \dfrac{a}{a+q} \quad \forall n \ge 1\ \forall q > 0$

(ii) $\quad \sum\limits_{n=1}^{\infty} n^{-2}\, b_n^2 < \infty \quad \Rightarrow \quad \frac{1}{n} \sum\limits_{j=1}^{n} b_j\, X_j \to 0$ a.s. and in 2-mean

Exc. 4.69: Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of real random variables such that $\sum_1^\infty Y_n$ is summable a.s. Show that

(i) $\quad \sum\limits_{n=1}^{\infty} P(X_n \ne Y_n) < \infty \quad \Rightarrow \quad \sum\limits_{n=1}^{\infty} X_n$ is summable a.s.

(ii) If there exist $\delta > 0$ such that $|Y_n(\omega)| \ge \delta$ whenever $Y_n(\omega) \ne X_n(\omega)$, then $\sum_1^\infty X_n$ is summable a.s.

Exc. 4.70: Let $0 < q \le 1$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of random variables such that $\sum_{n=1}^\infty E|X_n|^q < \infty$. Show that $\sum_1^\infty X_n$ is summable a.s. and in q-mean [cf. (4.36)].

Hint: Apply Beppo Levi's theorem (3.7) to show that $\sum_1 \infty |X_n|^q < \infty$ a.s. Recall that $\ell^q \subseteq \ell^1$ if $0 < q \le 1$ [see Exc.3.21]..

Exc. 4.71: *(The three-series theorem)* Let $\{X_n\}$ be a sequence of real random variables and let $\{Y_n\} \subseteq L^2(P)$ be a sequence of independent random variables such that the following three infinite series converge and are finite:

(*) $\quad \sum\limits_{n=1}^{\infty} \mathrm{var}(Y_n) \quad , \quad \sum\limits_{n=1}^{\infty} E\,Y_n \quad , \quad \sum\limits_{n=1}^{\infty} P(X_n \ne Y_n)$

Show that $\sum_1^\infty X_n$ is summable a.s.

Remark: In applications of the three-series theorem one usually assumes that $X_1, X_2, \ldots$ are independent and that Y_n is a *truncation* of the X_n, that is,

$$Y_n = \begin{cases} X_n & \text{if } a_n < X_n < b_n \\ 0 & \text{otherwise} \end{cases}$$

where $a_n, b_n \in \mathbf{R}$ are given numbers with $a_n < b_n$. But other choices may be relevant. In Exc.4.75 you will a partial converse of the three-series theorem.

Hint: Using (4.36) show that the convergence of the first two series implies that $\sum_1^\infty Y_n$ is summable a.s. Then using Exc.4.69 show that the convergence of the last series implies that $\sum_1^\infty X_n$ is summable a.s.

Exc. 4.72: *(The two-series theorem)* Let $\{X_n\}$ be a sequence of real random variables and let $\{Y_n\} \subseteq L^2(P)$ be a sequence of independent, real, random variables satisfying

$$(*) \qquad |Y_n(\omega)| \geq \delta \qquad \forall \omega \in \{X_n \neq Y_n\} \; \forall n \geq 1$$

for some $\delta > 0$. Suppose that the following two infinite series converge and are finite:

$$(**) \qquad \sum_{n=1}^{\infty} \operatorname{var}(Y_n) \quad , \quad \sum_{n=1}^{\infty} E\,Y_n$$

Show that $\sum_1^\infty X_n$ is summable a.s. [See the hint to Exc.4.71.]

Remark: In applications of the two-series theorem one usually assumes that $X_1, X_2, \ldots$ are independent and that Y_n is a *modular truncation* of the X_n , that is,

$$Y_n = \begin{cases} a_n & \text{if } X_n \leq a_n \\ X_n & \text{if } a_n < X_n < b_n \\ b_n & \text{if } X_n \geq b_n \end{cases}$$

where $a_n, b_n \in \mathbf{R}$ are given numbers with $a_n < b_n$. But other choices may be relevant. Note that if $\limsup a_n < 0 < \liminf b_n$ and Y_n is the modular truncation defined above, then (*) holds. In Exc.4.75 you will a partial converse of the two-series theorem.

Exc. 4.73: Let $\{X_n\}$ be a sequence of independent, real, random variables and let

$$S_n = \sum_{j=1}^{n} X_j \quad , \quad M = \sup_{n \geq 1} |S_n| \quad , \quad L = \sup_{j \geq 1} |X_j|$$

Suppose that there exists $a > 0$ such that $P(|S_n| > a) \leq \frac{1}{8}$ for all $n \geq 1$. Show that [apply (iv) and (v) in Exc.2.58 noting that $R(a) \leq \frac{1}{8}$]

(i) $\quad E\,M^q \leq 2^{q+1} \left((2a)^q + \liminf_{n \to \infty} E|S_n|^q \right) \qquad \forall\, 0 < q < \infty$

(ii) $\quad E\,M^q \leq 2^{q+1} \left((4a)^q + E\,L^q \right) \qquad \forall\, 0 < q < \infty$

Exc. 4.74: Let $0 < q < \infty$ and let $\{X_n\} \subseteq L^q(P)$ be a sequence of independent random variables. Let

$$S_n = \sum_{j=1}^{n} X_j \quad , \quad M = \sup_{n \geq 1} |S_n| \quad , \quad L = \sup_{j \geq 1} |X_j|$$

Suppose that $S_n \to S$ in pr for some random variable S and that

(a) either $E\,L^q < \infty$ or $\liminf_{n\to\infty} E|S_n|^q < \infty$

Show that [apply Exc.4.73]

(i) $S, M \in L^q(P)$ and $S_n \to S$ a.s. and in q-mean

(ii) $E\,S = \sum_{n=1}^{\infty} E\,X_j \neq \pm\infty$ if $1 \leq q < \infty$

(iii) $\mathrm{var}(S) = \sum_{j=1}^{\infty} \mathrm{var}(X_j) < \infty$ if $2 \leq q < \infty$

Exc. 4.75: *(The converse two-three-series theorem)* Let $\{(X_n, U_n)\}$ be a sequence of independent 2-dimensional random vectors such that $\sum_1^\infty X_n$ is summable a.s. and

(*) $E\Big\{ \sup_{j\geq 1} |U_j|^2 \Big\} < \infty$ and $U_j(\omega) = X_j(\omega) \quad \forall \omega \in \{\, |X_j| < \delta \,\}\ \forall j \geq 1$

for some $\delta > 0$. Show that the following three infinite series converge and are finite:

(i) $\sum_{j=1}^{\infty} \mathrm{var}(U_j) \quad , \quad \sum_{j=1}^{\infty} E\,U_j \quad , \quad \sum_{j=1}^{\infty} P(X_j \neq U_j)$

Suppose that $\{X_n\}$ is given sequence of independent random variables such that $\sum_1^\infty X_n$ is summable a.s. Show that either of the two sequences $\{Y_j\}$ and $\{Z_j\}$ given by

$$Y_j = \begin{cases} X_j & \text{if } |X_j| \leq 1 \\ 0 & \text{if } |X_j| > 1 \end{cases} \quad , \quad Z_j = \begin{cases} -1 & \text{if } X_j < -1 \\ X_j & \text{if } |X_j| \leq 1 \\ 1 & \text{if } X_j > 1 \end{cases}$$

satisfies (*). Show that the following six infinite sums converge and are finite

$$\sum_{j=1}^{\infty} \mathrm{var}(Y_j) \quad , \quad \sum_{j=1}^{\infty} E\,Y_j \quad , \quad \sum_{j=1}^{\infty} P(X_j \neq Y_j)$$

$$\sum_{j=1}^{\infty} \mathrm{var}(Z_j) \quad , \quad \sum_{j=1}^{\infty} E\,Z_j \quad , \quad \sum_{j=1}^{\infty} P(X_j \neq Z_j)$$

Hint: Let $F_n = \{U_n \neq X_n\}$. Show that $F_1, F_2, \ldots$ are independent and $\limsup F_n \subseteq \limsup\{|X_n| \geq \delta\}$. Then apply the second Borel-Cantelli lemma (2.11.2) to conclude that the third sum in (i) is finite. Apply Exc.4.69 to conclude that $\sum_1^\infty U_n$ is summable a.s., and use Exc.4.74 to show that the two first sums in (i) are convergent and finite.

Exc. 4.76: Let X be a random variable with a continuous distribution function $F(x)$. Show that $Y = F(X)$ is uniformly distributed over $[0,1]$; that is, show that $P(Y \leq t) = t$ for all $0 < t < 1$ [apply (3.29.4) with $g = 1_{]-\infty,t]}$].

Exc. 4.77: Let $X = (X_1, \ldots, X_k)$ and $Y = (Y_1, \ldots, Y_k)$ be independent random vectors. Prove (4.31.2).

Exc. 4.78: Let $X_1, X_2, \ldots$ be a sequence of independent random variables such that $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$ where $0 < p_n < 1$ for all $n \geq 1$. Let $\{a_n\}$ be a sequence of real numbers and show that

(i) $\quad EX_n = p_n \quad , \quad \mathrm{var}(X_n) = p_n(1-p_n) \qquad \forall n \geq 1$

(ii) $\quad \sum_{j=1}^{\infty} j^{-2} a_j^2 < \infty \quad \Rightarrow \quad \frac{1}{n} \sum_{j=1}^{n} a_j(X_j - p_j) \to 0$ a.s.

(iii) $\quad \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} p_j = q \quad \Rightarrow \quad \frac{1}{n} \sum_{j=1}^{n} X_j \to q$ a.s.

Exc. 4.79: *(Khinchine's inequality)* Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent Bernoulli random variables [see (4.29)] and let

$$\rho_\alpha(x_1, \ldots, x_n) = E \left\| \sum_{j=1}^{n} \varepsilon_j x_j \right\|^\alpha \qquad \forall x_1, \cdots, x_n \in \mathbf{R}^k \; \forall \alpha > 0$$

where $\|\cdot\|$ as usual denotes the euclidean norm in $\mathbf{R}^k$. Let $\alpha > 0$ and $x_1, y_1, \ldots, x_n, y_n \in \mathbf{R}^k$ be given. Show the following:

(i) $\quad \rho_\alpha(x_1, \ldots, x_n) \leq \rho_\beta(x_1, \ldots, x_n)^{\alpha/\beta} \qquad \forall \beta \geq \alpha$

(ii) $\quad \rho_2(x_1, \ldots, x_n) = \sum_{j=1}^{n} \|x_j\|^2$

(iii) $\quad \rho_\alpha(x_1 - y_1, \ldots, x_n - y_n) \leq 2^{(\alpha-1)^+}(\rho_\alpha(x_1, \ldots, x_n) + \rho_\alpha(y_1, \ldots, y_n))$

(iv) $\quad \rho_\alpha(x_1 \ldots, x_n) \leq C_\alpha \left\{ \sum_{j=1}^{n} \|x_j\|^2 \right\}^{\alpha/2}$

where C_α is given by (4.30.8). Let X_j , U_j and V_j be k-dimensional random vectors for $j = 1, \ldots, n$ such that

(a) $(U_1, \ldots, U_n)$ and $(\varepsilon_1, \ldots, \varepsilon_n)$ are independent

(b) $V_1, \ldots, V_n$ are *sign-symmetric*, that is, $(V_1, \ldots, V_n) \sim (\pm V_1, \ldots, \pm V_n)$ for all choices of signs $\pm$ [cf. (4.31)]

(c) $X_1, \ldots, X_n$ are independent and integrable with mean vectors $\mu_1, \ldots, \mu_n$

Show that:

(v) $E \left\| \sum_{j=1}^{n} \varepsilon_j U_j \right\|^\alpha = E\rho_\alpha(U_1, \ldots, U_n)$

(vi) $E \left\| \sum_{j=1}^{n} V_j \right\|^\alpha = E\rho_\alpha(V_1, \ldots, V_n)$

(vii) $E \left\| \sum_{j=1}^{n} (X_j - \mu_j) \right\|^\alpha \leq C_\alpha \, n^{(\alpha-2)^+/2} \sum_{j=1}^{n} E\|X_j - \mu_j\|^\alpha$

Remark: Note that (vii) is an extension of Khinchine's inequality (4.32.1) to random vectors.

Hint: (i): Apply (4.3.9). (iii): Apply the inequality $(u+v)^\alpha \leq 2^\lambda(u^\alpha + v^\alpha)$ where $\lambda = (\alpha - 1)^+$. (iv): If $0 < \alpha \leq 2$, then (iv) follows from (i) and (ii). Therefore, suppose that $\alpha > 2$ and let

$$W_i = \left| \sum_{j=1}^{n} \varepsilon_j \, x_{ij} \right|^2 \qquad \forall i = 1, \ldots, k$$

where $x_{1j}, \ldots, x_{kj}$ are the coordinates of x_j . Then apply Minkowski's inequality (3.12.1) with $p = \frac{\alpha}{2}$ to prove

$$\rho_\alpha(x_1, \ldots, x_n)^{1/p} = \left\| \sum_{i=1}^{k} W_i \right\|_p \leq \sum_{i=1}^{k} \|W_i\|_p = \sum_{i=1}^{k} \left(E|\, W_i|^p \right)^{1/p}$$

$$= \sum_{\gamma=1}^{k} \left(E \left| \sum_{j=1}^{n} \varepsilon_j \, x_{\gamma j} \right|^\alpha \right)^{2/\alpha}$$

and show that (iv) follows from (4.30.7). (v): Apply (4.5.2). (vi): Easy consequence of (v). (vii): Mimic the proof of Khinchine's inequality (4.32).

Chapter 5

CONVERGENCE IN LAW

5.1 Convergence in Law on Metric Spaces

In (4.22) I told you that Abraham de Moivre in 1733 proved that

$$\lim_{n\to\infty} \sum_{j\in\Gamma_n(s)} \binom{n}{j} 2^{-n} = \frac{2}{B} \sum_{k=0}^{\infty} (-1)^k \frac{2^k s^{2k-1}}{k!(2k+1)}$$

where $\Gamma_n(s) = \{j \mid \frac{n}{2} \le j \le \frac{n}{2} + s\sqrt{n}\}$, and B is a constant, which James Stirling, who was a close friend of de Moivre, found to be equal to $\sqrt{2\pi}$. Let $U \sim N(0, \frac{1}{4})$. Then expanding e^{-2x^2} in its Taylor series and integrating the terms, we find

$$P(0 \le U \le s) = \sqrt{\frac{2}{\pi}} \int_0^s e^{-2x^2}\, dx = \frac{2}{B} \sum_{k=0}^{\infty} (-1)^k \frac{2^k s^{2k+1}}{k!(2k+1)}$$

Moreover, if $X_1, X_2, \ldots$ are independent random variables taking the values 0 and 1 with probability $\frac{1}{2}$, then

$$P(0 \le U_n \le s) = \sum_{j\in\Gamma_n(s)} \binom{n}{j} 2^{-n}$$

where $U_n = (X_1 + \cdots + X_n - \frac{n}{2})/\sqrt{n}$. Thus, de Moivre's result gives

$$\lim_{n\to\infty} P(U_n \le s) = P(U \le s) \qquad \forall s \in \mathbf{R}$$

Hence, de Moivre established the first version of the so-called "central limit theorem" which, loosely speaking, states that the sum of "many small" independent random variables is approximately normally distributed [see (5.20) for the precise statement]. Abraham de Moivre did not find the normal distribution, and he did not view his result in the above manner. But he did realize the enormous importance of his work and he writes:

> *And thus, in all cases it will be found that altho' chance produces irregularities, still the odds will be infinitely great, that in process of time, those irregularities will bear no proportion to the recurrency of that Order which naturally results from the ORIGINAL DESIGN*

Behind the result of de Moivre hides the most important convergence notion of probability, namely, convergence in law of random variables. This chapter is devoted to a careful study of this notion. Note that Moivre's result tells us that the distribution function of U_n tends to the distribution function of U. However, it does **not** say anything about convergence of the values $U_1(\omega), U_2(\omega), \ldots$, but only something about convergence of the probabilities $P(U_1 \in A), P(U_2 \in A), \ldots$. We shall now formalize the underlying convergence notion.

Let (S, d) be a metric space and let $C(S)$ be the set of *all bounded continuous functions* $f : S \to \mathbf{R}$. Let μ and μ_n be finite Borel measures on S. Then we say that $\{\mu_n\}$ *converges in law* to μ and we write $\mu_n \xrightarrow{\sim} \mu$ if

$$\int_S f\, d\mu = \lim_{n\to\infty} \int_S f\, d\mu_n \qquad \forall f \in C(S) \tag{5.1.1}$$

Suppose that X and X_n are S-valued random functions and that μ is a Borel probability measure on S. Then we say that $\{X_n\}$ *converges in law to* μ and we write $X_n \xrightarrow{\sim} \mu$ if $P_{X_n} \xrightarrow{\sim} \mu$. Similarly, we say that $\{X_n\}$ *converges in law to* X and we write $X_n \xrightarrow{\sim} X$ if $P_{X_n} \xrightarrow{\sim} P_X$. Hence, by (5.1.1) and the first transformation theorem (3.15) we have

$$X_n \xrightarrow{\sim} \mu \iff \int_S f\, d\mu = \lim_{n\to\infty} Ef(X_n) \qquad \forall f \in C(S) \tag{5.1.2}$$

$$X_n \xrightarrow{\sim} X \iff Ef(X) = \lim_{n\to\infty} Ef(X_n) \qquad \forall f \in C(S) \tag{5.1.3}$$

Clearly we may replace $C(S)$ in (5.1.1)–(5.1.3) by the set of all bounded continuous complex-valued functions.

In analogy with definition (3.22.6) we say that $\{X_n\}$ *converges in probability* to X and we write $X_n \to X$ in *pr* if $\lim_{n\to\infty} P^*(d(X, X_n) > \varepsilon) = 0$ for

all $\varepsilon > 0$ where P^* is the outer P-measure [see (3.4)]. Note that if $d(X, X_n)$ is measurable for all $n \geq 1$, then we have

$$(5.1.4) \qquad X_n \to X \text{ in } pr \iff \lim_{n\to\infty} P(d(X_n, X) > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

By Exc.1.73 we have that $d(X, X_n)$ is measurable, whenever S is a separable metric space. However, in general $d(X, X_n)$ need not be measurable.

Suppose that $\{\mu_n\}$ is a sequence of finite Borel measures on S such that $\mu_n \overset{\sim}{\to} \mu$ and $\mu_n \overset{\sim}{\to} \nu$. Then

$$\int_S f\, d\mu = \lim_{n\to\infty} \int_S f\, d\mu_n = \int_S f\, d\nu \qquad \forall f \in C(S)$$

Hence, by the uniqueness theorem (3.40) we find that μ and ν coincide on $\sigma(C(S))$ and by (P.5.20+21), we know that every closed set belongs to $\sigma(C(S))$. Thus, $\sigma(C(S)) = \mathcal{B}(S)$ and so we have proved that

(5.1.5) If $\mu_n \overset{\sim}{\to} \mu$ and $\mu_n \overset{\sim}{\to} \nu$, then $\mu = \nu$

(5.1.6) If $X_n \overset{\sim}{\to} \mu$ and $X_n \overset{\sim}{\to} \nu$, then $\mu = \nu$

(5.1.7) If $X_n \overset{\sim}{\to} X$ and $X_n \overset{\sim}{\to} Y$, then $X \sim Y$

whenever μ_n, μ and ν are finite Borel measures on S, and X_n, X and Y are S-valued random functions. Note that if $X_n \overset{\sim}{\to} X$ and $X \sim Y$, then $X_n \overset{\sim}{\to} Y$. Hence, the limit random function X is far from being unique, but the distribution law of X is unique by (5.1.7).

In our study of convergence in law we need various function spaces, that is, sets of functions from S into $\overline{\mathbf{R}}$ or $\mathbf{C}$. Let $\mathcal{M}(S)$ denote the sets of *all measurable* from S into $\mathbf{C}$ or $\overline{\mathbf{R}}$, and let $M(S)$ denote the set of all *measurable bounded* functions from S into $\mathbf{C}$. If μ is a Borel measure on S, we let $\mathrm{Lsc}(\mu)$ denote the sets of all functions from S into $\overline{\mathbf{R}}$ which are *lower semicontinuous* at μ-a.a. points in S, we let $\mathrm{Usc}(\mu)$ denote the set of all functions which are *upper semicontinuous* at μ-a.a. points in S [see (P.5.23)–(P.5.26)], and we let $C(\mu)$ denote the set of all functions from S into $\mathbf{C}$ which are *continuous* at μ-a.a. points in S. Moreover, for any given function g from S into $\mathbf{C}$, we introduce the two function spaces $o(g)$ and $O(g)$ as follows:

$$o(g) = \{f : S \to \mathbf{C} \mid \forall \varepsilon > 0\ \exists b \in \mathbf{R}_+ \ :\ |f(s)| \leq b + \varepsilon |g(s)| \quad \forall s\}$$
$$O(g) = \{f : S \to \mathbf{C} \mid \exists b \in \mathbf{R}_+ \ :\ |f(s)| \leq b(1 + |g(s)|) \quad \forall s\}$$

If $f \in o(g)$, we say that f is *small o of g* ; and if $f \in O(g)$, we say that f is *big O of g* . Note that if $S = \mathbf{R}^k$ and f is continuous [or more generally, if f is bounded on all bounded subsets of $\mathbf{R}^k$], then we have

$$\limsup_{\|x\| \to \infty} \frac{|f(x)|}{1 + |g(x)|} = 0 \quad \Rightarrow \quad f \in o(g) \tag{5.1.8}$$

$$\limsup_{\|x\| \to \infty} \frac{|f(x)|}{1 + |g(x)|} < \infty \quad \Leftrightarrow \quad f \in O(g) \tag{5.1.9}$$

Let $\{X_n\}$ be a given sequence of S-valued random functions. Then we define

$$\mathcal{M}_u(X_n) := \left\{ f \in \mathcal{M}(S) \;\middle|\; \{f(X_n) | n \geq 1\} \text{ is uniformly integrable} \right\} \tag{5.1.10}$$

$$\mathcal{M}_b(X_n) := \left\{ f \in \mathcal{M}(S) \;\middle|\; \sup_{n \geq 1} E|f(X_n)| < \infty \right\} \tag{5.1.11}$$

Then. clearly, we have that

$$M(S) \subseteq \mathcal{M}_u(X_n) \subseteq \mathcal{M}_b(X_n) \subseteq \mathcal{M}(S) \tag{5.1.12}$$

Let $f, g \in \mathcal{M}(S)$ and suppose that $f \in O(g)$. Then there exists $b \in \mathbf{R}_+$ such that $|f| \leq b(1 + |g|)$. Hence, we have

$$\sup_n \left\{ \int_F |f(X_n)| \, dP \right\} \leq bP(F) + b \sup_n \int_F |g(X_n)| \, dP \qquad \forall F \in \mathcal{F}$$

Similarly, if $f \in o(g)$ and $\varepsilon > 0$ is given, then there exists $b \in \mathbf{R}_+$ such that $|f| \leq b + \varepsilon |g|$. Thus, we have

$$\sup_n \left\{ \int_F |f(X_n)| \, dP \right\} \leq bP(F) + \varepsilon \sup_n E|g(X_n)| \qquad \forall F \in \mathcal{F}$$

Hence, we find

$$\mathcal{M}(S) \cap O(g) \subseteq \mathcal{M}_u(X_n) \quad \text{if} \quad g \in \mathcal{M}_u(X_n) \tag{5.1.13}$$

$$\mathcal{M}(S) \cap o(g) \subseteq \mathcal{M}_u(X_n) \quad \text{if} \quad g \in \mathcal{M}_b(X_n) \tag{5.1.14}$$

$$\mathcal{M}(S) \cap O(g) \subseteq \mathcal{M}_b(X_n) \quad \text{if} \quad g \in \mathcal{M}_b(X_n) \tag{5.1.15}$$

Finally, we let $\mathrm{Lip}(S)$ denote the set of all functions $f : S \to \mathbf{R}$ satisfying a *Lipschitz' condition*, that is,

$$|f(s) - f(t)| \leq Cd(s,t) \qquad \forall s, t \in S \tag{5.1.16}$$

for some $C \in \mathbf{R}_+$. We let $\mathrm{Lip}_1^+(S)$ denote the set of all functions f satisfying (5.1.16) with $C = 1$ and such that $0 \le f(s) \le 1$ for all $s \in S$.

I have chosen to formulate the results in this chapter for random functions and Borel probability measures. However, most of the results below will hold for finite Borel measures.

5.2 The Portmanteau Lemma

Let (S, d) be a metric space and let $\{X_n\}$ be a sequence of S-valued random functions such that $X_n \xrightarrow{\sim} \mu$ where μ is a Borel probability measure on S . Let $f : S \to \overline{\mathbf{R}}$ and $h : S \to \mathbf{C}$ be two Borel functions. Then we have

$$f \in \mathrm{Lsc}(\mu) \text{ and } f^- \in \mathcal{M}_u(X_n) \;\Rightarrow\; \int_S f\,d\mu \le \liminf_{n\to\infty} Ef(X_n) \tag{5.2.1}$$

$$f \in \mathrm{Usc}(\mu) \text{ and } f^+ \in \mathcal{M}_u(X_n) \;\Rightarrow\; \int_S f\,d\mu \ge \limsup_{n\to\infty} Ef(X_n) \tag{5.2.2}$$

$$h \in \mathcal{C}(\mu) \cap \mathcal{M}_u(X_n) \;\Rightarrow\; \int_S h\,d\mu = \lim_{n\to\infty} Eh(X_n) \tag{5.2.3}$$

where $\mathcal{M}_u(X_n)$ is given by (5.1.10). Moreover, if $B \in \mathcal{B}(S)$, then we have

$$\mu(\mathrm{int}(B)) \le \liminf_{n\to\infty} P(X_n \in B) \le \limsup_{n\to\infty} P(X_n \in B) \le \mu(\mathrm{cl}(B)) \tag{5.2.4}$$

$$\mu(\mathrm{bd}(B)) = 0 \;\Rightarrow\; \mu(B) = \lim_{n\to\infty} P(X_n \in B) \tag{5.2.5}$$

where $\mathrm{int}(B)$ is the interior of B , $\mathrm{cl}(B)$ is the closure of B and $\mathrm{bd}(B) = \mathrm{cl}(B) \setminus \mathrm{int}(B)$ is the boundary of B . Finally, if $f \in \mathcal{C}(\mu)$ satisfies

$$E|f(X_n)| < \infty \quad \forall n \quad \text{and} \quad \int_S |f|\,d\mu = \lim_{n\to\infty} E|f(X_n)| < \infty \tag{5.2.6}$$

then $f \in \mathcal{M}_u(X_n)$.

Remarks: Portmanteau is **not** the name of a mathematician, but it means a large suitcase. The result has got its name because of its importance, and it is a result you should always carry in your suitcase when you are travelling in the realm of convergence in law.

Proof: (5.2.1): Let $\varepsilon > 0$ be given. Since $\{f^-(X_n)\}$ is uniformly integrable, then by (3.24.4) there exists $q > 0$ so that

$$\int_{\{X_n \in F\}} f^-(X_n)\, dP < \varepsilon \qquad \forall n \geq 1$$

where $F = \{s \mid f^-(s) > q\} = \{s \mid f(s) < -q\}$. Let $g = f \vee (-q)$. Then $g \geq -q$ and since $f \in \mathrm{Lsc}(\mu)$, there exists a μ-null set $N \in \mathcal{B}(S)$ such that g is lower semicontinuous at s for all $s \in S \setminus N$. Let g_k be defined by (P.5.31). Then $g_k \in C(S)$ and $g_k(s) \leq g(s)$ for all $s \in S$ by (P.5.32)–(P.5.34). Hence, we have

$$\int_S g_k\, d\mu = \lim_{n\to\infty} Eg_k(X_n) \leq \liminf_{n\to\infty} Eg(X_n) \qquad \forall k \geq 1$$

and since $g_k(s) \uparrow g(s)$ for all $s \in S \setminus N$ by (P.5.35), then by the monotone convergence theorem (3.4) we get

$$\int_S g\, d\mu = \lim_{k\to\infty} \int_S g_k\, d\mu \leq \liminf_{n\to\infty} Eg(X_n)$$

Note that $f \leq g \leq f + 1_F f^-$ and so we find

$$\begin{aligned}\int_S f\, d\mu &\leq \int_S g\, d\mu \leq \liminf_{n\to\infty} Eg(X_n)\\ &\leq \liminf_{n\to\infty} \left\{ Ef(X_n) + \int_{\{X_n \in F\}} f^-(X_n)\, dP \right\}\\ &\leq \varepsilon + \liminf_{n\to\infty} Ef(X_n)\end{aligned}$$

Letting $\varepsilon \downarrow 0$, we see that (5.2.1) holds.

(5.2.2): Follows from (5.2.1) applied to $(-f)$.

(5.2.3): Let $f_1 = \Re h$ and $f_2 = \Im h$ be the real and imaginary parts of h . Then f_1 and f_2 are continuous μ-a.s. and $f_1, f_2 \in \mathcal{M}_u(X_n)$. Hence, by (5.2.1) we have

$$\int_S |f_j|\, d\mu \leq \liminf_{n\to\infty} E|f_j(X_n)| < \infty \quad \text{for } j = 1, 2$$

Thus, $h = f_1 + if_2 \in L^1(\mu, \mathbf{C})$ and so (5.2.3) follows from (5.2.1+2).

(5.2.4+5): Let $G = \mathrm{int}\,(B)$ and $F = \mathrm{cl}(B)$. Then 1_G is lower semicontinuous, 1_F is upper semicontinuous and $1_G \leq 1_B \leq 1_F$. Hence (5.2.4) follows from (5.2.1+2) applied to 1_G and 1_F . Moreover, since 1_B is continuous at s for $s \notin \mathrm{bd}\,(B)$, we see that (5.2.5) follows from (5.2.3).

Suppose that $f \in \mathcal{C}(\mu)$ satisfies (5.2.6), and let $\varepsilon > 0$ be given. Since $f \in L^1(\mu)$, there exists $q \in \mathbf{N}$ such that

$$\int_{S \setminus G} |f| \, d\mu < \varepsilon \qquad \text{where} \qquad G = \{|f| < q\}$$

Let $Y_n = |f(X_n)|$ and $h = 1_G |f|$. Then $h \in \mathrm{Lsc}(\mu)$ and so by (5.2.1) and (5.2.6) we have

$$\begin{aligned}
\varepsilon > \int_{S \setminus G} |f| \, d\mu &= \int_S |f| \, d\mu - \int_S h \, d\mu \\
&\geq \limsup_{n \to \infty} E|f(X_n)| - \liminf_{n \to \infty} Eh(X_n) \\
&\geq \limsup_{n \to \infty} E\big(|f(X_n)| - h(X_n)\big) \\
&= \limsup_{n \to \infty} \int_{\{Y_n \geq q\}} Y_n \, dP
\end{aligned}$$

Hence, we can find an integer $m \in \mathbf{N}$ such that

$$\int_{\{Y_n \geq q\}} Y_n \, dP < \varepsilon \qquad \forall n \geq m$$

Thus, by (3.24.4) we conclude that $f \in \mathcal{M}_u(X_n)$. □

5.3 A Criterion for Convergence in Law

(1): Let (S, d) be a metric space and let $\nu, \nu_1, \nu_2, \ldots$ be Borel measures on S. Let $\mathcal{W}$ be a set of non-negative Borel functions on S, and let $\mathcal{H} \subseteq \mathrm{Lsc}(\nu)$ such that $0 \leq h \leq 1$ for all $h \in \mathcal{H}$ and

(5.3.1) $$\limsup_{n \to \infty} \int_S h \, d\nu_n \leq \int_S h \, d\nu < \infty \qquad \forall h \in \mathcal{H}$$

(5.3.2) $$\int_S g \, d\nu \leq \liminf_{n \to \infty} \int_S g \, d\nu_n \qquad \forall g \in \mathcal{W}$$

(5.3.3) $$\nu(G) = \sup\left\{ \int_S g \, d\nu \;\middle|\; g \in \mathcal{W}^* \;,\; g \leq 1_G \right\} \qquad \forall G \text{ open } \subseteq S$$

where $\mathcal{W}^* = con_\sigma(\mathcal{W})$ *is the smallest* $(\uparrow)$*-stable convex cone containing* $\mathcal{W}$ *[see (1.49)]. Set*

$$R(f) := \inf_{h\in\mathcal{H}}\left\{\limsup_{n\to\infty}\int_S (1-h(s))\,|f(s)|\,\nu_n(ds)\right\} \quad \forall f \in \mathcal{M}(S) \tag{5.3.4}$$

$$\mathcal{M}_u(\nu_n) = \left\{ f \in \mathcal{M}(S) \;\middle|\; \lim_{q\to\infty}\left(\limsup_{n\to\infty}\int_{\{|f|>q\}} |f|d\nu_n\right) = 0 \right\} \tag{5.3.5}$$

(1): If $f \in \mathcal{C}(\nu)\cap\mathcal{M}_u(\nu_n)$ *is a given function with* $R(f) < \infty$ *, then*

$$\int_S |f|\,d\nu \le \liminf_{n\to\infty}\int_S |f|\,d\nu_n \le \limsup_{n\to\infty}\int_S |f|\,d\nu_n < \infty \tag{5.3.6}$$

$$\limsup_{n\to\infty}\int_S |f|\,d\nu_n = R(f) + \sup_{h\in\mathcal{H}}\int_S |f|h\,d\nu \tag{5.3.7}$$

$$\lim_{n\to\infty}\int_S f\,d\nu_n = \int_S f\,d\nu \quad \text{if} \quad R(f) = 0 \tag{5.3.8}$$

(2): Let μ *be a Borel probability measure on* S *and let* $X_1, X_2, \ldots$ *be* S*-valued random functions. If* $\mathcal{G} \subseteq \mathcal{B}(S)$ *is a paving satisfying*

$$\mu(U) = \sup\{\,\mu(G) \mid G \in \hat{\mathcal{G}}\,,\; G \subseteq U\,\} \quad \forall U \text{ open } \subseteq S \tag{5.3.9}$$

where $\hat{\mathcal{G}}$ *is the smallest* $(\sum c, \uparrow)$*-stable paving containing* $\mathcal{G}$ *, then the following three statements are equivalent:*

$$X_n \xrightarrow{\sim} \mu \tag{5.3.10}$$

$$\int_S g\,d\mu \le \liminf_{n\to\infty} Eg(X_n) \qquad \forall g \in \mathrm{Lip}_1^+(S) \tag{5.3.11}$$

$$\mu(G) \le \liminf_{n\to\infty} P(X_n \in G) \qquad \forall G \in \mathcal{G} \tag{5.3.12}$$

Remarks: (1): Note that the measures $\nu, \nu_1, \nu_2, \ldots$ are not assumed to be finite. Hence, we may no longer define $\nu_n \xrightarrow{\sim} \nu$ by (5.1.1) because a bounded continuous function need no longer be integrable.

(2): Note that every bounded measurable function $f : S \to \mathbf{R}$ belongs to $\mathcal{M}_u(\nu_n)$. Moreover, if $1_S \in \mathcal{H}$, then $R(f) = 0$ for all $f \in \mathcal{M}(S)$.

Proof: (1): Let $\mathcal{L}$ be the set of all non-negative Borel functions satisfying (5.3.2). Then $\mathcal{L}$ is a convex cone and if $f : S \to [0,\infty]$ is a given non-negative Borel function, then I claim that we have

(i) $\displaystyle\int_S f\,d\nu = \sup\left\{ \int_S g\,d\nu \;\middle|\; g \in \mathcal{L}\,,\ g \le f \right\} \quad \Rightarrow \quad f \in \mathcal{L}$

(ii) $\displaystyle f \in \mathrm{Lsc}(\nu) \quad \Rightarrow \quad \int_S f\,d\nu \le \liminf_{n\to\infty} \int_S f\,d\nu_n$

(iii) $\displaystyle f \in \mathrm{Usc}(\nu) \cap \mathcal{M}_u(\nu_n) \text{ and } R(f) < \infty \quad \Rightarrow \quad \limsup_{n\to\infty} \int_S f\,d\nu_n < \infty$

(iv) $\displaystyle f \in \mathrm{Usc}(\nu) \cap \mathcal{M}_u(\nu_n) \quad \Rightarrow \quad \limsup_{n\to\infty} \int_S f\,d\nu_n \le R(f) + \sup_{h\in\mathcal{H}} \int_S fh\,d\nu$

(i): Let $a < \int_S f\,d\nu$. Then by assumption there exists $g \in \mathcal{L}$ such that $g \le f$ and $\int_S g\,d\nu > a$. Hence,

$$a < \int_S g\,d\nu \le \liminf_{n\to\infty} \int_S g\,d\nu_n \le \liminf_{n\to\infty} \int_S f\,d\nu_n$$

Letting $a \uparrow \int_S f\,d\nu$, we see that (i) holds.

(ii): Let $f \in \mathrm{Lsc}(\nu)$ be a given non-negative function. By (i) and the monotone convergence theorem we see that $\mathcal{L}$ is $(\uparrow)$ -stable. Hence, by (5.3.2+3) and (i) we have $\mathcal{L} \supseteq \mathcal{W}^*$ and $1_G \in \mathcal{L}$ for all open sets $G \subseteq S$. Moreover, by (P.5.31)–(P.5.35) there exists an increasing sequence of bounded, non-negative continuous functions $\{f_k\}$ such that $g \le f$ and $g = f$ ν-a.s. where $g = \sup_k f_k$. Then g is a non-negative lower semicontinuous Borel function and so $G_{jn} := \{g > j2^{-n}\}$ is open. By (1.46.2), we have

$$2^{-n} \sum_{j=1}^{\infty} 1_{G_{jn}}(s) \uparrow g(s) \qquad \forall s \in S$$

and since $\mathcal{L}$ is a $(\uparrow)$-stable convex cone containing 1_G for all open set sets $G \subseteq S$, we see that $g \in \mathcal{L}$. Thus, (ii) follows from (i).

(iii)+(iv): Let $f \in \mathrm{Usc}(\nu) \cap \mathcal{M}_u(\nu_n)$ $\varepsilon >$ be given. If $R(f) = \infty$ then (iv) holds trivially. So suppose that $R(f) < \infty$ and let $\varepsilon > 0$ and $c > R(f)$ be given. Then we choose $q > 0$ and $h_0 \in \mathcal{H}$ such that

$$\limsup_{n\to\infty} \int_{\{|f|>q\}} |f|\,d\nu_n < \varepsilon \quad \text{and} \quad \limsup_{n\to\infty} \int_S f(1-h_0)\,d\nu_n < c$$

Setting $f_q = f \wedge q$, we have that $h_0(q - f_q)$ is non-negative and belongs to $\mathrm{Lsc}(\nu)$. Moreover, qh_0 and $h_0 f_q$ are ν-integrable by (5.3.1). Note that

$fh_0 \le f_q h_0 + 1_{\{f>q\}} f$. Hence, by (5.3.1) and (ii) we find

$$\begin{aligned}
\limsup_{n\to\infty} \int_S f\,d\nu_n &\le \limsup_{n\to\infty} \int_S f(1-h_0)\,d\nu_n + \limsup_{n\to\infty} \int_S f h_0\,d\nu_n \\
&\le c + \limsup_{n\to\infty} \int_S f_q h_0\,d\nu_n + \limsup_{n\to\infty} \int_{\{f>q\}} f\,d\nu_n \\
&\le \varepsilon + c + \limsup_{n\to\infty} \Big\{ q\int_S h_0\,d\nu_n - \int_S (q-f_q)h_0\,d\nu_n \Big\} \\
&\le \varepsilon + c + \limsup_{n\to\infty} q\int_S h_0\,d\nu_n - \liminf_{n\to\infty} \int_S (q-f_q)h_0\,d\nu_n \\
&\le \varepsilon + c + q\int_S h_0\,d\nu - \int_S (q-f_q)h_0\,d\nu \\
&= \varepsilon + c + \int_S f_q h_0\,d\nu \le \varepsilon + c + \sup_{h\in\mathcal{H}} \int_S fh\,d\nu
\end{aligned}$$

Thus, (iii) holds and (iv) follows by letting $\varepsilon \downarrow 0$ and $c \downarrow R(f)$.

(5.3.6)–(5.3.8): Suppose that $f \in \mathcal{C}(\nu) \cap \mathcal{M}_u(\nu_n)$ such that $R(f) < \infty$. Then $|f| \in \mathcal{C}(\nu) \cap \mathcal{M}_u(\nu_n)$ and so by (ii) and (iii) we have

$$\int_S |f|\,d\nu \le \liminf_{n\to\infty} \int_S |f|\,d\nu_n \le \limsup_{n\to\infty} \int_S |f|\,d\nu_n < \infty$$

Thus, (5.3.6) is proved. If $h \in \mathcal{H}$, then $|f|h \in \mathrm{Lsc}(\nu)$. So by (ii), (P.4.19) and definition of $R(f)$, we get

$$\begin{aligned}
\limsup_{n\to\infty} \int_S |f|\,d\nu_n &\ge \liminf_{n\to\infty} \int_S |f|h\,d\nu_n + \limsup_{n\to\infty} \int_S |f|(1-h)\,d\nu_n \\
&\ge \int_S |f|h\,d\nu + R(f)
\end{aligned}$$

Hence, taking supremum over $h \in \mathcal{H}$, we obtain $\ge$ in (5.3.7) and the converse inequality follows from (iv). Moreover, if $R(f) = 0$ and $f_1 = (\Re f)^+$, $f_2 = (\Re f)^-$, $f_3 = (\Im f)^+$ and $f_4 = (\Im f)^-$, then $f_j \in \mathcal{C}(\nu) \cap \mathcal{M}_u(\nu_n)$ and $R(f_j) = 0$ for $j = 1,2,3,4$. Recall that $0 \le h \le 1$ for all $h \in \mathcal{H}$. Hence, by (ii) and (iv) we get

$$\int f_j\,d\nu = \lim_{n\to\infty} \int f_j\,d\nu_n \quad \text{for} \quad j = 1,2,3,4$$

and since $f = (f_1 - f_2) + i(f_3 - f_4)$ we see that (5.3.8) holds.

(2): Set $\nu = \mu$, $\nu_n = P_{X_n}$ and $\mathcal{H} = \{1_S\}$. Then (5.3.1) holds and $R(f) = 0$ for all Borel functions f . Moreover, by (5.3.9) we have that $\mathcal{W} := \{1_G \mid G \in \mathcal{G}\}$ satisfies (5.3.3) and by (P.5.31-35) and the monotone convergence theorem (3.4), we have that $\mathcal{W} := \mathrm{Lip}_1^+(S)$ satisfies (5.3.3). Thus, the equivalence of (5.3.10)–(5.3.12) follows from (5.3.8). □

Remarks: (1): Note that the measures $\nu, \nu_1, \nu_2, \ldots$ may be infinite; thus, we can no longer define $\nu_n \xrightarrow{\sim} \nu$ by (5.1.1) because bounded continuous functions no longer need to be integrable.

(2): Observe that every bounded Borel function belongs to $\mathcal{M}_u(\nu_n)$ and that

(5.3.12) $\mathcal{M}_u(P_{X_n}) = \mathcal{M}_u(X_n)$ whenever $X_1, X_2, \ldots$ are S-valued random functions

Moreover, if $1_S \in \mathcal{H}$, then evidently we have $R(f) = 0$ for all $f \in \mathcal{M}(S)$.

5.4 The Calculus of Convergence in Law

Let (S, d) and (T, δ) be metric spaces, let $X, X_1, X_2, \cdots$ be S-valued random functions, and let $\varphi : S \to T$ be a Borel function such that φ is continuous at P_X-a.a. points in S . Then we have the following:

(5.4.1) *If $X_n \xrightarrow{\sim} X$, then $\varphi(X_n) \xrightarrow{\sim} \varphi(X)$*

(5.4.2) *If $X_n \xrightarrow{\sim} X$ and X is degenerated, then $X_n \to X$ in pr*

(5.4.3) *If $X_n \to X$ in pr, then $\varphi(X_n) \to \varphi(X)$ in pr*

(5.4.4) *If $X_n \to X$ in pr, then $X_n \xrightarrow{\sim} X$*

Let $S \times T$ be metrized in the usual way [see (P.5.11)] and let $Z = (X, Y)$ and $Z_n = (X_n, Y_n)$ be $(S \times T)$-valued random functions. Suppose that ψ is a Borel function from $S \times T$ into the metric space (M, ρ) such that ψ is continuous at P_Z-a.a. points in $S \times T$. Then we have the following:

(5.4.5) *If $X_n \to X$ in pr and $Y_n \to Y$ in pr, then $Z_n \to Z$ in pr and $\psi(Z_n) \to \psi(Z)$ in pr*

(5.4.6) *If $X_n \xrightarrow{\sim} X$, $Y_n \xrightarrow{\sim} Y$ and X is degenerated, then $Z_n \xrightarrow{\sim} Z$ and $\psi(Z_n) \xrightarrow{\sim} \psi(Z)$*

(5.4.7) *Suppose that S is separable, and that X and Y are independent and X_n and Y_n are independent for all $n \geq 1$. If $X_n \xrightarrow{\sim} X$ and $Y_n \xrightarrow{\sim} Y$, then $Z_n \xrightarrow{\sim} Z$ and $\psi(Z_n) \xrightarrow{\sim} \psi(Z)$*

Warning: Convergence in law behaves differently from the other convergence notions, such as convergence a.s. or convergence in q-mean. For instance, suppose that $X_n \xrightarrow{\sim} X$ where $X \sim N(0,1)$ and let $Y_n = -X_n$, then $Y_n \xrightarrow{\sim} X$, but $X_n + Y_n \xrightarrow{\sim} 0$, and so $\{X_n + Y_n\}$ does **not** converge in law to $X + X$.

Proof: Let $\varphi : S \to T$ be P_X-a.s. continuous and let $U_n = \varphi(X_n)$ and $U = \varphi(X)$. Then U_n and U are T-valued random functions.

(5.4.1): Let $f \in C(T)$; then $g = f \circ \varphi$ is bounded and belongs to $\mathcal{C}(P_X)$. Hence, by (5.2.3) we have

$$Ef(U) = Eg(X) = \lim_{n\to\infty} Eg(X_n) = \lim_{n\to\infty} Ef(U_n)$$

Thus, $U_n \xrightarrow{\sim} U$ and (5.4.1) is proved.

(5.4.2): Since X is degenerated, there exists $c \in S$ such that $X = c$ a.s. Let $f(s) = \arctan d(s,c)$. Then $f \in C(S)$ and $f(X) = f(c) = 0$ a.s. Hence,

$$E\{\arctan d(X_n, X)\} = Ef(X_n) \to Ef(X) = 0$$

Thus, by (4.3.8) with $\varphi = \arctan$ we have that $X_n \to X$ in *pr.*

(5.4.3): Let C be the set of continuity points of φ and let $\varepsilon > 0$ be given. If we set

$$C_k = \{s \in S \mid \delta(\varphi(s'), \varphi(s'')) \le \varepsilon \quad \forall s', s'' \in b(s, 2^{-k})\} \quad , \quad C_\infty = \bigcup_{k=1}^{\infty} C_k$$

then it is easily seen that C_k is closed and $C_k \uparrow C_\infty$. Moreover, by definition of continuity we have that $C \subseteq C_\infty$. Hence, $P(X \notin C_k) \to P(X \notin C_\infty) = 0$ because φ is continuous at P_X-a.a. points. By definition of C_k , we have

$$\{\delta(U, U_n) > \varepsilon\} \subseteq \{X \notin C_k\} \cup \{d(X, X_n) \ge 2^{-k}\}$$

Thus, by (1.24.3) we find

$$P^*(\delta(U, U_n) > \varepsilon) \le P(X \notin C_k) + P^*(d(X, X_n) \ge 2^{-k})$$

Fix $k \in \mathbf{N}$ and recall that $X_n \to X$ in *pr.* Hence, letting $n \to \infty$ we get

$$\limsup_{n\to\infty} P^*(\delta(U, U_n) > \varepsilon) \le P(X \notin C_k) \qquad \forall k \ge 1$$

and since $P(X \notin C_k) \to 0$, then letting $k \to \infty$, we see that $U_n \to U$ in *pr.*

(5.4.4): Let $f \in C(S)$. Then $f(X_n) \to f(X)$ in pr by (5.4.3) and $\{f(X_n) \mid n \geq 1\}$ is uniformly integrable because f is bounded. Hence, by (3.25.6) we have that $Ef(X_n) \to Ef(X)$ and so $X_n \xrightarrow{\sim} X$.

(5.4.5): Let d_0 be the metric on $S \times T$ defined in (P.5.11). Then we have

$$d_0((s_1,t_1),(s_2,t_2)) = \sqrt{d(s_1,s_2)^2 + \delta(t_1,t_2)^2} \leq d(s_1,s_2) + \delta(t_1,t_2)$$

Hence, $d_0(Z,Z_n) \leq d(X,X_n) + \delta(Y,Y_n)$ and so

$$P^*(d_0(Z,Z_n) > \varepsilon) \leq P^*(d(X,X_n) > \tfrac{\varepsilon}{2}) + P^*(\delta(Y,Y_n) > \tfrac{\varepsilon}{2})$$

Thus, we see that $Z_n \to Z$ in pr and so $\psi(Z_n) \to \psi(Z)$ in pr by (5.4.3).

(5.4.6): Let $f \in \mathrm{Lip}_1^+(S \times T)$. By assumption there exists $c \in S$ such that $X = c$ a.s. Set $h(s) = 1 \wedge d(s,c)$ and $g(t) = f(c,t)$, then $|f(s,t) - f(c,t)| \leq h(s)$ and h and g are bounded and continuous. Hence, we find

$$\begin{aligned} |Ef(Z_n) - Ef(Z)| &\leq E|f(Z_n) - f(c,Y_n)| + |Eg(Y_n) - Eg(Y)| \\ &\leq Eh(X_n) + |Eg(Y_n) - Eg(Y)| \end{aligned}$$

and since $h(X) = h(c) = 0$ a.s., we see that $Ef(Z_n) \to Ef(Z)$ for all $f \in \mathrm{Lip}_1^+(S \times T)$. Thus, by the equivalence of (5.3.10) and (5.3.11), we have that $Z_n \xrightarrow{\sim} Z$ and so $\psi(Z_n) \xrightarrow{\sim} \psi(Z)$ by (5.4.1).

(5.4.7): Let $f \in \mathrm{Lip}_1^+(S \times T)$, and set $g(s) = Ef(s,Y)$ and $g_n(s) = Ef(s,Y_n)$. Then g and g_n belong to $\mathrm{Lip}_1^+(S)$. If we set $L = \{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \cdots\}$ with its usual euclidean metric, then L is a metric space and

$$h(s,v) = \begin{cases} g_n(s) & \text{if } v = \frac{1}{n} \\ g(s) & \text{if } v = 0 \end{cases}$$

is a bounded continuous function from $S \times L$ into $[0,1]$. Let $V_n(\omega) = \frac{1}{n}$ and $V(\omega) = 0$ for all $\omega \in \Omega$. Then V_n and V are L-valued random functions such that $V_n \xrightarrow{\sim} V$ and V is degenerated. Hence, $(X_n, V_n) \xrightarrow{\sim} (X,V)$ by (5.4.6) and $h \in C(S \times L)$. Recall that S is separable by assumption. Hence f is $(\mathcal{B}(S) \otimes \mathcal{B}(T))$-measurable by Exc.1.90 and so by (4.5.2), we get

$$\begin{aligned} Ef(X,Y) = Eg(X) = Eh(X,V) &= \lim_{n\to\infty} Eh(X_n,V_n) \\ &= \lim_{n\to\infty} Eg_n(X_n) = \lim_{n\to\infty} Ef(X_n,Y_n) \end{aligned}$$

Thus, by the equivalence of (5.3.10) and (5.3.11) we have that $Z_n \xrightarrow{\sim} Z$ and so $\psi(Z_n) \xrightarrow{\sim} \psi(Z)$ by (5.4.1). □

5.5 Eventual Tightness

The main virtue of Cauchy's criterion for convergence of sequences of real numbers is that without knowing the limit we can actually decide if a given sequence of real numbers is convergent or divergent. We shall now consider the same problem of deciding when a sequence $\{X_n\}$ of S-valued random functions converges in law without knowing the limit measure. The tool to do this is a certain tightness condition.

Let (S, d) be a metric space and let $\{X_n\}$ be a sequence of S-valued random functions, then we say that $\{X_n\}$ is *eventually tight* if for every $\varepsilon > 0$ there exists a compact set $K \subseteq S$ satisfying

(5.5.1) $$\limsup_{n\to\infty} P(X_n \notin G) < \varepsilon \qquad \forall G \text{ open } \subseteq S \text{ such that } G \supseteq K$$

Suppose that $X_n \xrightarrow{\sim} \mu$ where μ is a Radon probability on S and let $\varepsilon > 0$ be given. Then there exists a compact set $K \subseteq S$ such that $\mu(S \setminus K) < \varepsilon$. Let G be an open set with $G \supseteq K$, then $S \setminus G$ is closed and $S \setminus G \subseteq S \setminus K$. Hence, by (5.2.4) we have

$$\limsup_{n\to\infty} P(X_n \notin G) \le \mu(S \setminus G) \le \mu(S \setminus K) < \varepsilon$$

Thus, we have:

(5.5.2) If $X_n \xrightarrow{\sim} \mu$ where μ is a Radon probability measure, then $\{X_n\}$ is eventually tight

The next theorem gives a partial converse to (5.5.2).

5.6 The Eventual Tightness Theorem

Let (S, d) be a metric space and let $\{X_n\}$ be an eventually tight sequence of S-valued random functions. Then there exist a Radon probability measure μ_0 on S and integers $\sigma(1) < \sigma(2) < \cdots$ such that $X_{\sigma(n)} \xrightarrow{\sim} \mu_0$.

Proof: By eventual tightness we can find compact sets $C_j \subseteq S$ such that $C_1 \subseteq C_2 \subseteq \cdots$ and

(i) $$\limsup_{n\to\infty} P(X_n \notin G) < 2^{-j} \qquad \forall G \text{ open with } G \supseteq C_j$$

Let $S_0 = \cup_1^\infty C_j$. Then S_0 is separable by (P.5.4) and so there exists a countable $(\cup f)$-stable paving $\mathcal{G}$ satisfying (P.5.5) such that every set in $\mathcal{G}$ is open. Let $\mu_n(B) = P(X_n \in B)$ be the law of X_n . Since $\mathcal{G}$ is countable, then by (P.6.10) there exists a function $\lambda : \mathcal{G} \to [0,1]$ and integers $\sigma(1) < \sigma(2) < \cdots$ such that $\mu_{\sigma(n)}(G) \to \lambda(G)$ for all $G \in \mathcal{G}$. Let $\mathcal{K}$ be the paving of all compact subsets of S and set

$$\begin{aligned}
&\lambda_j(K) = \inf\{\lambda(G) \mid G \in \mathcal{G}\,,\, G \supseteq C_j \cap K\} \qquad \forall K \in \mathcal{K} \\
&\mu(K) = \sup\{\lambda_j(K) \mid j \geq 1\} \qquad \forall K \in \mathcal{K} \\
&\mu_0(B) = \sup\{\mu(K) \mid K \in \mathcal{K}\,,\; K \subseteq B\} \qquad \forall B \in \mathcal{B}(S)
\end{aligned}$$

Since $C_j \uparrow S_0$, we have that $\lambda_j(K) \uparrow \mu(K)$ for all $K \in \mathcal{K}$. Moreover, I claim that we have

(ii) μ is superadditive on $\mathcal{K}$

(iii) μ is inner tight on $\mathcal{K}$

(iv) μ_0 is a Radon measure with $\mu_0(S \setminus S_0) = 0$

(v) $X_{\sigma(n)} \xrightarrow{\sim} \mu_0$

(ii): Let K and L be disjoint compact sets and let $G \in \mathcal{G}$ such that $G \supseteq C_j \cap (K \cup L)$. Since $C_j \cap K$ and $C_j \cap L$ are disjoint compact sets and G is open, there exists two disjoint open sets U and V such that $C_j \cap K \subseteq U \subseteq G$ and $C_j \cap L \subseteq V \subseteq G$. Then by (P.5.5) and compactness of $C_j \cap K$ and $C_j \cap L$, it follows easily that there exists $G_0, G_1 \in \mathcal{G}$ satisfying

$$C_j \cap K \subseteq G_0 \subseteq U \subseteq G \quad , \quad C_j \cap L \subseteq G_1 \subseteq V \subseteq G$$

Since $U \cap V = \emptyset$, we have that $G_0 \cap G_1 = \emptyset$ and so

$$\begin{aligned}
\lambda(G) &= \lim_{n\to\infty} \mu_{\sigma(n)}(G) \geq \lim_{n\to\infty} \mu_{\sigma(n)}(G_0 \cup G_1) \\
&= \lim_{n\to\infty} \{\mu_{\sigma(n)}(G_0) + \mu_{\sigma(n)}(G_1)\} \\
&= \lambda(G_0) + \lambda(G_1) \geq \lambda_j(K) + \lambda_j(L)
\end{aligned}$$

Taking inf over all $G \in \mathcal{G}$ with $G \supseteq C_j \cap (K \cup L)$, we see that λ_j is superadditive on $\mathcal{K}$. Thus, $\mu = \lim \lambda_j$ is superadditive on $\mathcal{K}$.

(iii): Let $K, L \in \mathcal{K}$, $j \in \mathbf{N}$ and $\varepsilon > 0$ be given. Then there exist $G_0 \in \mathcal{G}$ such that $G_0 \supseteq K \cap L \cap C_j$ and

$$\lambda(G_0) \leq \varepsilon + \lambda_j(K \cap L) \leq \varepsilon + \mu(K \cap L)$$

If $M = (K \cap C_j) \setminus G_0$, then $M \in \mathcal{K}$ and $M \subseteq K \setminus L$. Hence, there exists $G_1 \in \mathcal{G}$ such that $G_1 \supseteq M \cap C_j$ and

$$\lambda(G_1) \leq \varepsilon + \lambda_j(M) \leq \varepsilon + \mu(M) \leq \varepsilon + \mu_*(K \setminus L)$$

Let $G = G_0 \cup G_1$. Then $G \supseteq K \cap C_j$ and $G \in \mathcal{G}$ because $\mathcal{G}$ is $(\cup f)$-stable. Hence, we have

$$\begin{aligned}\lambda_j(K) \leq \lambda(G) = \lim_{n\to\infty} \mu_{\sigma(n)}(G_0 \cup G_1) &\leq \lambda(G_0) + \lambda(G_1) \\ &\leq 2\varepsilon + \mu_*(K \cap L) + \mu_*(K \setminus L)\end{aligned}$$

Letting $\varepsilon \downarrow 0$ and $j \uparrow \infty$, we see that

$$\mu(K) \leq \mu_*(K \cap L) + \mu_*(K \setminus L) \qquad \forall K, L \in \mathcal{K}$$

The converse inequality follows from superadditivity of μ and μ_* . Thus, (iii) is proved.

(iv): Evidently, we have that $\mathcal{K}$ is a compact $(\cup f, \cap c)$-stable paving on S . Hence, by (iii) and (1.31) we know that (1.31.1)-(1.31.4) hold. Let F be a closed set. Then $F \cap K \in \mathcal{K} \subseteq \mathcal{M}(\mu_*)$ for all $K \in \mathcal{K}$ and so by (1.31.3+4), we have that $\mathcal{M}(\mu_*)$ is a σ-algebra containing all closed sets. Thus, $\mathcal{M}(\mu_*) \supseteq \mathcal{B}(S)$. Note that $\mu_0(B) = \mu_*(B)$ for all $B \in \mathcal{B}(S)$ by (1.31.1). Hence, by (1.31.3) and definition of μ_0 we see that μ_0 is a Radon measure on S such that $\mu_0(S_0) \leq \mu_0(S) \leq 1$. By (i) we have

$$\lambda_j(C_i) \geq \lambda_i(C_i) \geq 1 - 2^{-i} \qquad \forall 1 \leq i \leq j$$

Letting $j \to \infty$, we see that $\mu_0(C_i) = \mu(C_i) \geq 1 - 2^{-i}$ and so

$$1 \geq \mu_0(S) \geq \mu_0(S_0) = \lim_{i\to\infty} \mu_0(C_i) = 1$$

Hence, μ_0 is a Radon probability on S with $\mu_0(S \setminus S_0) = 0$.

(v): Let $\hat{\mathcal{G}}$ be the smallest $(\sum c, \uparrow)$-stable paving containing $\mathcal{G}$ and let U be a given open set. By (P.5.5) there exists an increasing sequence $\{G_n\} \subseteq \mathcal{G}$ such that $G \subseteq U$ and $G \cap S_0 = U \cap S_0$ where $G = \cup_1^\infty G_n$. Then $G \in \hat{\mathcal{G}}$ and since $U \setminus G \subseteq S \setminus S_0$ and $\mu_0(S \setminus S_0) = 0$, then by (1.3.1) we have

$$\mu_0(U) = \mu_0(G) + \mu_0(U \setminus G) \leq \mu_0(G) + \mu_0(S \setminus S_0) = \mu_0(G)$$

Hence, $\mathcal{G}$ satisfies (5.3.9) and if $G \in \mathcal{G}$, then $\mu(K) \leq \lambda(G)$ for every compact set $K \subseteq G$. So by (1.31.1), we get

$$\mu_0(G) \leq \lambda(G) = \lim_{n\to\infty} P(X_{\sigma(n)} \in G) \qquad \forall G \in \mathcal{G}$$

Thus, by the equivalence of (5.3.10) and (5.3.12) we have $X_{\sigma(n)} \xrightarrow{\sim} \mu_0$. □

5.7 The Continuity Theorem

Let (S,d) be a metric space and let $\{X_n\}$ be a sequence of S-valued random functions. Let $(T,\mathcal{B})$ be a measurable space and let Γ be a kernel on $S \times T$ [see (4.14)] such that $\Gamma(\cdot\,,t)$ is continuous for all $t \in T$. Set

$$T_b = \left\{ t \in T \,\middle|\, \sup_{n\geq 1} E\,|\Gamma(X_n,t)| < \infty \right\}$$

$$T_u = \left\{ t \in T \,\middle|\, \{\Gamma(X_n,t)\}_1^\infty \quad \textit{is uniformly integrable} \right\}$$

If $X_n \xrightarrow{\sim} \mu$ for some Borel probability measure μ on S, then we have

$$(5.7.1) \qquad T_u \subseteq T_b \subseteq \mathcal{D}(\Gamma_\mu) \quad , \quad \Gamma_\mu(t) = \lim_{n\to\infty} \Gamma_{X_n}(t) \quad \forall t \in T_u$$

Conversely, if $\{X_n\}$ is eventually tight and $\lim_{n\to\infty} \Gamma_{X_n}(t) = \gamma(t)$ for all $t \in T_0$, where $T_0 \subseteq T$ and $\gamma : T_0 \to \mathbf{C}$ is a function such that there exists at most one Radon probability measure μ on S satisfying

$$(5.7.2) \qquad t \in \mathcal{D}(\Gamma_\mu) \quad \textit{and} \quad \Gamma_\mu(t) = \gamma(t) \quad \forall t \in T_0 \cap T_u$$

then there exist a unique Radon probability measure μ on S satisfying

$$(5.7.3) \qquad X_n \xrightarrow{\sim} \mu \quad \textit{and} \quad \Gamma_\mu(t) = \gamma(t) \quad \forall t \in T_0 \cap T_u$$

Remarks: (1): Note that if $\Gamma(\cdot\,,t)$ is bounded for all $t \in T$, then $T_b = T_u = T$. Recall that the uniqueness theorem for transforms (4.15) provides a powerful tool for verifying that there exists at most one Radon probability satisfying (5.7.2).

(2): Suppose that $\Gamma_{X_n}(t) \to \gamma(t)$ for all $t \in T_0$. Then below we shall see that some sort continuity of the function γ often will imply eventual tightness of $\{X_n\}$. This is the reason for the name "the continuity theorem".

Proof: (5.7.1): Let $t \in T_b$ and set $f(s) = |\Gamma(s,t)|$. Then f is continuous, and $f \geq 0$. Since $t \in T_b$, by (5.2.1) we have

$$\int_S f\,d\mu \leq \liminf_{n\to\infty} Ef(X_n) \leq \sup_{n\geq 1} Ef(X_n) < \infty$$

Thus, $\Gamma(\cdot\,,t) \in L^1(\mu)$ and so $t \in \mathcal{D}(\Gamma_\mu)$. Let $t \in T_u$ be given. Then by (5.2.3) we have that

$$\Gamma_\mu(t) = \int_S \Gamma(s,t)\mu(ds) = \lim_{n\to\infty} E\,\Gamma(X_n,t)$$

and so (5.7.1) holds.

(5.7.3): By the eventual tightness theorem (5.6) there exists a Radon probability measure μ on S, and integers $\sigma(1) < \sigma(2) < \cdots$ such that $X_{\sigma(n)} \xrightarrow{\sim} \mu$. I claim that we have $X_n \xrightarrow{\sim} \mu$. Suppose this is not so. Then there exists $f \in C(S)$ such that $Ef(X_n) \not\to \int_S f\,d\mu$. But then there exists $\varepsilon_0 > 0$ and integers $\tau(1) < \tau(2) < \cdots$ such that

$$(*) \qquad \left| Ef(X_{\tau(n)}) - \int_S f\,d\mu \right| \geq \varepsilon_0 \qquad \forall n \geq 1$$

By the eventual tightness theorem (5.6) we may assume that $X_{\tau(n)} \xrightarrow{\sim} \nu$ for some Radon measure ν. By (5.7.1) we have that μ and ν both satisfy (5.7.2). Hence, by assumption we have $\mu = \nu$, and so we conclude that $X_{\tau(n)} \xrightarrow{\sim} \mu$. But this contradicts (*) and thus, we must have that $X_n \xrightarrow{\sim} \mu$. □

5.8 Criteria for Eventual Tightness on $\mathbf{R}^k$

Let $\{X_n\}$ be a sequence of k-dimensional random vectors. Then the following four statements are equivalent:

(5.8.1) $\{X_n \mid n \geq 1\}$ *is eventually tight*

(5.8.2) $\forall \varepsilon > 0\ \exists a > 0$ *so that* $\limsup_{n\to\infty} P(||X_n|| \geq a) \leq \varepsilon$

(5.8.3) $\forall \varepsilon > 0\ \exists a > 0$ *so that* $P(||X_n|| > a) \leq \varepsilon \quad \forall n \geq 1$

(5.8.4) $\forall \varepsilon > 0\ \exists g : \mathbf{R}^k \to [0,\infty] \ : \ \limsup_{n\to\infty} Eg(X_n) < \varepsilon \cdot \liminf_{||x||\to\infty} g(x)$

Proof: (5.8.1) $\Rightarrow$ (5.8.2.): Let $\varepsilon > 0$ be given. Since $\{X_n\}$ is eventually tight there exists a compact set $K \subseteq \mathbf{R}^k$ so that $\limsup P(X_n \notin G) < \varepsilon$ for every open set $G \supseteq K$, and since K is bounded there exist $a > 0$ such that $K \subseteq G$, where $G = \{x \mid ||x|| < a\}$. Then G is open and so

$$\limsup_{n\to\infty} P(||X_n|| \geq a) \leq \limsup_{n\to\infty} P(X_n \notin G) < \varepsilon$$

(5.8.2.) $\Rightarrow$ (5.8.3.): Let $\varepsilon > 0$ be given. Then by (5.8.2) there exists $c > 0$ and $q \in \mathbf{N}$ such that $P(||X_n|| > c) \leq \varepsilon$ for all $n \geq q$, and since $P(||X_n|| > a) \to 0$ as $a \to \infty$ there exists $a \geq c$ such that $P(||X_n|| > a) \leq \varepsilon$ for all $n = 1, 2, \ldots, q$. Hence, (5.8.3) holds.

(5.8.3) $\Rightarrow$ (5.8.4): Let $0 = a_0 < a_1 < a_2 < \cdots$ be chosen such that $a_n \to \infty$ and $P(||X_n|| \geq a_k) \leq 4^{-k}$ for all $n \geq 1$ and all $k \geq 0$. If we set $g(x) = 2^k$

when $a_k \le ||x|| < a_{k+1}$, then g is a Borel function from $\mathbf{R}^k$ into $\mathbf{R}_+$ such that $g(x) \to \infty$ as $||x|| \to \infty$. Moreover,

$$E\,g(X_n) = \sum_{k=0}^{\infty} 2^k\, P(a_k \le ||X_n|| < a_{k+1}) \le \sum_{n=0}^{\infty} 2^{-k} = 2$$

Hence, we get

$$\limsup_{n\to\infty} Eg(X_n) \le 2 < \varepsilon \cdot \infty = \varepsilon \cdot \liminf_{||x||\to\infty} g(x)$$

for all $\varepsilon > 0$. Thus, (5.8.4) holds.

(5.8.4) $\Rightarrow$ (5.8.1): Let $\varepsilon > 0$ be given and choose $g : \mathbf{R}^k \to [0, \infty]$ according to (5.8.4). Then by assumption we can find numbers $c \in \mathbf{R}$ and $a > 0$ satisfying

$$\limsup_{n\to\infty} Eg(X_n) < c < \varepsilon \cdot g(y) \qquad \forall\, ||y|| \ge a$$

Set $K = \{x \in \mathbf{R}^k \mid ||x|| \le a\}$. Then K is compact and if G is an open set with $G \supseteq K$, then we have that $\frac{\varepsilon}{c}\, g(x) \ge 1$ for all $x \in \mathbf{R}^k \setminus G$. Hence, $P(X_n \notin G) \le \frac{\varepsilon}{c} Eg(X_n)$ and so

$$\limsup_{n\to\infty} P(X_n \notin G) \le \frac{\varepsilon}{c} \limsup_{n\to\infty} Eg(X_n) \le \varepsilon$$

which shows that $\{X_n\}$ is eventually tight. □

5.9 Convergence of Distribution Functions

Let X and X_n be real random variables with distribution functions $F(x)$ and $F_n(x)$ for all $n \ge 1$. If C_F is the set of all continuity points of F , then the following four statements are equivalent:

(5.9.1) $\quad X_n \xrightarrow{\sim} X$

(5.9.2) $\quad F(x-) \le \liminf_{n\to\infty} F_n(x) \le \limsup_{n\to\infty} F_n(x) \le F(x) \qquad \forall\, x \in \mathbf{R}$

(5.9.3) $\quad F(x) = \lim_{n\to\infty} F_n(x) \qquad \forall\, x \in C_F$

(5.9.4) $\quad \exists D$ *dense in* $\mathbf{R}$ *so that* $F(x) = \lim_{n\to\infty} F_n(x) \qquad \forall x \in D$

Proof: (5.9.1) $\Rightarrow$ (5.9.2): Note that $\operatorname{int}(]-\infty,x]) =]-\infty,x[$ and $\operatorname{cl}(]-\infty,x]) =]-\infty,x]$. Hence, by (5.2.4) we see that (5.9.1) implies (5.9.2).

(5.9.2) $\Rightarrow$ (5.9.3): Note that $F(x-) = F(x)$ for all $x \in C_F$. Hence (5.9.2) implies (5.9.3).

(5.9.3) $\Rightarrow$ (5.9.4): By (2.14.13) and (2.14.14) we have that the set of discontinuity points is at most countable. Hence, $D = C_F$ is dense in $\mathbf{R}$. Thus, (5.9.4) follows from (5.9.3).

(5.9.4) $\Rightarrow$ (5.9.1): Let D be a dense subset of $\mathbf{R}$ satisfying the hypothesis of (5.9.4) and let $\mathcal{G} = \{]a,b] \mid a,b \in D\}$. Then $P(X_n \in G) \to P(X \in G)$ for all $G \in \mathcal{G}$, and since every open subset of $\mathbf{R}$ is a countable union of disjoint open intervals, and every open interval is an increasing union of intervals from $\mathcal{G}$, we see that (5.3.9) holds. Hence, by the equivalence of (5.3.10) and (5.3.12) we conclude, that $X_n \xrightarrow{\sim} X$. □

5.10 Convergence of Characteristic Functions

Let X and X_n be k-dimensional random vectors with characteristic functions $\varphi(t)$ and $\varphi_n(t)$ for all $n = 1,2,\ldots$ and let $\gamma : \mathbf{R}^k \to \mathbf{C}$ be a function such that γ is continuous at 0 . If $\hat{\mu}$ denotes the Fourier transform of μ [see (4.18.18)], then we have the following:

(5.10.1) *If $\lim_{n\to\infty} \varphi_n(t) = \gamma(t)$ for all $t \in \mathbf{R}^k$, then there exists a Borel probability measure μ on $\mathbf{R}^k$ such that $X_n \xrightarrow{\sim} \mu$ and $\hat{\mu}(t) = \gamma(t)$ for all $t \in \mathbf{R}^k$*

(5.10.2) $$X_n \xrightarrow{\sim} X \iff \lim_{n\to\infty} \varphi_n(t) = \varphi(t) \quad \forall t \in \mathbf{R}^k$$

Remark: The condition that γ is continuous at 0 is essential and cannot be removed. It is this continuity condition that has baptized (5.7) "the continuity theorem."

Proof: (5.10.1): Let $\varepsilon > 0$ be given and let $U = (U_1,\ldots,U_k) \sim N_k(0,I_k)$ be a k-dimensional normally distributed random vector where I_k denotes the $(k \times k)$-identity matrix. Since $\gamma(0) = 1$ and $|\gamma(t)| \le 1$ for all t by (4.18.2), then by continuity of γ at 0 and Lebesgue's dominated convergence theorem (3.6) there exists $a > 0$ so that $|1 - E\gamma(aU)| < \varepsilon$. Moreover, since $\varphi_n \to \gamma$ and $|\varphi_n| \le 1$, then by Lebesgue's dominated convergence theorem (3.6) there exists $q \in \mathbf{N}$ such that $|1 - E\varphi_n(aU)| < \varepsilon$ for all $n \ge q$. If Ψ is the characteristic function of U , then by Parseval's equation (4.18.4) we have

$E\varphi_n(aU) = E\Psi(aX_n)$ and by (4.22.9), we have that $\Psi(t) = \exp(-\frac{1}{2}||t||^2)$. Set $g(t) = 1 - \Psi(at)$. Then $\lim_{||t||\to\infty} g(t) = 1$ and

$$Eg(X_n) = 1 - E\Psi(aX_n) = 1 - E\varphi_n(aU) < \varepsilon \qquad \forall n \geq q$$

Hence, (5.8.4) holds and so $\{X_n\}$ is eventually tight. Thus, by (4.18.8) and the continuity theorem (5.7) we conclude that $X_n \xrightarrow{\sim} \mu$ for some Borel probability measure with $\hat{\mu}(t) = \gamma(t)$.

(5.10.2): If $X_n \xrightarrow{\sim} X$, then $\varphi_n(t) \to \varphi(t)$ by (5.1.3). Conversely, if $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbf{R}^k$. Then by (4.18.3) and (5.10.1) there exists a Borel probability measure μ such that $X_n \xrightarrow{\sim} \mu$ and $\hat{\mu} = \varphi$. But then $\mu = P_X$ by (4.18.8) and so $X_n \xrightarrow{\sim} X$. □

5.11 Convergence of Probability Mass Functions

Let X and X_n be integer-valued random variables with probability mass functions $p(x)$ and $p_n(x)$ for all $n \in \mathbf{N}$. If $B(\mathbf{R})$ is the set of all bounded Borel functions from $\mathbf{R}$ into $\mathbf{R}$, then the following three statements are equivalent:

(5.11.1) $\quad X_n \xrightarrow{\sim} X$

(5.11.2) $\quad Ef(X) = \lim_{n\to\infty} Ef(X_n) \qquad \forall f \in B(\mathbf{R})$

(5.11.3) $\quad p(k) = \lim_{n\to\infty} p_n(k) \qquad \forall k \in \mathbf{Z}$

Warning: This equivalence does **not** hold if we only assume that X_n and X are discrete random variables. For instance, if $X_n = \frac{1}{n}$ and $X = 0$, then $X_n \xrightarrow{\sim} X$, but $p_n(0) = 0$ and $p(0) = 1$.

Proof: (5.11.1) $\Rightarrow$ (5.11.2). Let $f \in B(\mathbf{R})$. It is then easily seen that there exists $g \in C(\mathbf{R})$ such that $g(n) = f(n)$ for all $n \in \mathbf{Z}$. Since X_n and X are integer-valued, we have $f(X) = g(X)$ and $f(X_n) = g(X_n)$. Hence, (5.11.2) follows from (5.11.1).

(5.11.2) $\Rightarrow$ (5.11.3): Apply (5.11.2) with $f = 1_{\{k\}}$

(5.11.3) $\Rightarrow$ (5.11.1): Let $\mathcal{G} = \{\{k\} \mid k \in \mathbf{Z}\}$. Then $P(X_n \in G) \to P(X \in G)$ for all $G \in \mathcal{G}$ by (5.11.3). If U is an open subset of $\mathbf{R}$, then $V := U \cap \mathbf{Z}$ is a disjoint countable union of singletons belonging to $\mathcal{G}$ and $P(X \in U \setminus V) = 0$. Hence, $X_n \xrightarrow{\sim} X$ by the equivalence of (5.3.10) and (5.3.12). □

5.12 Convergence of Absolute Moments

Let X and X_n be random variables with absolute moment transforms $m(q)$ and $m_n(q)$. Then the sets

$$T_b = \left\{ q \geq 0 \,\middle|\, \sup_n E|X_n|^q < \infty \right\}$$
$$T_u = \left\{ q \geq 0 \,\middle|\, \{ |X_n|^q \mid n \geq 1 \} \quad \textit{is uniformly integrable} \right\}$$

are intervals containing 0 such that

(5.12.1) $\quad \operatorname{int}(T_b) \subseteq T_u \subseteq T_b$

(5.12.2) $\quad$ *If $T_b \neq \{0\}$, then $\{X_n\}$ is eventually tight*

Moreover, we have the following:

(1): If $X_n \xrightarrow{\sim} X$ and $f : \mathbf{R} \to \mathbf{C}$ is a continuous function, then we have

(5.12.3) $\quad T_b \subseteq \mathcal{D}(m) \quad$ *and* $\quad \lim_{n\to\infty} m_n(q) = m(q) \qquad \forall q \in T_u$

(5.12.4) $\quad$ *If $q \in T_b$ and $\lim_{n\to\infty} m_n(q) = m(q)$, then $q \in T_u$*

(5.12.5) $\quad Ef(X_n) \to Ef(X) \quad$ *if $f \in O(|x|^r)$ and $r \in T_u$*

(5.12.6) $\quad Ef(X_n) \to Ef(X) \quad$ *if $f \in o(|x|^q)$ and $q \in T_b$*

(2): Suppose that $\{m_n(q) \mid n \geq 1\}$ converges for all $q \in I$ for some open non-empty interval $I \subseteq T_b$. Then there exists a Borel probability measure μ on $[0, \infty[$ satisfying

(5.12.7) $\quad |X_n| \xrightarrow{\sim} \mu \quad$ *and* $\quad m_n(q) \to \int_0^\infty x^q \mu(dx) \quad \forall q \in T_u$

(3): Suppose that $m_n(q) \to m(q)$ for all $q \in I$ for some open non-empty interval $I \subseteq T_b$. Then $|X_n| \xrightarrow{\sim} |X|$.

Proof: Since $|x|^q \in O(|x|^r)$ for all $0 \leq q \leq r$, then by (5.1.13) and (5.1.15) we have that T_b and T_u are intervals which evidently contains 0.

(5.12.1): Since $|x|^q \in o(|x|^r)$ for all $0 \leq q < r$, then by (5.1.14) we have that $\operatorname{int}(T_b) \subseteq T_u$ and clearly $T_u \subseteq T_b$.

(5.12.2): Since $|x|^q \to \infty$ as $|x| \to \infty$ for $q > 0$, we see that (5.8.4) holds whenever $T_b \neq \{0\}$. Hence, $\{X_n\}$ is eventually tight by (5.8).

(1) follows easily from (5.1.13)–(5.1.15) and (5.2).

(2): By (5.12.2) we have that $\{X_n \mid n \geq 1\}$ is eventually tight, and so (2) follows from (4.20.1) and the continuity theorem (5.7).

(3): By (2) we have that $|X_n| \xrightarrow{\sim} \mu$ for some probability measure μ on $[0, \infty[$ with $m(q) = \int_0^\infty x^q \, d\mu$ for all $q \in I$. But then $P_{|X|} = \mu$ by (4.20.1), and so $|X_n| \xrightarrow{\sim} |X|$. □

5.13 Convergence of Moments

Let X and X_n be random variables with moment transforms $\mu(k)$ and $\mu_n(k)$, and let $a_0, a_1, a_2, \ldots$ be a sequence of real numbers satisfying

$$\limsup_{k\to\infty} \left\{ \frac{|a_{2k}|}{(2k)!} \right\}^{1/2k} < \infty \tag{5.13.1}$$

(1): If $X_n \xrightarrow{\sim} X$ and $0 < q < \infty$ is a positive number with $\sup_n E|X_n|^q < \infty$, then $\mu(k) = \lim_{n\to\infty} \mu_n(k)$ for all $k \in \mathbf{N}_0$ with $k < q$.

(2): If $\mathcal{D}(\mu_n) = \mathbf{N}_0$ for all $n \geq 1$ and $a_k = \lim_{n\to\infty} \mu_n(k)$ for all $k = 0, 1, 2, \ldots$, then there exists a probability measure ν on $\mathbf{R}$ satisfying

$$X_n \xrightarrow{\sim} \nu \quad \text{and} \quad a_k = \int_{-\infty}^{\infty} x^k \nu(dx) \quad \forall k = 0, 1, 2, \ldots \tag{5.13.2}$$

(3): Suppose that $\mathcal{D}(\mu_n) = \mathcal{D}(\mu) = \mathbf{N}_0$ for all $n \geq 1$. If $\mu(k) = \lim_{n\to\infty} \mu_n(k)$ for all $k = 0, 1, 2, \ldots$ and

$$\limsup_{k\to\infty} \left\{ \frac{\mu(2k)}{(2k)!} \right\}^{1/2k} < \infty \tag{5.13.3}$$

then $X_n \xrightarrow{\sim} X$.

Remark: By the remark to (4.20.2), we may replace condition (5.13.1) by the slightly weaker condition

$$\sum_{k=1}^{\infty} |a_{2k}|^{-1/2k} = \infty \tag{5.13.4}$$

Similarly, we may replace condition (5.13.3) by

$$\sum_{k=1}^{\infty} \mu(2k)^{-1/2k} = \infty \tag{5.13.5}$$

Proof: (1) follows directly from (5.12.3).

(2): Since $\mu_n(2) = EX_n^2$ is convergent and thus, bounded, we have that $\{X_n\}$ is eventually tight by (5.12.2). But then (2) follows from (4.20.2) and the continuity theorem (5.7).

(3): By (2) we have that $X_n \xrightarrow{\sim} \nu$ for some probability measure ν with $\mu(k) = \int_{-\infty}^{\infty} x^k \, d\nu$ for all $k \in \mathbf{N}_0$. But then $P_X = \nu$ by (4.20.2) and so $X_n \xrightarrow{\sim} X$. □

5.14 Convergence of Laplace Transforms

Let X and X_n be random variables with Laplace transforms $L(z)$ and $L_n(z)$ and let

$$S_b = \left\{ s \in \mathbf{R} \,\middle|\, \sup_n E(e^{sX_n}) < \infty \right\}$$

$$S_u = \left\{ s \in \mathbf{R} \,\middle|\, \{ e^{sX_n} \mid n \geq 1 \} \text{ is uniformly integrable} \right\}$$

Then S_b and S_u are intervals containing 0 such that

(5.14.1) $\quad \operatorname{int}(S_b) \subseteq S_u \subseteq S_b$

Moreover, we have the following:

(1): If $X_n \xrightarrow{\sim} X$ and $f : \mathbf{R}^k \to \mathbf{C}$ is a continuous function, then we have

(5.14.2) $\quad S_b \subseteq \mathcal{D}(L)$ *and* $L(z) = \lim_{n\to\infty} L_n(z) \quad \forall z \in \mathbf{C}$ *with* $\Re z \in S_u$

(5.14.3) $\quad$ *If* $s \in S_b$ *and* $L(s) = \lim_{n\to\infty} L_n(s)$, *then* $s \in S_u$

(5.14.4) $\quad Ef(X_n) \to Ef(X) \quad$ *if* $f \in O(e^{ax} + e^{bx})$ *and* $a, b \in S_u$

(5.14.5) $\quad Ef(X_n) \to Ef(X) \quad$ *if* $f \in o(e^{\alpha x} + e^{\beta x})$ *and* $\alpha, \beta \in S_b$

(2): If $\{L_n(s)\}$ converges for all $s \in I$ for some open non-empty interval $I \subseteq S_b$, then there exists a Borel measure μ on $\mathbf{R}$ satisfying

(5.14.6) $\quad \mu(\mathbf{R}) \leq 1 \quad$ *and* $\quad \tilde{\mu}(s) = \lim_{n\to\infty} L_n(s) \quad \forall s \in S_u \setminus \{0\}$

(5.14.7) $\quad \mu(\mathbf{R}) = 1 \quad \Rightarrow \quad X_n \xrightarrow{\sim} \mu$

where $\tilde{\mu}$ denotes the Laplace transform of μ [see (4.19.22)].

(3): Let $I \subseteq S_b$ be a given open non-empty interval, then we have

$$(5.14.8) \qquad X_n \xrightarrow{\sim} X \qquad \Longleftrightarrow \qquad L(s) = \lim_{n\to\infty} L_n(s) \quad \forall s \in I$$

Proof: Since $e^{cx} \in O(e^{ax} + e^{bx})$ for $a \le c \le b$, we see that S_b and S_u are intervals containing 0.

(5.14.1): Since $e^{cx} \in o(e^{ax} + e^{bx})$ for $a < c < b$, we see that $\operatorname{int}(S_b) \subseteq S_u$ and clearly $S_u \subseteq S_b$.

(1): Follows directly from (5.1.13)–(5.1.15) and (5.2).

(2): Let us choose $a \in I \setminus \{0\}$ and assume that $a > 0$ [the case $a < 0$ is treated similarly]. Let $m_n(q)$ be the absolute moment transform of $Y_n = e^{aX_n}$. Then $m_n(q) = L_n(aq)$, and so $\{m_n(q)\}$ converges for q in a neighborhood of 1. Hence, by (5.12) there exists a non-negative random variable Y such that $Y_n \xrightarrow{\sim} Y$. Let $f(x) = \frac{1}{a}\log x$ for $x > 0$ and $f(0) = -\infty$. Then $f : \mathbf{R}_+ \to [-\infty, \infty[$ is continuous and so by (5.4.1) we have that $X_n = f(Y_n) \xrightarrow{\sim} Z$, where $Z = f(Y)$ is a random variable with values in $[-\infty, \infty[$. Let $\mu(A) = P(Z \in A)$ for $A \in \mathcal{B}(\mathbf{R})$. Then μ is a measure on $\mathbf{R}$ with $\mu(\mathbf{R}) = P(Z \ne -\infty) \le 1$. Let $s \in S_u \setminus \{0\}$. Setting $e^{\infty} = \infty$ and $e^{-\infty} = 0$, we see that e^{sx} is continuous on $[-\infty, \infty]$ and $e^{sx} \in \mathcal{M}_u(X_n)$. Since $s \in S_u$, then by (5.2.3) we get

$$\lim_{n\to\infty} L_n(s) = Ee^{sZ} = \int_{-\infty}^{\infty} e^{sx}\mu(ds) + e^{-s\infty}P(Z = -\infty) < \infty$$

and since $s \ne 0$, this is only possible if the last term is 0. Hence, $L_n(s) \to \tilde{\mu}(s)$ and if $\mu(\mathbf{R}) = 1$, then $Z \sim \mu$ and $X_n \xrightarrow{\sim} \mu$.

(3): Suppose that $L_n(s) \to L(s)$ for all $s \in I$. Then by (2) there exists a finite measure μ on $\mathbf{R}$ with $L_n(s) \to \tilde{\mu}(s)$ for all $s \in S_u \setminus \{0\}$; in particular, $\tilde{\mu}(s) = L(s)$ for all $s \in I \setminus \{0\}$. Hence, $\mu = P_X$ by (4.19.21), and so $\mu(\mathbf{R}) = 1$ and $X_n \xrightarrow{\sim} X$ by (2). The converse follows directly from (1). □

5.15 Convergence of Generating Functions

Let X and X_n be non-negative random variables with generating functions $q(t)$ and $q_n(t)$ and let $\gamma :]0,1] \to \mathbf{R}$ be a function such that γ is continuous at 1. Then we have the following:

(5.15.1) *If $\lim_{n\to\infty} q_n(t) = \gamma(t)$ for all $t \in]0,1]$, then there exists a Borel probability measure μ on $[0,\infty[$ such that $X_n \xrightarrow{\sim} \mu$ and*

$$\gamma(t) = \int_0^\infty t^x \, \mu(dx) \qquad \forall\, 0 < t \le 1$$

(5.15.2) $$X_n \xrightarrow{\sim} X \quad \Longleftrightarrow \quad q(t) = \lim_{n\to\infty} q_n(t) \quad \forall\, 0 < t \le 1$$

Proof: (5.15.1): Let $\varepsilon > 0$ be given. Since $\gamma(1) = \lim q_n(1) = 1$ there exists $a \in]0,1[$ such that $\gamma(a) > 1 - \varepsilon$. Let $g(x) = 1 - a^x$. Then $\liminf_{|x|\to\infty} g(x) = 1$ and

$$\limsup_{n\to\infty} Eg(X_n) = \limsup_{n\to\infty} (1 - q_n(a)) = 1 - \gamma(a) < \varepsilon$$

Hence, (5.8.4) holds, and so $\{X_n\}$ is eventually tight by (5.8). But then (2) follows from (4.21.14) and the continuity theorem (5.7).

(5.15.2): If $X_n \xrightarrow{\sim} X$, then $q_n(t) \to q(t)$ because $x \curvearrowright t^x$ is continuous and bounded for all $0 < t \le 1$. The converse follows easily from (5.15.1) and (4.21.14). □

5.16 Preview of the Central Limit Theorem

In (5.1) and (4.22), I told you how Abraham de Moivre stumbled over the first version of the central limit theorem. Let me now finish the story. In the period 1770–1820, Laplace extended the result of de Moivre to the case where X_n takes the values 1 and 0 with probabilities p and $1 - p$ and recognized the limit as the normal distribution, which was discovered a in 1808–1809 by Adrain, Legendre and Gauss. Laplace included his final version of the central limit theorem in the third edition of his book *Théorie Analytique des Probabilités* (1820). General random variables were first introduced by Lvovich Panufty Chebyshev ($\approx$ 1870), and Chebyshev proved the first central limit theorem for random variables in 1887. His original result is as follows:

Let $X_1, X_2, \cdots$ be independent random variables with mean 0 and variances $\sigma_1^2, \sigma_2^2, \cdots$ such that for some $c \in \mathbf{R}_+$, we have $|X_n| \le c$ for all $n \ge 1$. Then

(5.16.1) $$\lim_{n\to\infty} P\Big(s_n^{-1} \sum_{j=1}^n X_j \le a\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}x^2} \, dx$$

where $s_n = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$.

Chebyshev's theorem is **not** correct as it stands; but if $s_n \to \infty$ then (5.16.1) is true. In his proof, Chebyshev does not show (5.16.1) directly, but he shows that all the moments of U_n converge to the moments of the $N(0,1)$-distribution, and he then claims that this will imply (5.16.1). Today, we know that the claim of

Chebyshev is correct [see (5.13) and (5.9)] but that was not known at that time. This incompleteness was noticed by Markov, who was a student of Chebyshev, and he gave a complete proof of Chebyshev's result. Later Lyapounov, also a student of Chebyshev, showed that the boundedness assumption: $|X_n| \leq c$, could be replaced by a much weaker condition, which we today call Lyapounov's condition. Lyapounov introduced the characteristic functions in probability for the purpose of the proof of his central limit theorem. Markov insisted on using the boundedness condition, and in order to treat general (unbounded) random variables he introduced the so-called *truncation technique*, which is still widely used in probability [see Exc.4.71 – Exc.4.75]. In the period 1890–1940, the central limit was extended by many probabilists, for instance, Markov, Lyapounov, Levy, Lindeberg, Khinchine, Feller and Raikov until the final form [see (5.21)] was found.

As you may have guessed by now, a central limit theorem states that under such and such conditions the sum of "many but small" independent random variables is approximately normally distributed. To formalize the idea of sums of "many but small" independent random variables let us consider a *triangular scheme* of random variables:

$$\begin{array}{l} X_{11} \\ X_{21}\,,\, X_{22} \\ X_{31}\,,\, X_{32}\,, X_{33} \\ \vdots \\ X_{n1}\,,\, X_{n2}\,, \ldots\,, X_{nn} \\ \vdots \end{array}$$

We say that the collection $\{X_{nj} \mid 1 \leq j \leq n\}$ is an *independent, triangular scheme* if the random variables in each row are independent; that is, if for each $n \geq 1$, we have that $X_{n1}, \ldots, X_{nn}$ are independent. However, variables in different rows may very well be dependent. We say that the triangular scheme $\{X_{nj} \mid 1 \leq j \leq n\}$ satisfies *the uniform asymptotic negligibility condition*, often abbreviated to *the uan condition* if

$$\text{(5.16.2)} \qquad \lim_{n\to\infty} \Big\{ \max_{1\leq j\leq n} P(|X_{nj}| > \varepsilon) \Big\} = 0 \qquad \forall \varepsilon > 0$$

which means that even the largest variable in the nth row is small when n is large. Let $\{X_{nj} \mid 1 \leq j \leq n\}$ be a given independent, triangular scheme and let $m_{nj} \in \mathbf{R}$ be given numbers. We shall then study the limit distribution of *the centered row sums*:

$$\text{(5.16.3)} \qquad U_n = \sum_{j=1}^{n} (X_{nj} - m_{nj})$$

By a *central limit theorem* we shall understand a result, stating that under such and such conditions we have that $U_n \xrightarrow{\sim} F$, where F is some specified distribution on $\mathbf{R}$. In the classical case, the distribution F is a normal distribution with mean 0 and positive variance. However, as we shall see in Section 5.20, many other distributions may emerge as the limit distribution of U_n .

I shall use the original idea of Lyapounov in the proof of the general central limit theorem [see (5.20)]. The idea is simple: Suppose that $\{X_{nj} \mid 1 \le j \le n\}$ is an independent triangular scheme and let $\{m_{nj} \mid 1 \le j \le n\}$ be given numbers to be determined later. If φ_{nj} and φ_n are the characteristic functions of X_{nj} and U_n , then by (4.18.5+7) we have

$$\varphi_n(t) = \prod_{j=1}^{n} \left(\varphi_{nj}(t) \exp(-itm_{nj})\right) \qquad \forall t \in \mathbf{R} \tag{5.16.4}$$

and by (5.10.2) we know that $U_n \xrightarrow{\sim} F$ for some distribution F if and only if $\varphi_n(t)$ converges to the characteristic function $\varphi(t)$ of F for all $t \in \mathbf{R}$. Thus, we just have to identify the limit of φ_n as the characteristic function of some known distribution, such as the normal distribution. In your elementary calculus course you have certainly met the following result

$$a_n \to a \quad \Rightarrow \quad \left(1 + \frac{a_n}{n}\right)^n \to e^a \tag{5.16.5}$$

whenever $\{a_n\}$ is a sequence of complex numbers. In order to handle the limit of the product in (5.16.4), we need the following extension of (5.16.5):

(5.16.6) Let a_{nj} and b_{nj} be complex numbers for $1 \le j < \infty$ satisfying

$$\lim_{n\to\infty} \sum_{j=1}^{n} |a_{nj} - 1|^2 = 0 \ , \quad \lim_{n\to\infty} \sum_{j=1}^{n} (a_{nj} - 1 - b_{nj}) = a$$

for some complex number $a \in \mathbf{C}$. Then we have

$$\lim_{n\to\infty} \prod_{j=1}^{n} \left(a_{nj} \exp(-b_{nj})\right) = e^a$$

[The proof is straightforward, and you can find an outline of the proof in Exc.5.25.] Note that we obtain (5.16.5) from (5.16.6) by setting $b_{nj} = 0$ and $a_{nj} = 1 + \frac{1}{n} a_n$. In (5.20) you will see how to use (5.16.4) and (5.16.6) to obtain the limit distribution of U_n .

The proof of (5.20) is based on the simple idea above but it is easily lost in technicalities. So let me first show the central limit theorem in a simple, but important case, which clearly demonstrates the idea. Suppose that $X_1, X_2, \cdots$ are independent identically distributed random variables with mean 0 and variance

$0 \le \sigma^2 < \infty$. Let φ be the common characteristic function of $X_1, X_2, \ldots$ and let φ_n be the characteristic function of $U_n = (X_1 + \cdots + X_n)/\sqrt{n}$. Then $\varphi(0) = 1$, $\varphi'(0) = 0$ and $\varphi''(0) = -\sigma^2$ by (4.18.13). Hence, by l'Hospital's rule we have

$$\lim_{s \to 0} \frac{\varphi(ts) - 1}{s^2} = \lim_{s \to 0} \frac{t^2 \varphi''(ts)}{2} = -\frac{t^2 \sigma^2}{2}$$

Setting $s = n^{-1/2}$, we have $n(\varphi(\frac{t}{\sqrt{n}}) - 1) \to -\frac{1}{2} t^2 \sigma^2$ for all $t \in \mathbf{R}$ and by (4.18.5+7) and (5.16.5), we get

$$\varphi_n(t) = \varphi\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 + \frac{n(\varphi(\frac{t}{\sqrt{n}}) - 1)}{n}\right)^n \to e^{-\frac{1}{2} t^2 \sigma^2}$$

Thus, by (4.22.3) we have that the limit of φ_n is the characteristic function of $N(0, \sigma^2)$. Hence, by (5.10.2) we get the following version of *the normal central limit theorem*:

(5.16.7) If $\{X_n\}$ are independent, identically distributed random variables with mean 0 and variance $0 \le \sigma^2 < \infty$, then we have

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j \xrightarrow{\sim} N(0, \sigma^2)$$

Setting $m_{nj} = 0$ and $X_{nj} = n^{-\frac{1}{2}} X_j$ we see that $U_n = (X_1 + \cdots + X_n)/\sqrt{n}$ equals the U_n defined in (5.16.3), and that $\{X_{nj} \mid 1 \le j \le n\}$ is an independent, triangular scheme satisfying the uan condition (5.16.2).

At this point it is natural to ask if the normal distribution really is the only distribution which can occur as the limit distribution of the centered row sums is of an independent, infinitesimal, triangular scheme? The answer to this question is that this is a property which the normal distribution shares with many, but not all, other distributions, such as the Poisson distributions, the negative binomial distributions, the gamma distributions, the stable distributions, but not the binomial distributions, nor the beta distributions, nor the rectangular distributions. In the next section I shall introduce the class of distributions on $\mathbf{R}$ having this property.

5.17 The Lévy-Khinchine Distributions

Let ν be a Borel measure on $\mathbf{R}$ such that $0 < \nu(\mathbf{R}) < \infty$ and set $\lambda = \nu(\mathbf{R})$. Then $\pi(B) = \lambda^{-1} \nu(B)$ is a probability measure on $\mathbf{R}$. Let $\xi_1, \xi_2, \cdots$ be independent identically distributed random variables with distribution π and let

N be a $Po(\lambda)$-distributed random variable which is independent of $\{\xi_n\}$. Then by Exc.5.26 we have that $U = \sum_{1\le j\le N} \xi_j$ has characteristic functions $\varphi_U(t)$ given by

$$\varphi_U(t) = \exp\left\{\int_{-\infty}^{\infty} (e^{itx} - 1)\,\nu_0(dx)\right\} \qquad \forall t \in \mathbf{R}$$

where $\nu_0(B) = \nu(B \setminus \{0\})$ for all $B \in \mathcal{B}(\mathbf{R})$. More generally, suppose that ν is a Borel measure on $\mathbf{R}$ satisfying

$$\text{(5.17.1)} \qquad \nu(\{0\}) = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} (1 \wedge |x|)\,\nu(dx) < \infty$$

Since $|e^{itx} - 1| \le (2 + |t|)\,(1 \wedge |x|)$, we may define the function φ as follows:

$$\text{(5.17.2)} \qquad \varphi(t) = \exp\left\{\int_{-\infty}^{\infty} (e^{itx} - 1)\,\nu(dx)\right\} \qquad \forall t \in \mathbf{R}$$

and by Lebesgue's dominated convergence theorem (3.6), we have that φ is continuous. Setting $J_n = \{x \mid |x| > \frac{1}{n}\}$ and $\nu_n(B) = \nu(B \cap J_n)$, we see that ν_n is a finite measure on $\mathbf{R}$. From the argument above, we have that

$$\varphi_n(t) = \exp\left\{\int_{J_n} (e^{itx} - 1)\,\nu(dx)\right\}$$

is a characteristic function such that $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbf{R}$. Hence, by (5.10.1) we see that the function φ defined by (5.17.1) is a characteristic function for some distribution on $\mathbf{R}$, whenever ν is a Borel measure on $\mathbf{R}$ satisfying (5.17.2).

A distribution having characteristic function φ given by (5.17.2) for some Borel measure ν on $\mathbf{R}$ satisfying (5.17.1) is called a *compound Poisson distribution* and is denoted $Po[\nu]$. A straightforward computation shows that we have

$$\text{(5.17.3)} \qquad Po(\lambda) = Po[\lambda\,\delta_1] \quad \text{where } \delta_1 \text{ is the Dirac measure [see (1.15)]}$$

$$\text{(5.17.4)} \qquad Bi^-(\alpha, p) = Po[\nu] \quad \text{where} \quad \nu(B) = \sum_{n\in B\cap\mathbf{N}} \frac{\alpha p^n}{n}$$

$$\text{(5.17.5)} \qquad \Gamma(\alpha, \beta) = Po[\nu] \quad \text{where} \quad \nu(B) = \int_{B\cap[0,\infty[} \frac{\alpha e^{-x/\beta}}{x}\,dx$$

Thus, the Poisson distribution $Po(\lambda)$, the negative binomial distribution $Bi^-(\alpha,p)$ and the gamma distribution $\Gamma(\alpha,\beta)$ are all compound Poisson distributions.

Let ν be a Borel measure on $\mathbf{R}$ and let $q\colon \mathbf{R} \to \mathbf{R}$ be a Borel function. Then we say that (q,ν) is a *Lévy–Khinchine pair* if

$$(5.17.6) \qquad \nu(\{0\}) = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} (1 \wedge x^2)\, \nu(dx) < \infty$$

$$(5.17.7) \qquad q \in \mathcal{C}(\nu) \quad , \quad q(0) = 0 = \lim_{x\to 0} q(x) \quad , \quad \sup_{|x|\le a} |q(x)| < \infty \quad \forall\, a \in \mathbf{R}_+$$

$$(5.17.8) \qquad \int_I |x - q(x)|\, \nu(dx) < \infty \qquad \text{and} \qquad \int_{\mathbf{R}\setminus I} |q(x)|\, \nu(dx) < \infty$$

where $I = [-1,1]$ and $\mathcal{C}(\nu)$ is the set of all functions from $\mathbf{R}$ into $\mathbf{R}$ which are continuous at ν-a.a. points. A Borel measure ν on $\mathbf{R}$ satisfying (5.17.6) is called a *Lévy measure*. Suppose that (q,ν) is a Lévy-Khinchine pair and set $h(t,x) = e^{itx} - 1 - it\,q(x)$. Then it is easily seen that

$$|h(t,x)| \le 2 + |t|\,|q(x)| \quad , \quad |h(t,x)| \le |t|\,|x - q(x)| + \tfrac{1}{2}\, t^2\, x^2$$

Hence, by (5.17.6) and (5.17.8) we see that $h(t,\cdot) \in L^1(\nu)$ for all $t \in \mathbf{R}$. Thus, we may define

$$(5.17.9) \qquad \varphi(t) = \exp\left\{ it\mu - \frac{t^2\sigma^2}{2} + \int_{-\infty}^{\infty} (e^{itx} - 1 - itq(x))\, \nu(dx) \right\}$$

whenever $\mu \in \mathbf{R}$, $\sigma \in [0,\infty[$ and (q,ν) is a Lévy-Khinchine pair. Exactly as above one shows that φ is the characteristic function of a distribution on $\mathbf{R}$. This distribution is called a *Lévy-Khinchine distribution* and it is denoted $Lev(\mu,\sigma,q,\nu)$. Note that ν satisfies (5.17.2) if and only if $(0,\nu)$ is a Lévy-Khinchine pair; and if so, then

$$(5.17.10) \qquad Po[\nu] \;=\; Lev(0,0,0,\nu)$$

Thus, every compound Poisson distribution is a Lévy-Khinchine distribution. If $q : \mathbf{R} \to \mathbf{R}$ is a Borel function with $q(0) = \lim_{x\to\infty} q(x) = 0$ and $\mathbf{0}$ denotes the measure which is identically equal to zero, then by (4.22.3) we have

$$(5.17.11) \qquad N(\mu,\sigma^2) \;=\; Lev(\mu,\sigma,q,\mathbf{0})$$

Moreover, it is easily checked, that we have the following:

(5.17.12) Let ν be a Lévy measure, and let $q : \mathbf{R} \to \mathbf{R}$ be a Borel function satisfying (5.17.7) and

$$\limsup_{x \to 0} \frac{|x - q(x)|}{x^2} < \infty \quad , \quad \int_{\mathbf{R} \setminus [-a,a]} |q(x)| \, \nu(dx) < \infty$$

for some $0 < a < \infty$, then (q, ν) is a Lévy-Khinchine pair

(5.17.13) Let ν be a Lévy measure, and let $q_1, q_2 : \mathbf{R} \to \mathbf{R}$ be Borel functions satisfying (5.17.7) such that $q_1 - q_2 \in L^1(\nu)$ and (q_1, ν) is a Lévy-Khinchine pair. Then (q_2, ν) is a Lévy-Khinchine pair and $Lev(\mu_1, \sigma, q_1, \nu) = Lev(\mu_2, \sigma, q_2, \nu)$ where

$$\mu_2 = \mu_1 + \int_{-\infty}^{\infty} (q_2(x) - q_1(x)) \, \nu(dx)$$

For instance, if ν is a Lévy measure, and q is any of the functions

$$q(x) = \frac{x}{1 + x^2} \quad , \quad q(x) = \sin x$$
$$q(x) = \arctan x \quad , \quad q(x) = x \, 1_{[-a,a]}(x) \quad (a > 0 \ , \ \nu(\{\pm a\}) = 0)$$

then (q, ν) is a Lévy-Khinchine pair. Note that (5.17.13) tells us that the representation $Lev(\mu, \sigma, q, \nu)$ is **not** unique. However, if we fix the function q (for instance, to be one of the four functions mentioned above), then the representation is unique [see Exc.5.16].

5.18 The Stable Distribution $Stab(\mu, \sigma, p, \gamma)$

I shall now introduce an important subclass of the Lévy-Khinchine distributions called the *stable distributions*. Let $\mu \in \mathbf{R}$, $\sigma \in [0, \infty[$, $0 < p < 2$ and $-1 \leq \gamma \leq 1$ be given numbers and set

(5.18.1) $$\nu_p(B) = a \int_{B \cap [0,\infty[} x^{-p-1} dx + b \int_{B \cap]-\infty,0]} |x|^{-p-1} dx$$

(5.18.2) $$q_p(x) = \begin{cases} 0 & \text{if } 0 < p < 1 \\ \sin x & \text{if } p = 1 \\ x & \text{if } 1 < p < 2 \end{cases}$$

where a and b are given by [see (3.37.14+15)]

$$a = \frac{\sigma^p (1-\gamma)}{2\,C(p)} \quad , \quad b = \frac{\sigma^p (1+\gamma)}{2\,C(p)} \quad \text{if } 0 < p < 2$$

$$C(p) = \begin{cases} \dfrac{\Gamma(2-p)\cos\frac{p\pi}{2}}{p(1-p)} & \text{if } p \neq 1\,,\ 0 < p < 2 \\ \frac{\pi}{2} & \text{if } p = 1 \end{cases}$$

$$\mathcal{K}(p,t) = \begin{cases} (\operatorname{sign} t)\tan\frac{p\pi}{2} & \text{if } p \neq 1\,,\ 0 < p \leq 2 \\ \frac{2}{\pi}(\operatorname{sign} t)\log|t| & \text{if } p = 1 \end{cases}$$

Then (q_p, ν_p) is a Lévy-Khinchine pair and by (3.37.9)–(3.37.17), we have that the characteristic function φ of $Lev(\mu, 0, q_p, \nu_p)$ is given by

(5.18.3) $$\varphi(t) = \exp\left\{i\mu t - |\sigma t|^p(1 + i\gamma\mathcal{K}(p,t))\right\}$$

A distribution having characteristic function φ given by (5.18.3) for some $\mu \in \mathbf{R}$, some $\sigma \in [0,\infty[$, some $p \in]0,2]$ and some $\gamma \in [-1,1]$, is called a *stable distribution* and it is denoted $Stab\,(\mu,\sigma,p,\gamma)$. Since $\mathcal{K}(2,t) = 0$ for all $t \in \mathbf{R}$ and $q_p(x) = 0$ for $0 < p < 1$ we find

(5.18.4) $$Stab(\mu,\sigma,2,\gamma) = N(\mu, 2\sigma^2) \qquad \forall\,\mu \in \mathbf{R}\ \forall\sigma \geq 0\ \forall -1 \leq \gamma \leq 1$$

(5.18.5) $$Stab(a,b,1,0) = C(a,b) \qquad \forall\,a \in \mathbf{R}\ \forall b > 0$$

(5.18.6) $$Stab(0,\sigma,p,\gamma) = Po[\nu_p] \qquad \forall\,0 < p < 1\ \forall -1 \leq \gamma \leq 1$$

(5.18.7) $$Stab(\mu,0,p,\gamma) = \delta_\mu \qquad \forall\,\mu \in \mathbf{R}\ \forall 0 < p \leq 2\ \ \forall -1 \leq \gamma \leq 1$$

where $C(a,b)$ is the Cauchy distribution and δ_μ is the Dirac measure [see (1.15)].

Let $\mathcal{K}(p,t)$ be the function defined in (3.37.15). Then evidently, we have that $\mathcal{K}(1,at) = \operatorname{sign}(at)\log|a| + \operatorname{sign}(a)\,\mathcal{K}(1,t)$ and $\mathcal{K}(p,at) = \operatorname{sign}(a)\,\mathcal{K}(p,t)$ for $p \neq 1$. Hence, by (4.18.5) and (5.18.3) we get

(5.18.8) If $a, b \in \mathbf{R}$ are real numbers and X is a random variable with $X \sim Stab(\mu,\sigma,p,\gamma)$, then $aX + b \sim Stab(\tilde{\mu},\tilde{\sigma},p,\tilde{\gamma})$ where $\tilde{\sigma} = \sigma\,|a|$, $\tilde{\gamma} = \gamma\operatorname{sign}(a)$ and

$$\tilde{\mu} = \begin{cases} a\mu + b & \text{if } p \neq 1 \\ a\mu + b - \frac{2\gamma\sigma}{\pi}\,a\log|a| & \text{if } p = 1 \end{cases}$$

Similarly, by (4.18.7) and (5.18.3) we get

(5.18.9) If X_1 and X_2 are independent random variables such that $X_j \sim Stab(\mu_j, \sigma_j, p, \gamma_j)$ for $j = 1,2$, then we have that $X_1 + X_2 \sim Stab(\mu, \sigma, p, \gamma)$ where

$$\mu = \mu_1 + \mu_2 \ , \ \sigma = ({\sigma_1}^p + {\sigma_2}^p)^{1/p} \ , \ \gamma = \frac{\gamma_1 {\sigma_1}^p + \gamma_2 {\sigma_2}^p}{{\sigma_1}^p + {\sigma_2}^p}$$

5.19 The Upper and Lower Tail Functionals

Let $\{X_{nj} \mid 1 \le j \le n\}$ be a triangular scheme and let $f: \mathbf{R} \to \mathbf{C}$ be a Borel function. Then we define the lower and upper tail functionals of $\{X_{nj}\}$ as follows:

(5.19.1) $$\rho_0(f) = \lim_{\delta \to 0} \left\{ \limsup_{n \to \infty} \sum_{j=1}^{n} \int_{\{|X_{nj}| < \delta\}} |f(X_{nj})| \, dP \right\}$$

(5.19.2) $$\rho_\infty(f) = \lim_{a \to \infty} \left\{ \limsup_{n \to \infty} \sum_{j=1}^{n} \int_{\{|X_{nj}| > a\}} |f(X_{nj})| \, dP \right\}$$

Let $f, g : \mathbf{R} \to \mathbf{R}$ be two Borel functions, let $a \in \mathbf{C}$ be a complex number, and let $k \in \mathbf{N}$ be an integer. Then we have the following:

(5.19.3) $$\rho_0(af) = |a| \, \rho_0(f) \quad , \quad \rho_0(f+g) \le \rho_0(f) + \rho_0(g)$$

(5.19.4) $$\rho_\infty(af) = |a| \, \rho_\infty(f) \quad , \quad \rho_\infty(f+g) \le \rho_\infty(f) + \rho_\infty(g)$$

(5.19.5) $$\rho_0(f \cdot g) \le \rho_0(g) \limsup_{x \to 0} |f(x)|$$

(5.19.6) $$\rho_\infty(f \cdot g) \le \rho_\infty(g) \limsup_{|x| \to \infty} |f(x)|$$

(5.19.7) *If $\rho_\infty(\mathbf{1}) = 0$ and g is bounded, then $\rho_\infty(f+g) = \rho_\infty(f)$ and $\rho_\infty(g) = 0$*

where $\mathbf{1}$ is the function which is identically equal to 1, and where we in (5.19.5+6) use the **unusual** *convention: $0 \cdot \infty = \infty$. If f and g are bounded on $[-\delta, \delta]$ for some $\delta > 0$, then we have*

(5.19.8) $$\rho_0(g - f) = 0 \quad \Rightarrow \quad \rho_0(f^k) = \rho_0(g^k) \quad \forall k = 1, 2, \ldots$$

Moreover, if $\{X_{nj}\}$ satisfies the uan condition (5.16.2) and the following condition:

$$\sup_{n\geq 1}\left\{\sum_{j=1}^{n} P(|X_{nj}|>\delta)\right\}<\infty \qquad \forall\,\delta>0 \tag{5.19.9}$$

and if f is bounded on all compact intervals, then we have

$$\limsup_{n\to\infty}\left\{\max_{1\leq j\leq n} E|f(X_{nj})|\right\}\leq \rho_\infty(f)+\limsup_{x\to 0}|f(x)| \tag{5.19.10}$$

$$\limsup_{n\to\infty}\left\{\sum_{j=1}^{n}\left(E|f(X_{nj})|\right)^2\right\}\leq 3(\rho_0(|f|^2)+\rho_\infty(f)^2) \tag{5.19.11}$$

Proof: Propositions (5.19.3)–(5.19.7) follows easily from the definition of the upper and lower functionals.

(5.19.8): If $k=1$, then (5.19.8) follows from (5.19.3). So suppose that $k\geq 2$ then $f^k-g^k=h(f-g)$ where $h=f^{k-1}+f^{k-2}g+\cdots+fg^{k-2}+g^{k-1}$. By assumption, we have that $\limsup_{x\to 0} h(x)<\infty$. Hence, by (5.19.5) we see that $\rho_0(f^k-g^k)=0$ and so (5.19.8) follows from (5.19.3).

(5.19.10+11): Suppose that f is bounded on all compact intervals and let $0<\delta<a<\infty$ be given. Then there exists $0<k<\infty$ such that $|f(x)|\leq k$ for all $|x|\leq a$. Hence, if $Y_{nj}=|f(X_{nj})|$, then we get

$$\begin{aligned} EY_{nj} &= \int_{\{|X_{nj}|\leq\delta\}} Y_{nj}\,dP+\int_{\{\delta<|X_{nj}|\leq a\}} Y_{nj}\,dP+\int_{\{|X_{nj}|>a\}} Y_{nj}\,dP \\ &\leq \sup_{|x|\leq\delta}|f(x)|+k\,P(|X_{nj}|>\delta)+\int_{\{|X_{nj}|>a\}}|f(X_{nj})|\,dP \end{aligned}$$

Thus, taking maximum over $j=1,\ldots,n$ and letting $n\to\infty$, $\delta\to 0$ and $a\to\infty$, we see that (5.19.10) follows from (5.16.2) and the definition of ρ_∞. If we set

$$\begin{aligned} A_{nj} &= \int_{\{|X_{nj}|<\delta\}}|f(X_{nj})|^2\,dP \quad , \quad B_{nj}=k^2\,P(|X_{nj}|>\delta)^2 \\ C_{nj} &= \int_{\{|X_{nj}|>a\}}|f(X_{nj})|\,dP \end{aligned}$$

then splitting the integral $E|f(X_{nj})|$ as above and applying (4.3.9) with $r=1$ and $q=2$, we find

$$E|f(X_{nj})|=E\big(1_{\{|X_{nj}|\leq\delta\}}Y_{nj}\big)+\sqrt{B_{nj}}+C_{nj}\leq\sqrt{A_{jn}}+\sqrt{B_{nj}}+C_{nj}$$

Note that $x \curvearrowright x^2$ is convex and $x \curvearrowright x^{1/2}$ is concave; hence, we have

$$|x+y+z|^2 \le 3(|x|^2+|y|^2+|z|^2)$$
$$|x_1|^2+\cdots+|x_n|^2 \le (|x_1|+\cdots+|x_n|)^2$$

Thus, we get

$$\sum_{j=1}^{n}(E|f(X_{nj})|)^2 \le 3\left\{\sum_{j=1}^{n}A_{nj}+\sum_{j=1}^{n}B_{nj}+\Big(\sum_{j=1}^{n}C_{nj}\Big)^2\right\}$$
$$\sum_{j=1}^{n}B_{nj} \le k^2 \max_{1\le j\le n} P(|X_{nj}|>\delta)\left\{\sum_{j=1}^{n}P(|X_{nj}|>\delta)\right\}$$

By definition of ρ_∞ and ρ_0, we have:

$$\rho_0(f^2)=\lim_{\delta\to 0}\left\{\limsup_{n\to\infty}\sum_{j=1}^{n}A_{nj}\right\} \quad , \quad \rho_\infty(f)=\lim_{a\to\infty}\left\{\limsup_{n\to\infty}\sum_{j=1}^{n}C_{nj}\right\}$$

and by (5.16.2) and (5.19.9), we have:

$$\lim_{\delta\to 0}\left\{\limsup_{n\to\infty}\sum_{j=1}^{n}B_{nj}\right\}=0$$

Thus, (5.19.11) follows. □

5.20 The General Central Limit Theorem

Let (q,ν) be a Lévy-Khinchine pair and let $\{X_{nj} \mid 1\le j\le n\}$ be an independent, triangular scheme satisfying the uan condition (5.16.2). Let $w:\mathbf{R}\to[0,\infty[$ be a Borel function such that w is continuous ν-a.s. and w is bounded on all bounded intervals. Suppose that $m_{nj}=Eq(X_{nj})$ exists and is finite for all $1\le j\le n$ and that D_1 and D_2 are a dense subsets of $]0,\infty[$ such that

(5.20.1) $$\lim_{n\to\infty}\sum_{j=1}^{n}P(X_{nj}>x)=\nu(]x,\infty[) \qquad \forall x\in D_1$$

(5.20.2) $$\lim_{n\to\infty}\sum_{j=1}^{n}P(X_{nj}<-x)=\nu(]-\infty,-x[) \qquad \forall x\in D_2$$

(5.20.3) $$\rho_\infty(q)=\rho_\infty(w)=\rho_0(\tilde{q})=\rho_0(\tilde{w})=0$$

(5.20.4) $\alpha = \lim_{n\to\infty} \sum_{j=1}^{n} Ew(X_{nj})$ *exists and is finite*

(5.20.5) $\lim_{n\to\infty} \sum_{j=1}^{n} m_{nj}^2 = 0$

where $\tilde{q}(x) = x - q(x)$, $\tilde{w}(x) = x^2 - w(x)$ *and* ρ_0 *and* ρ_∞ *are the lower and upper functionals [see (5.19.1+2)]. Then* $w \in L^1(\nu)$, *and if* $\sigma = \sqrt{\rho_0(w)}$, *then we have*

(5.20.6) $\alpha = \sigma^2 + \int_{-\infty}^{\infty} w(x)\,\nu(dx)$

(5.20.7) $U_n = \sum_{j=1}^{n} (X_{nj} - m_{nj}) \xrightarrow{\sim} Lev(0, \sigma, q, \nu)$

Remark: Note that (5.20.1+2) implies (5.19.9). Hence, by (5.19.11) we have that (5.20.1+2) and $\rho_\infty(q) = \rho_0(q^2) = 0$ implies (5.20.5). Moreover, by (5.19.8) we have that $\rho_0(\tilde{q}) = \rho_0(\tilde{w}) = \rho_0(w) = 0$ implies $\rho_0(q^2) = 0$.

Proof: Let us define

$$\nu_n(B) = \sum_{j=1}^{n} P(X_{nj} \in B) \quad \forall B \in \mathcal{B}(\mathbf{R})$$

$$\mathcal{W} = \{ 1_{]a,b]} \mid a, b \in D_1 \} \cup \{ 1_{[-a,-b[} \mid a, b \in D_2 \}$$

$$\mathcal{H} = \{ 1_{]a,\infty[} + 1_{]-\infty,-b[} \mid a \in D_1 \ , \ b \in D_2 \}$$

then ν_n is a finite Borel measure on $\mathbf{R}$ satisfying (5.3.1+2) and by the standard proof (3.14.4), we have

$$\int_{-\infty}^{\infty} f\, d\nu_n = \sum_{j=1}^{n} Ef(X_{nj}) \quad \text{in the usual sense}$$

Let G be an open subset of $\mathbf{R}$, then $G \setminus \{0\}$ is a disjoint union of open intervals not containing 0 . Since $\nu(\{0\}) = 0$ and every open interval not containing 0 is an increasing union of intervals of the form used in the definition of $\mathcal{W}$, we conclude that (5.3.3) holds. Let

$$\mathcal{L} := \{ f \in \mathcal{C}_0(\nu) \mid \rho_0(f) < \infty \ , \ \rho_\infty(f) = 0 \} \ , \ \mathcal{L}_0 := \{ f \in \mathcal{L} \mid \rho_0(f) = 0 \}$$

where $C_0(\nu)$ is the set of all functions $f \in C(\nu)$ such that f is bounded on all compact intervals. Let $f \in \mathcal{L}$ and $\varepsilon > 0$ be given. Since $\rho_\infty(f) = 0$, there exists $a > 0$ such that

$$\limsup_{n\to\infty}\left\{\int_{\{|f|>b\}} |f|\,d\nu_n\right\} \le \limsup_{n\to\infty}\left\{\sum_{j=1}^{n}\int_{\{|X_{nj}|>a\}} |f(X_{nj})|\,dP\right\} < \varepsilon$$

where $b = \sup\{|f(x)| \mid |x| \le a\} < \infty$. Thus, we conclude that $\mathcal{L} \subseteq \mathcal{M}_u(\nu_n)$. Since $R(f) = \rho_0(f)$ where $R(f)$ is given by (5.3.4), then by (5.3.7+8) we get

(i) $$\lim_{n\to\infty}\sum_{j=1}^{n} Ef(X_{nj}) = \int_{-\infty}^{\infty} f\,d\nu \quad \forall f \in \mathcal{L}_0$$

(ii) $$\limsup_{n\to\infty}\sum_{j=1}^{n} E|f(X_{nj})| = \rho_0(f) + \int_{-\infty}^{\infty} |f|\,d\nu < \infty \quad \forall f \in \mathcal{L}$$

(5.20.6): By (5.20.3+4), we have that $w \in \mathcal{L}$. Hence, $w \in L^1(\nu)$ and (5.20.6) follows from (ii), (5.20.4) and non-negativity of w.

(5.20.7): Let φ_{nj} and φ_n be the characteristic functions of X_{nj} and U_n, respectively. Let $t \in \mathbf{R}$ be fixed for the rest of this proof and set

$$h(x) = e^{itx} - 1 - itq(x) \quad , \quad g(x) = h(x) + \tfrac{1}{2}t^2\,w(x)$$

Then $h, g \in C_0(\nu)$ and since $\nu(x \mid |x| > a) \to 0$ as $a \to \infty$, then by (5.20.1+2) we see that $\rho_\infty(\mathbf{1}) = 0$. Hence, by (5.19.7) and (5.20.3) we have that $\rho_\infty(h) = \rho_\infty(g) = 0$. Recalling that $|e^{iy} - 1 - iy| \le \frac{1}{2}|y|^2$ and $x^2 = w(x) + \tilde{w}(x)$, we get

$$|h(x)| \le \tfrac{1}{2}t^2\,(w(x) + \tilde{w}(x)) + |t|\,|\tilde{q}(x)|$$

Hence, by (5.19.3) and (5.20.3) we see that $\rho_0(h) \le \frac{1}{2}t^2\,\rho_0(w) < \infty$ and since $h(x) \to 0$ as $x \to 0$, then by (5.19.5) we have $\rho_0(h^2) = 0$. Note that $Eh(X_{nj}) = \varphi_{nj}(t) - 1 - it\,m_{nj}$. Thus, using (5.19.11), (5.20.5) and the well-known inequality $|u+v|^2 \le 2\,(|u|^2 + |v|^2)$, we find

$$\limsup_{n\to\infty}\sum_{j=1}^{n} |\varphi_{nj}(t) - 1|^2 \le 2\lim_{n\to\infty}\sum_{j=1}^{n}\left(|it\,m_{nj}|^2 + (E|h(X_{nj})|)^2\right) = 0$$

Recalling that $|e^{iy} - 1 - iy + \frac{1}{2}y^2| \le \frac{1}{6}|y|^3$ for all $y \in \mathbf{R}$, we get

$$|g(x)| \le \tfrac{1}{6}|t|^3\,|x|^3 + \tfrac{1}{2}t^2|\,\tilde{w}(x)| + |t|\,|\tilde{q}(x)|$$

and as above we conclude that $\rho_0(g) = 0$. Hence, $g \in \mathcal{L}_0$ and since $h(x) = g(x) - \frac{1}{2}t^2 w(x)$, then by (5.20.4), (5.20.6) and (i) we find

$$\begin{aligned}
-\tfrac{1}{2}\sigma^2 t^2 + \int_{\mathbf{R}} h\,d\nu &= -\tfrac{1}{2}\sigma^2 t^2 + \int_{\mathbf{R}} (e^{it} - 1 - it\,q(x))\,d\nu \\
&= -\tfrac{1}{2}\alpha t^2 + \int_{\mathbf{R}} g\,d\nu \\
&= \lim_{n\to\infty} \left\{ \sum_{j=1}^{n} Eg(X_{nj}) - \tfrac{1}{2}t^2 \sum_{j=1}^{n} Ew(X_{nj}) \right\} \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} Eh(X_{nj}) \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} (\varphi_{nj}(t) - 1 - it\,m_{nj})
\end{aligned}$$

Thus, (5.20.7) follows from (5.16.6), (5.10.2) and (5.17.9). □

5.21 The Normal Central Limit Theorem I

Let $\{X_{nj} \mid 1 \le j \le n\}$ *be an independent triangular scheme, and let* $q : \mathbf{R} \to \mathbf{R}$ *be a Borel function such that* $m_{nj} = Eq(X_{nj})$ *and* $\sigma_{nj}^2 = \operatorname{var} q(X_{nj})$ *exists and are finite for all* $1 \le j \le n$ *and*

(5.21.1) $$\lim_{n\to\infty} \sum_{j=1}^{n} m_{n_j}^2 = 0 \quad \textit{and} \quad \lim_{n\to\infty} \sum_{j=1}^{n} \sigma_{nj}^2 = \sigma^2$$

(5.21.2) $$q(0) = 0 = \lim_{x\to 0} \frac{x - q(x)}{x^2} \quad \textit{and} \quad \sup_{|x|\le a} |q(x)| < \infty \qquad \forall\, a \in \mathbf{R}_+$$

(5.21.3) $$\lim_{n\to\infty} \left\{ \sum_{j=1}^{n} \int_{\{|X_{nj}|>\varepsilon\}} (1 + q(X_{nj})^2)\,dP \right\} = 0 \qquad \forall\, \varepsilon > 0$$

for some $0 \le \sigma < \infty$. *Then we have*

(5.21.4) $$U_n = \sum_{j=1}^{n} (X_{nj} - m_{nj}) \xrightarrow{\sim} N(0, \sigma^2)$$

Proof: Let $\nu \equiv 0$ and $w(x) = q(x)^2$. Then $Ew(X_{nj}) = m_{nj}^2 + \sigma_{nj}^2$. Hence, by (5.21.1) and (5.21.3) we see that $\{X_{nj}\}$ satisfies (5.16.2), (5.20.1+2) and (5.20.4+5) with $D_1 = D_2 =]0, \infty[$ and $\alpha = \sigma^2$. By (5.17.11+12), we have that (q, ν) is a Lévy–Khinchine pair with $N(0, \sigma^2) = Lev(0, \sigma, q, \nu)$. By (5.21.3), we see that $\rho_\infty(q) = \rho_\infty(w) = 0$. Moreover, by (5.21.2) it follows easily that

$$\lim_{x \to 0} \frac{x - q(x)}{w(x)} = \lim_{x \to 0} \frac{x^2 - w(x)}{w(x)} = 0$$

and since $\rho_0(w) < \infty$, then by (5.19.5) we have $\rho_0(\tilde{w}) = \rho_0(\tilde{q}) = 0$ where $\tilde{w}(x) = x^2 - w(x)$ and $\tilde{q}(x) = x - q(x)$. Thus, (5.21) follows from (5.20). □

Remarks: (1): Suppose that q is twice differentiable at 0 [see Exc.5.15] with $q(0) = q''(0) = 0$ and $q'(0) = 1$. Then by l'Hospital's rule [see Exc.5.15], we have that q satisfies (5.21.2). In applications of the normal central limit theorem you should first decide which function q to use. The following five functions, which all satisfy (5.21.2), seem to be the most popular

$$q(x) = x \quad , \quad q(x) = \frac{x}{1 + x^2} \quad , \quad q(x) = \sin x$$
$$q(x) = \arctan x \quad , \quad q(x) = x\, 1_{[-a,a]}(x) \qquad (a > 0)$$

But of course many other choices are possible.

(2): Having chosen the function q the crucial point is to verify (5.21.3). To do this various conditions have been invented and let me here mention the three most important, called *Raikov's*, *Lindeberg's* and *Lyapounov's condition*, respectively:

(Raikov) $$\lim_{n \to \infty} \left\{ \sum_{j=1}^{n} P(|X_{nj}| > \varepsilon) \right\} = 0 \qquad \forall \varepsilon > 0$$

(Lindeberg) $$\lim_{n \to \infty} \sum_{j=1}^{n} \int_{\{|X_{nj}| > \varepsilon\}} q(X_{nj})^2 \, dP = 0 \qquad \forall \varepsilon > 0$$

(Lyapounov) $$\exists \alpha > 2 \ : \ \lim_{n \to \infty} \sum_{j=1}^{n} E|q(X_{nj})|^\alpha = 0$$

Evidently, we have

(5.21.5) $\quad$ (Raikov) + (Lindeberg) $\iff$ (9.21.3)

Let $m(\varepsilon) = \inf_{|x|>\varepsilon} |q(x)|$ and $M = \sup_{x\in\mathbf{R}} |q(x)|$. Then we have

$$m(\varepsilon)^2\, P(|X| > \varepsilon) \le \int_{\{|X|>\varepsilon\}} q(X)^2\, dP \le M^2\, P(|X| > \varepsilon)$$

$$\int_{\{|X|>\varepsilon\}} q(X)^2\, dP \le m(\varepsilon)^{2-\alpha}\, E|q(X)|^\alpha \qquad \forall\, \alpha > 2$$

Hence we get

(5.21.6) If $\inf\limits_{|x|>\varepsilon} |q(x)| > 0$ for all $\varepsilon > 0$, then we have

$$\text{(Lyapounov)} \Rightarrow \text{(Lindeberg)} \Rightarrow \text{(Raikov)} + (9.21.3)$$

(5.21.7) If q is bounded, then we have

$$\text{(Raikov)} \Rightarrow \text{(Lindeberg)} + (9.21.3)$$

5.22 The Normal Central Limit Theorem II

Let $\{X_{nj} \mid 1 \le j \le n\}$ *be an independent, triangular scheme such that* $\mu_{nj} = EX_{nj}$ *and* $\sigma_{nj}^2 = \operatorname{var} X_{nj}$ *exists and are finite for all* $1 \le j \le n$. *Set* $s_n = \sqrt{\sigma_{n1}^2 + \cdots + \sigma_{nn}^2}$ *and suppose that* $s_n > 0$ *for all* $n \ge 1$. *If Lindeberg's condition:*

$$\text{(5.22.1)} \qquad \lim_{n\to\infty} \left\{ s_n^{-2} \sum_{j=1}^{n} \int_{\{|X_{nj}-\mu_{nj}|>\varepsilon s_n\}} (X_{nj} - \mu_{nj})^2\, dP \right\} = 0 \quad \forall\, \varepsilon > 0$$

holds, then we have

$$\text{(5.22.2)} \qquad U_n = s_n^{-1} \sum_{j=1}^{n} (X_{nj} - \mu_{nj}) \xrightarrow{\sim} N(0,1)$$

Moreover, Lyapounov's condition:

$$\text{(5.22.3)} \qquad \exists \alpha > 2 \;:\; \lim_{n\to\infty} \left\{ s_n^{-\alpha} \sum_{j=1}^{n} E|X_{nj} - \mu_{nj}|^\alpha \right\} = 0$$

implies Lindeberg's condition (5.22.1) and (5.22.2). Similarly, the conditions:

$$\left\{ \frac{(X_{nj} - \mu_{nj})^2}{\sigma_{nj}^2} \;\middle|\; n \in \mathbf{N}, 1 \le j \le n \right\} \text{ is uniformly integrable} \tag{5.22.4}$$

$$\lim_{n\to\infty} \left\{ s_n^{-1} \max_{1\le j\le n} \sigma_{nj} \right\} = 0 \tag{5.22.5}$$

implies Lindeberg's condition (5.22.1) and (5.22.2).

Proof: Set $Y_{nj} = (X_{nj} - \mu_{nj})/s_n$ and $q(x) = x$. Then by (5.21) and (5.21.6) we have that (5.22.1) implies (5.22.2) and (5.22.3) implies (5.22.1).

Suppose that (5.22.4+5) holds and let $\varepsilon > 0$ and $\delta > 0$ be given. Then by (5.22.4) and (3.24.4), there exists $a > 0$ such that

$$\int_{\{|X_{nj}-\mu_{nj}|>a\sigma_{nj}\}} (X_{nj} - \mu_{nj})^2 \, dP < \delta\sigma_{nj}^2$$

for all $n \in \mathbf{N}$ and all $j = 1, \ldots, n$. Moreover, by (5.22.5) there exists $n_0 \in \mathbf{N}$ such that $a\sigma_{nj} \le \varepsilon s_n$ for all $n \ge n_0$ and all $j = 1, \ldots, n$. Hence, if $n \ge n_0$, then

$$\begin{aligned} & s_n^{-2} \sum_{j=1}^{n} \int_{\{|X_{nj}-\mu_{nj}|>\varepsilon s_n\}} (X_{nj} - \mu_{nj})^2 \, dP \\ & \le s_n^{-2} \sum_{j=1}^{n} \int_{\{|X_{nj}-\mu_{nj}|>a\sigma_{nj}\}} (X_{nj} - \mu_{nj})^2 \, dP \\ & \le s_n^{-2} \sum_{j=1}^{n} \sigma_{nj}^2 \, \delta = \delta \end{aligned}$$

and so (5.22.1) holds. □

5.23 The Normal Central Limit Theorem III

Let $X_1, X_2, \cdots$ be independent random variables such that $\mu_n = EX_n$ and $\sigma_n^2 = \operatorname{var} X_n$ exists and are finite. Set $s_n = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$ and suppose that $\sigma_1^2 > 0$. If Lindeberg's condition:

$$\lim_{n\to\infty} \left\{ s_n^{-2} \sum_{j=1}^{n} \int_{\{|X_j-\mu_j|>\varepsilon s_n\}} (X_j - \mu_j)^2 \, dP \right\} = 0 \qquad \forall \varepsilon > 0 \tag{5.23.1}$$

holds, then we have

$$(5.23.2) \qquad U_n = s_n^{-1} \sum_{j=1}^{n} (X_j - \mu_j) \xrightarrow{\sim} N(0,1)$$

Moreover, Lyapounov's condition:

$$(5.23.3) \qquad \exists \alpha > 2 \ : \ \lim_{n\to\infty} \left\{ s_n^{-\alpha} \sum_{j=1}^{n} E|X_j - \mu_j|^\alpha \right\} = 0$$

implies Lindeberg's condition (5.23.1) and (5.23.2). Similarly, the conditions:

$$(5.23.4) \qquad \left\{ \frac{(X_j - \mu_j)^2}{\sigma_j^2} \;\middle|\; n \in \mathbf{N}, 1 \le j \le n \right\} \text{ is uniformly integrable}$$

$$(5.23.5) \qquad \lim_{n\to\infty} \left\{ s_n^{-1} \max_{1\le j\le n} \sigma_j \right\} = 0$$

implies Lindeberg's condition (5.23.1), and (5.23.2).

Proof: Apply (5.22) with $X_{nj} = X_j$. □

Remark: Suppose that $X_1, X_2, \cdots$ are independent, identically distributed random variables with mean μ and variance $0 < \sigma^2 < \infty$. Then evidently (5.23.1) holds and so (5.16.7) is a special case of the central limit theorem III.

5.24 The Poisson Central Limit Theorem

Let $\{X_{nj} \mid 1 \le j \le n\}$ *be an independent triangular scheme satisfying the uan condition (5.16.2), and let* $0 < \lambda < \infty$ *be a positive finite number satisfying*

$$(5.24.1) \qquad \lim_{n\to\infty} \sum_{j=1}^{n} \{P(X_{nj} > 1 + x) + P(X_{nj} < -x)\} = 0 \qquad \forall x > 0$$

$$(5.24.2) \qquad \lim_{n\to\infty} \sum_{j=1}^{n} P(X_{nj} > x) = \lambda \qquad \forall 0 < x < 1$$

$$(5.24.3) \qquad \forall \varepsilon > 0 \; \exists \delta > 0 \ : \ \limsup_{n\to\infty} \left\{ \sum_{j=1}^{n} \int_{\{|X_{nj}|<\delta\}} |X_{nj}| \, dP \right\} \le \varepsilon$$

Then we have

$$(5.24.4) \qquad U_n = \sum_{j=1}^{n} X_{nj} \xrightarrow{\sim} Po(\lambda)$$

Proof: Let $q(x) = w(x) = 0$ and $\nu = \lambda\,\delta_1$, where δ_1 is the Dirac measure at 1. Then by (5.24.1+2) we have that (5.20.1+2) and (5.20.4+5) hold with $D_1 = D_2 =]0,\infty[\setminus\{1\}$ and $\alpha = 0$, and by (5.24.3) we have that (5.20.3) holds. Moreover, since $Po(\lambda) = \mathrm{Lev}(0,0,q,\nu)$, we see that (5.24) follows from (5.20). □

5.25 The Stable Central Limit Theorem

Let $X_1, X_2, \cdots$ be independent, identically distributed random variables such that

$$(5.25.1) \qquad \lim_{x\to\infty} x^p P(X_1 > x) = \alpha \quad , \qquad \lim_{x\to\infty} x^p P(X_1 < -x) = \beta$$

where $0 < p < 2$, $\alpha, \beta \in [0,\infty[$ and $\alpha + \beta > 0$. Set

$$m_n = \begin{cases} 0 & \text{if } 0 < p < 1 \\ n\,E\{\sin(\frac{1}{n}X_1)\} & \text{if } p = 1 \\ EX_1 & \text{if } 1 < p < 2 \end{cases}$$

$$\sigma = \{p(\alpha+\beta)\,C(p)\}^{1/p} \quad , \quad \gamma = \frac{\beta-\alpha}{\beta+\alpha}$$

where $C(p)$ is given by (3.37.14) or (3.37.16). Then we have

$$(5.25.2) \qquad U_n = n^{-1/p}\sum_{j=1}^{n}(X_j - m_n) \xrightarrow{\sim} Stab(0,\sigma,p,\gamma)$$

where $Stab(0,\sigma,p,\gamma)$ is the stable distribution with parameters $(0,\sigma,p,\gamma)$ [see (5.18)].

Proof: Let $X_{nj} = n^{-1/p} X_j$, and let ν_p and q_p be given by (5.18.1+2). Then $Eq_p(X_{nj}) = n^{-1/p} m_n$, $Stab(0, \sigma, p, \gamma) = Lev\,(0, 0, q_p, \nu_p)$ and (5.20.1+2) follow from (5.25.1). Let $r > p$ and $\delta > 0$ be given numbers and set $Y = |X_1|\ 1_{\{|X_1| < \delta n^{1/p}\}}$. Then by (5.25.1) there exists $0 \le c < \infty$ so that $P(Y > t) \le P(|X_1| > t) \le ct^{-p}$ for all $t > 0$ and $P(Y > t) = 0$ for all $t \ge \delta n^{1/p}$. Since $X_1, X_2, \ldots$ are identically distributed, then by (4.2.7) we get

$$\sum_{j=1}^{n} \int_{\{|X_{nj}| < \delta\}} |X_{nj}|^r \, dP = n^{1-\frac{r}{p}} EY^r = rn^{1-\frac{r}{p}} \int_0^\infty t^{r-1} P(Y > t)\, dt$$

$$\le crn^{1-\frac{r}{p}} \int_0^{\delta n^{1/p}} t^{r-1-p}\, dt = \tfrac{cr}{r-p}\, \delta^{r-p}$$

Thus, we see that $\rho_0(x^r) = 0$ whenever $r > p$. Similarly, we find that $\rho_\infty(x^r) = 0$ whenever $r < p$ and by (5.19.7), we have that $\rho_\infty(q_1) = 0$. Hence setting $w(x) = 0$ and $q = q_p$, we see that (5.20.3)–(5.20.5) hold with $\alpha = 0$. Thus, (5.25) follows from (5.20). □

5.26 A Review of the Central Limit Theorem

We have now seen that every Lévy-Khinchine distribution $Lev(\mu, \sigma, q, \nu)$ has its own central limit theorem, but the normal central limit theorems (5.21)–(5.23) are by far the most important central limit theorems. The research in the central limit theorem is far from finished and today, probabilists work with the central limit theorems in many different directions. For instance,

What about higher dimensions? Unexpected problems occur here, partly due to the fact that not all matrices commute and not all matrices are invertible. Even in infinite dimensional spaces there exists central limit theorems, and here a surprising connection between the geometric structure of the unit ball and the central limit theorem has emerged.

What about dependent random variables? The recent research has given series of central limit theorems for "weakly dependent" random variables, such as martingale differences, interchangeable variables, etc.

How fast is the convergence? This question is of vital interest in statistics where many results about the asymptotic distribution of estimators and testors depend on the central limit theorem, and the validity of the statistical method depends on how good the asymptotic distribution approximates the real distribution. The best possible answer to this question that probability theory today can provide, is disappointingly weak:

(5.26.1) Suppose that $\{X_n\} \subseteq L^3(P)$ are independent identically distributed random variables with mean μ and variance 1 . Let $\Phi(x)$

be the distribution function of the normal $N(0,1)$-distribution and set $U_n = (X_1 + \cdots + X_n)/\sqrt{n}$. Then we have

$$|P(U_n \le x) - \Phi(x)| \le \frac{1.322 \cdot E|X_1|^3}{\sigma^3 \sqrt{n}} \quad \forall x \in \mathbf{R}\, \forall n \ge 1$$

However, if $\beta := EX_n^3 \ne 0$, then by Exc.4.40 and (4.2.8) we have

$$\beta = \sqrt{n}\, EU_n^3 = \sqrt{n} \int_0^\infty 2x^2 \left(1 - P(U_n \le x) - P(U_n \le -x)\right) dx$$

and since $\Phi(-x) = 1 - \Phi(x)$ there exists $0 < a < \infty$ such that

$$\begin{aligned} \tfrac{1}{2}\,|\beta| &\le \sqrt{n} \left| \int_0^a 2x^2 \left(1 - P(U_n \le x) - P(U_n \le -x)\right) dx \right| \\ &\le \sqrt{n}\, 2a^3 \sup_{x \in \mathbf{R}} |P(U_n \le x) - \Phi(x)| \end{aligned}$$

Thus, we get

(5.26.2) Suppose that $\{X_n\} \subseteq L^3(P)$ are independent identically distributed random variables with mean μ , variance 1 and $EX_n^3 \ne 0$. Let $\Phi(x)$ and U_n be defined as above. Then there exists a positive number $c > 0$ satisfying:

$$\sup_{x \in \mathbf{R}} |P(U_n \le x) - \Phi(x)| \ge \frac{c}{\sqrt{n}} \qquad \forall n \ge 1$$

Estimates of this kind are called *Berry-Essén estimates*. Note that (5.26.1) and (5.26.2) say that n has to be of the order 10,000 before we know that $\Phi(x)$ and $P(U_n \le x)$ coincide on the first two decimal places. Prudent statisticians claim that the normal approximation works if $n \ge 10$, and bolder statisticians allow n's between 5 and 10. Moreover, computational experiments prove their point. This is far from $n \sim 10,000$, which is the best that probabilists can do today. It is fair to say that the Berry-Essén estimates are a defeat and challenge for today's probability theory.

5.27 The Inversion Formulas

(1): Let $\alpha > 0$ be a given positive number, and let W be a random variable, which only assumes the values $0, \pm\alpha, \pm 2\alpha, \pm 3\alpha, \ldots$. Let ψ be the characteristic

function of W *and let* $p_k = P(W = k\alpha)$ *for all* $k \in \mathbf{Z}$. *If* $\beta = \frac{2\pi}{\alpha}$, *then* ψ *is periodic with period* β *and we have*

$$(5.27.1) \qquad \psi(t) = \sum_{n=-\infty}^{\infty} p_n\, e^{itn\alpha} \quad , \quad p_k = \frac{1}{\beta}\int_c^{c+\beta} e^{-ixk\alpha}\, \psi(x)\, dx$$

for all $t, c \in \mathbf{R}$ *and all* $k \in \mathbf{Z}$. *Moreover, if* $\xi, \eta : [0, \beta[\to \mathbf{C}$ *are* λ*-integrable on* $[0, \beta[$ *where* λ *denotes the Lebesgue measure on* $\mathbf{R}$, *then* $\xi(t) = \eta(t)$ *for* λ*-a.a.* $t \in [0, \beta[$ *if and only if*

$$(5.27.2) \qquad \int_0^{\beta} e^{-ixk\alpha}\, \xi(x)\, dx = \int_0^{\beta} e^{-ixk\alpha}\, \eta(x)\, dx \qquad \forall k \in \mathbf{Z}$$

(2): Let X *be a random variable with characteristic function* $\varphi(t)$ *and distribution function* F *and let* $f_n(v)$ *be defined by*

$$(5.27.3) \qquad f_n(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2n^2}\right) e^{-ixv}\, \varphi(x)\, dx \qquad \forall v \in \mathbf{R}\, \forall n \geq 1$$

Then f_n *is a continuous bounded density function and if* C_F *is the set of all continuity points of* F , *then we have*

$$(5.27.4) \qquad F(a) = \lim_{n\to\infty} \int_{-\infty}^{a} f_n(v)\, dv \qquad \forall a \in C_F$$

Moreover, if φ *is* λ*-integrable on* $\mathbf{R}$, *then* X *is absolutely continuous with a bounded, continuous density function* $f(x)$ *given by*

$$(5.27.5) \qquad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixs}\, \varphi(s)\, ds = \frac{1}{\pi}\Re\left\{\int_0^{\infty} e^{-ixs}\, \varphi(s)\, ds\right\} \quad \forall x \in \mathbf{R}$$

Remark: Note that formulas (5.27.1) (5.27.4) and (5.27.5) express the probability mass function, the distribution function and the density function in terms of the characteristic function, and for this reason the formulas goes under the name *the inversion formulas for characteristic functions.*

Proof: (1): The first formula in (5.27.1) follows directly from (4.2.10). If $k, n \in \mathbf{Z}$, then

$$\int_c^{c+\beta} e^{-ixk\alpha}\, e^{ixn\alpha}\, dx = \begin{cases} 0 & \text{if } k \neq n \\ \beta & \text{if } k = n \end{cases}$$

for all $c \in \mathbf{R}$. Hence, the second formula in (5.27.1) follows from the first and Beppo Levi's theorem (3.32.3).

Suppose that (ξ, η) satisfies (5.27.2). Taking complex conjugates in (5.27.2) we see that $(\bar{\xi}, \bar{\eta})$ satisfies (5.27.2), and since $\Re z = \frac{1}{2}(z+\bar{z})$ and $\Im z = \frac{i}{2}(\bar{z}-z)$, we see that $(\Re\xi, \Re\eta)$ and $(\Im\xi, \Im\eta)$ satisfies (5.27.2). Thus, it is no loss of generality to assume, that ξ and η are real-valued, but then $(\tilde{\xi}, \tilde{\eta})$ satisfies (5.27.2) where $\tilde{\xi} = \xi^+ + \eta^-$ and $\tilde{\eta} = \eta^+ + \xi^-$. Let $\Gamma(x,k) = e^{-ixk\alpha}$ for $(x,k) \in [0,\beta[\times\mathbf{Z}$. Then Γ is a kernel on $[0,\beta[\times\mathbf{Z}$, and since $\alpha\beta = 2\pi$, we see that

$$x^* := -x \quad \text{and} \quad x\nabla y := \begin{cases} x+y & \text{if } x+y<\beta \\ x+y-\beta & \text{if } x+y \geq \beta \end{cases}$$

satisfies (3.41.1+2). Let $d\mu = \tilde{\xi}\, d\lambda$ and $d\nu = \tilde{\eta}\, d\lambda$. Since $(\tilde{\xi}, \tilde{\eta})$ satisfies (5.27.2) and $x \curvearrowright \Gamma(x,1) = e^{-ix\alpha}$ is continuous and injective, we see that (3.41.3)–(3.41.6) holds with $T_0 = T_1 = \mathbf{Z}$ and $\sigma(\Gamma(\cdot,k) \mid k \in \mathbf{Z}) = \mathcal{B}([0,\beta[)$. Hence, by (3.41) we get

$$\int_B \tilde{\xi}(x)\, dx = \int_B \tilde{\eta}(x)\, dx \qquad \forall B \in \mathcal{B}([0,\beta[)$$

and so we have $\tilde{\xi} = \tilde{\eta}$ λ-a.s. by (3.17.3). But then it follows easily that $\xi = \eta$ λ-a.s. And the converse is evident.

(2): Let $U \sim N(0,1)$ be a normally distributed random variable such that U is independent of X and set $X_n = X + \frac{1}{n}U$. By (4.22.1) and (4.22.3), we have that $\frac{1}{n}U$ has density function $g_n(x) = n(2\pi)^{-1/2}\varphi_{n\cdot U}(x)$ and by (4.6.6) and (4.18.4+5), we have that X_n has density $f_n(v)$ given by

$$\begin{aligned} f_n(v) &= \frac{n}{\sqrt{2\pi}} E\varphi_{n\cdot U}(X-v) = \frac{n}{\sqrt{2\pi}} E\varphi_{X-v}(nU) \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\Big(-\frac{x^2}{2n^2}\Big) e^{-ivx}\,\varphi(x)\, dx \end{aligned}$$

Hence, the function defined by (5.27.3) is a density function and by Lebesgue's dominated convergence theorem (3.6), we see that f_n is bounded and continuous. Note that $X_n \to X$ a.s. and so $X_n \xrightarrow{\sim} X$. Hence, by (5.9.3) we see that (5.27.4) holds. Suppose that φ is Lebesgue integrable. Since $\exp(-s^2/(2n^2)) \to 1$ as $n \to \infty$ and $e^{-ivx}\,\varphi(x) + e^{ivx}\,\varphi(-x) = \Re(e^{-ivx}\,\varphi(x))$, then by integrability of φ and Lebesgue's dominated convergence theorem (3.6) we find

$$\lim_{n\to\infty} f_n(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ivx}\,\varphi(x)\, dx = \frac{1}{\pi}\Re\Big\{\int_0^{\infty} e^{-ivx}\,\varphi(x)\, dx\Big\}$$

$$P(a < X \leq b) = \lim_{n\to\infty} P(a < X_n \leq b) = \int_a^b f(v)\, dv$$

for all $v \in \mathbf{R}$ and all $a, b \in C_F$ with $a < b$ where f is given by (5.27.5). Thus, we conclude that X has density f and by Lebesgue's dominated convergence theorem (3.6), we see that f is bounded and continuous. $\square$

5.28 Infinitely Divisible Distributions

Let X be a random variable with distribution F and characteristic function φ. Then we say that X (or F or φ) is *infinitely divisible* if and only if

(5.28.1) For every $n \geq 1$ there exists independent identically distributed random variables $X_{n1}, \ldots, X_{nn}$ with $X \sim X_{n1} + \cdots + X_{nn}$

By (4.6.3), we have that the distribution function F is infinitely divisible if and only if

(5.28.2) $\forall n \geq 1 \; \exists F_n$ a distribution function : $F = (F_n)^{\star n}$

where $(F_n)^{\star n} = F_n \star \cdots \star F_n$ denotes the n-fold convolution of F_n with itself. By (4.18.7), we have that the characteristic function φ is infinitely divisible if and only if

(5.28.3) $\forall n \geq 1 \; \exists \varphi_n$ a characteristic function : $\varphi(t) = \varphi_n(t)^n \quad \forall t \in \mathbf{R}$

Let us look at (5.28.3). Since $\varphi(t) = \varphi_n(t)^n$, we have that $\varphi_n(t)$ is an nth root of $\varphi(t)$ and since $\varphi(t)$ is a complex number, we know that $\varphi(t)$ has n different n^{th} roots, and (5.28.3) says that φ is infinitely divisible if for every n we can choose an nth root of $\varphi(t)$ which is a characteristic function.

Let X be infinitely divisible. Then by the very definition (5.28.1), we see that X has the same distribution as the sum of "many but small" independent identically distributed random variables. Thus, the distribution of X may occur as the limit distribution in a central limit theorem. As a matter of fact, the infinitely divisible distributions are the only distributions, which can occur as the limit distributions in a central limit theorem. More precisely:

(5.28.4) Let $\{X_{nj} \mid 1 \leq j \leq n\}$ be an independent triangular scheme satisfying the uan condition (5.16.2) such that the centered row sums $U_n = (X_{n1} - m_{n1}) + \cdots + (X_{nn} - m_{nn})$ converges in law to a distribution F for some centering constants $\{m_{nj}\}$. Then F is infinitely divisible

But I shall abstain from the proof of this fact here.

Consider the Lévy-Khinchine distribution $Lev(\mu, \sigma, q, \nu)$. Then its characteristic function φ is given by (5.17.9) and by (5.28.3) it follows easily that $Lev(\mu, \sigma, q, \nu)$ is infinitely divisible. Conversely, if F is an infinitely divisible distribution, then there exists $\mu \in \mathbf{R}$, $\sigma^2 \in [0, \infty[$ and a Lévy-Khinchine pair (q, ν) such that $F = Lev(\mu, \sigma, q, \nu)$. In Exc.4.16 you will find an outline of the proof of this fact.

The Lévy-Khinchine distribution is given in terms of its characteristic function φ [see (5.17.16)]. In general there is no elementary formula for the distribution function or the density function (if it exists) of $Lev(\mu, \sigma, q, \nu)$. However, if $F(x)$ is the distribution function of the Lévy-Khinchine distribution $Lev(\mu, \sigma, q, \nu)$, then by (5.27.4) we find

$$(5.28.5) \qquad F(a) = \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\infty}^{a} dv \int_{-\infty}^{\infty} e^{-ivs-\frac{1}{2}s^2n^{-2}} \varphi(s)\, ds \quad \forall a \in C_F$$

where φ is given by (5.17.16) and C_F is the set of all continuity points of F . Moreover, if φ is Lebesgue integrable on $\mathbf{R}$, then $Lev(\mu, \sigma, q, \nu)$ has a continuous bounded density function given by (5.27.5).

5.29 Stable Distributions

Let X be a random variable with distribution F and characteristic function φ . Then we say that X (or F or φ) is *stable* if and only if

$$(5.29.1) \qquad \forall n \geq 1\, \exists a_n > 0\, \exists b_n \in \mathbf{R} \;:\; X_1 + \cdots + X_n \sim a_n X + b_n$$

where $X_1, X_2, \ldots$ are independent random variables with the same distribution as X . By (4.6.6), we have that the distribution function F is stable if and only if

$$(5.29.2) \qquad \forall n \geq 1\, \exists a_n > 0\, \exists b_n \in \mathbf{R} \;:\; F^{\star n}(x) = F\Big(\frac{x - b_n}{a_n}\Big) \quad \forall x \in \mathbf{R}$$

where $F^{\star n} = F \star \cdots \star F$ is the n-fold convolution of F with it self. By (4.18.7), we have that the characteristic function φ is stable if and only if

$$(5.29.3) \qquad \forall n \geq 1\, \exists a_n > 0\, \exists b_n \in \mathbf{R} \;:\; \varphi(t)^n = e^{itb_n}\, \varphi(a_n t) \quad \forall t \in \mathbf{R}$$

Suppose that X is degenerated, then $X = c$ a.s. for some number $c \in \mathbf{R}$. By (5.18.7), we have that $X \sim Stab(c, 0, p, \gamma)$ for any $0 < p \leq 2$ and any $-1 \leq \gamma \leq 1$. Moreover, we have that (5.29.1) holds whenever $a_n > 0$ and $b_n = (n - a_n)\, c$. Thus, every degenerated random variable is stable in the sense of (5.29.1) and has a stable distribution. We shall now see that the stable distributions $Stab(\mu, \sigma, p, \gamma)$ are exactly the stable distributions in the sense of (5.29.2).

5.30 Characterization of Stable Distributions

Let X be a stable non-degenerated random variable and let $a_n > 0$ and b_n be chosen such that (5.29.1) holds. Then there exist numbers $\mu \in \mathbf{R}$, $0 < \sigma < \infty$, $0 < p \leq 2$ and $-1 \leq \gamma \leq 1$ satisfying

$$X \sim Stab(\mu, \sigma, p, \gamma) \tag{5.30.1}$$

$$a_n = n^{1/p} \quad , \quad b_n = \begin{cases} (n - n^{1/p})\,\mu & \text{if } p \neq 1 \\ \frac{2\gamma\sigma}{\pi}\, n \log n & \text{if } p = 1 \end{cases} \tag{5.30.2}$$

Conversely, if $X \sim Stab(\mu, \sigma, p, \gamma)$ and if a_n and b_n are defined by (5.30.2), then (5.29.1) holds and so X is stable in the sense defined in (5.29.1). Thus, the set of all stable distributions $Stab(\mu, \sigma, p, \gamma)$ is exactly the set of all distributions satisfying (5.29.2)

Proof: Suppose that X is a non-degenerated random variable satisfying (5.29.1) for some $a_n > 0$ and some $b_n \in \mathbf{R}$, and let φ be the characteristic function of X. By Exc.5.27, we have that X is infinitely divisible and there exists $0 < p \leq 2$ such that $a_n = n^{1/p}$ for all $n \geq 1$. Hence, by Exc.5.12 we know that $\varphi(t) \neq 0$ for all $t \in \mathbf{R}$. Thus, by Exc.5.13 we have a continuous logarithm $\psi := \operatorname{Log}\varphi$ of φ.

Let me first treat the case: $p \neq 1$. By Exc.5.27, there exists $\mu \in \mathbf{R}$ such that $b_n = \mu\,(n - n^q)$ for all $n \geq 1$ where $q = \frac{1}{p}$. Then (5.30.2) holds and we just have to prove (5.30.1). Taking logarithms in (5.29.3) [see Exc.5.13], we find

$$n\,\psi(t) = it\mu\,(n - n^q) + \psi(n^q t) \qquad \forall\, t \in \mathbf{R}\ \forall\, n \in \mathbf{N}$$

Let $r = \frac{n}{m}$ be a given positive rational number, where $m, n \in \mathbf{N}$. Then by the above equation we get

$$\begin{aligned} \psi(r^q) &= i\mu\,\frac{(n^q - n)}{m^q} + n\,\psi(m^{-q}) \\ &= i\mu\,\frac{(n^q - n)}{m^q} + i\mu\,\frac{n(m - m^q)}{m^{q+1}} + r\,\psi(1) \\ &= i\mu\,(r^q - r) + r\,\psi(1) \end{aligned}$$

Hence, if $a = -\Re\psi(1)$ and $b = \Im\psi(1)$, then by continuity of ψ we see that $\psi(r^q) = i\mu r^q - ar - i(\mu - b)r$ for all $r \geq 0$; and since X is non-degenerated we have $a > 0$. Recall that $\varphi(t) = e^{\psi(t)}$ and $\varphi(-t) = \overline{\varphi(t)} = e^{\overline{\psi(t)}}$. Thus, setting $r = |t|^p$, we get

$$\varphi(t) = \exp\Big\{i\mu t - |\sigma t|^p(1 + i\,\gamma\,\mathcal{K}(p,t))\Big\} \qquad \forall\, t \in \mathbf{R}$$

where $\sigma := a^{1/p}$, $\gamma := \frac{\mu-b}{a} \cdot (\tan \frac{p\pi}{2})^{-1}$ and $\mathcal{K}(p,t)$ is given by (3.37.15). Hence, φ is on the form (5.18.3) and it only remains to show that $|\gamma| \leq 1$. By (7), in Exc.5.16 there exists a finite Borel measure ξ on $\mathbf{R}$ with Fourier transform given by

$$\hat{\xi}(t) = \frac{1}{2} \int_{-1}^{1} (\psi(t) - \psi(t+y))\, dy \qquad \forall t \in \mathbf{R}$$

Let $q_p(x)$ is given by (5.18.2) and

$$g(x) = \begin{cases} \dfrac{\sigma^p(1-\gamma)}{2\,C(p)\,|x|^{p+1}} & \text{if } x > 0 \\ \dfrac{\sigma^p(1+\gamma)}{2\,C(p)\,|x|^{p+1}} & \text{if } x < 0 \end{cases}$$

Then by (3.37.12+13) we have

$$\psi(v) = i\mu v + \int_{-\infty}^{\infty} \left(e^{isv} - 1 - iv\, q_p(s)\right) g(s)\, ds \qquad \forall v \in \mathbf{R}$$

Hence we find

$$\begin{aligned} \hat{\xi}(t) &= \frac{1}{2} \int_{-1}^{1} dy \int_{-\infty}^{\infty} \left(e^{its}\left(1 - e^{iys}\right) + iy\, q_p(s)\right) g(s)\, ds \\ &= \frac{1}{2} \int_{-\infty}^{\infty} ds \int_{-1}^{1} \left(e^{its}\left(1 - e^{iys}\right) + iy\, q_p(s)\right) g(s)\, dy \\ &= \int_{-\infty}^{\infty} e^{its} \left(1 - \tfrac{\sin s}{s}\right) g(s)\, ds \end{aligned}$$

Thus, by the uniqueness theorem for characteristic functions (4.18.8) we see that ξ is absolutely continuous with density $(1 - \frac{\sin s}{s})\, g(s)$. In particular, we have $g \geq 0$ Lebesgue a.s. and so $-1 \leq \gamma \leq 1$.

Next let us consider the case: $p = 1$. Then $a_n = n$ for all $n \geq 1$ and by Exc.5.27, we have that $b_1 = 0$ and $b_{n \cdot k} = n\, b_k + k\, b_n$ for all $n, k \geq 1$. Let $r = \frac{n}{m}$ be a given rational number. Then, as above, we find

$$\psi(r) = r\, \psi(1) + i\, \frac{n\, b_m - m\, b_n}{m^2}$$

Hence, if $f(t) = -i\,(\psi(t) - t\,\psi(1))/t$ for $t > 0$, then f is a continuous on $]0, \infty[$ and $f(\frac{n}{m}) = (mn)^{-1}\,(n\, b_m - m\, b_n)$ for all $n, m \in \mathbf{N}$. If $r = \frac{n}{m}$ and $s = \frac{k}{l}$ are two positive rational numbers, then using the equation $b_{n \cdot k} = n\, b_k + k\, b_n$ we get

$$f(rs) = \frac{nk\, b_{ml} - ml\, b_{nk}}{nkml} = \frac{n\, b_m - m\, b_n}{mn} + \frac{k\, b_l - l\, b_k}{kl} = f(r) + f(s)$$

Hence, by continuity of f we have that $f(uv) = f(u) + f(v)$ for all $u, v > 0$. But it is well-known that a continuous function satisfying this equation is a multiple of $\log u$. Thus, there exists $\beta \in \mathbf{C}$ such that $f(u) = \beta \log u$ for all $u > 0$. Recall that $b_1 = 0$ and $\psi(t) = it\, f(t) + t\,\psi(1)$. Hence, we find

$$\begin{aligned} b_n &= (1\, b_n - n\, b_1) = n\, f(\tfrac{1}{n}) = -\beta\, n \log n \qquad \forall n \geq 1 \\ \psi(t) &= t\,\psi(1) + it\beta \log t \qquad \forall t > 0 \end{aligned}$$

But then $\beta \in \mathbf{R}$ and (5.30.1+2) follow with $\mu = \Im\psi(1)$, $\sigma = -\Re\psi(1)$ and $\gamma = -\frac{\beta\pi}{2\sigma}$ in the same manner as above.

Conversely, if $X \sim Stab(\mu, \sigma, p, \gamma)$, then by (5.18.3) and (3.37.15) it follows easily that (5.29.3) holds with a_n and b_n given by (5.30.2). □

5.31 The Mellin Transform of $\exp\left(-us - z\, s^p\right)$

By (5.18.7) we know that $Stab(\mu, 0, p, \gamma)$ is the Dirac distribution δ_μ, and by (5.18.4+5) we know that $stab(\mu, \sigma, 2, \gamma)$ is a normal distribution $N(\mu, \sigma^2)$ and $stab(\mu, \sigma, 1, 0)$ is a Cauchy distribution $C(\mu, \sigma)$. Thus, if $\sigma > 0$ and $p = 2$, or if $\sigma > 0$ and $(p, \gamma) = (1, 0)$, then the density function of $stab(\mu, \sigma, p, \gamma)$ can be expressed in terms of well-known functions. In the next section, we shall see that if $\sigma > 0$, then the density of $stab(\mu, \sigma, \frac{1}{2}, \pm 1)$ can likewise be expressed in terms of well-known functions. In many textbooks it is claimed that the density of $stab(\mu, \sigma, p, \gamma)$ cannot be expressed in terms of known functions in any other case. This is a truth with modifications (it depends on which functions you know).

Let $X \sim stab(\mu, \sigma, p, \gamma)$ where $\mu \in \mathbf{R}$, $\sigma > 0$, $0 < p \leq 2$ and $-1 \leq \gamma \leq 1$, and let φ be the characteristic function of X. Then by (5.18.3) we have that $|\varphi(t)| = \exp(-|\sigma t|^p)$ and so φ is Lebesgue integrable. Hence, by (5.27) we have that X is absolutely continuous with a continuous bounded density $f(x)$ given by (5.27.5). Setting $c = 1 + i\gamma \tan\frac{p\pi}{2}$ if $p \neq 1$, and $\kappa = \frac{2\sigma\gamma}{\pi}$ if $p = 1$. Then inserting the expression (5.18.3) for φ in (5.27.5), we see that the density function for X is given by

$$(5.31.1) \qquad f(x) = \begin{cases} \frac{1}{\pi}\,\Re\left\{\int_0^\infty e^{-is(x-\mu)-\sigma^p c s^p}\, ds\right\} & \text{if } p \neq 1 \\ \frac{1}{\pi}\,\Re\left\{\int_0^\infty e^{-s(\sigma+i(x-\mu))-i\kappa s \log s}\, ds\right\} & \text{if } p = 1 \end{cases}$$

for all $x \in \mathbf{R}$. Thus, we "only" have to compute the integrals above in order to obtain an expression for f . The first integral in (5.31.1) can be expressed in terms the so-called incomplete hypergeometric function ${}_1G_0(\alpha,\beta,z)$, which is defined for $\alpha > 0$ and $0 < \beta \le 1$ as follows:

$$(5.31.2)\qquad {}_1G_0(\alpha,\beta,z) = \begin{cases} \dfrac{1}{\Gamma(\alpha)} \displaystyle\sum_{n=0}^{\infty} \frac{\Gamma(\alpha+n\beta)}{n!}\, z^n & \text{if } 0<\beta<1 \\ (1-z)^{-\alpha} & \text{if } \beta = 1 \text{ and } z \ne 1 \end{cases}$$

[It is easily seen that the radius of convergence of the power series (5.31.2) equals $+\infty$ if $0 < \beta < 1$, and 1 if $\beta = 1$.] Moreover, using $\frac{\Gamma(\alpha+n)}{n!\,\Gamma(\alpha)} = (-1)^n \binom{-\alpha}{n}$, we see that the two expressions in (5.31.2) agree for $\beta = 1$ and $|z| < 1$.

Recall that z^p denotes the main p^{th} root of z , that is, [see (3.37.8)],

$$(5.31.3)\qquad z^p := |z|^p\, e^{ip\,\mathrm{Arg}(z)}$$

where $\mathrm{Arg}(z) = \theta$ is the unique angle $\theta \in]-\pi,\pi]$ satisfying $z = |z|\,e^{i\theta}$ whenever $z \ne 0$. It is easily checked that we have

$$(5.31.4)\qquad u^{p+q} = u^p\,u^q \ , \quad v^{-p} = \left(\tfrac{1}{v}\right)^p = \tfrac{1}{v^p} \text{ and } \overline{(v)^p} = (\bar{v})^p \text{ if } \mathrm{Arg}(v) \ne \pi$$

$$(5.31.5)\qquad u^p\,v^p = (uv)^p \quad \text{if } -\pi < \mathrm{Arg}(u) + \mathrm{Arg}(v) \le \pi \text{ or if } p \in \mathbf{Z}$$

$$(5.31.6)\qquad u^{pq} = (u^p)^q \quad \text{if } -\pi < p\,\mathrm{Arg}(u) \le \pi \text{ or if } q \in \mathbf{Z}$$

whenever $u, v \in \mathbf{C} \setminus \{0\}$ and $p, q \in \mathbf{R}$.

Note that coefficients in the series defining ${}_1G_0(\alpha,\beta,z)$ are real. Hence, if $\alpha > 0$ and $0 < \beta < 1$, then we have

$$(5.31.7)\qquad \Im\{{}_1G_0(\alpha,\beta,x)\} = 0 \qquad \forall x \in \mathbf{R}$$

$$(5.31.8)\qquad \Im\{{}_1G_0(\alpha,\beta,z)\} = -\Im\{{}_1G_0(\alpha,\beta,\bar{z})\} \qquad \forall z \in \mathbf{C}$$

and since $\beta n\,\Gamma(\beta n) = \Gamma(1+\beta n)$, we have

$$\begin{aligned} {}_1G_0(\beta,\beta,z) &= \frac{1}{z\,\Gamma(\beta)} \sum_{n=1}^{\infty} \frac{\Gamma(\beta n)}{(n-1)!}\, z^n = \frac{1}{z\,\Gamma(\beta)} \sum_{n=1}^{\infty} \frac{\Gamma(1+\beta n)}{n!}\, z^n \\ &= \frac{1}{z\,\Gamma(\beta)} \Big({}_1G_0(\alpha,\beta,z) - 1 \Big) \end{aligned}$$

Hence, if $0 < \beta < 1$ and $z \in \mathbf{C} \setminus \{0\}$, or if $\beta = 1$ and $z \in \mathbf{C} \setminus \{0,1\}$, then we have

$$(5.31.9)\qquad {}_1G_0(\beta,\beta,z) = \frac{1}{z\,\Gamma(\beta)} \Big({}_1G_0(1,\beta,z) - 1 \Big)$$

Let $p > 0$ be a positive number and let $u, z \in \mathbf{C}$ be two complex numbers. If we set $g(s) := \exp(-us - z\,s^p)$, then it is easily checked that the Mellin transform $M_g(\alpha)$ [see (3.36)] exists for all $\alpha > 0$, provided that $(u, z) \in \Delta_p$ where $\Delta_p \subseteq \mathbf{C} \times \mathbf{C}$ is given by

$$\Delta_p = \{\, (u, z) \in \mathbf{C} \times \mathbf{C} \mid \Re u > 0 \text{ or } \Re u = 0 \text{ and } \Re z > 0 \,\} \quad \text{if } p < 1$$

$$\Delta_1 = \{\, (u, z) \in \mathbf{C} \times \mathbf{C} \mid \Re(u + z) > 0 \,\}$$

$$\Delta_p = \{\, (u, z) \in \mathbf{C} \times \mathbf{C} \mid \Re u > 0 \text{ or } \Re u = 0 \text{ and } \Re z > 0 \,\} \quad \text{if } p > 1$$

Hence, we may define the function $M(\alpha, p, u, z)$ as follows:

$$M(\alpha, p, u, z) := \int_0^\infty s^{\alpha-1}\, e^{-us-zs^p}\, ds \qquad \forall\, \alpha, p > 0\ \forall\, (u, z) \in \Delta_p \tag{5.31.10}$$

Let $\alpha, p > 0$ and $(u, z) \in \Delta_p$ be given, then I claim that we have the following formulas:

$$M(\alpha, 1, u, z) = \frac{\Gamma(\alpha)}{(u + z)^\alpha} \quad \text{if} \quad \Re(u + z) > 0 \tag{5.31.11}$$

$$M(\alpha, p, u, 0) = \frac{\Gamma(\alpha)}{u^\alpha} \quad \text{if} \quad \Re u > 0 \tag{5.31.12}$$

$$M(\alpha, p, 0, z) = \frac{\Gamma(\frac{1}{p})}{p\, z^{\alpha/p}} \quad \text{if} \quad \Re z > 0 \tag{5.31.13}$$

$$M(\alpha, p, u, z) = \frac{\Gamma(\alpha)}{u^\alpha}\, {}_1G_0(\alpha, p, -z\, u^{-p}) \quad \text{if } p < 1\,,\ u \neq 0 \tag{5.31.14}$$

$$M(\alpha, p, u, z) = \frac{\Gamma(\frac{\alpha}{p})}{p\, z^{\alpha/p}}\, {}_1G_0\left(\tfrac{\alpha}{p}, q, -u\, z^{-\frac{1}{p}}\right) \quad \text{if } p > 1\,,\ z \neq 0 \tag{5.31.15}$$

$$M(1, p, u, z) = \tfrac{1}{u}\left(1 - {}_1G_0\left(\tfrac{1}{p}, q, -u\, z^{-\frac{1}{p}}\right)\right) \quad \text{if } p > 1, u \neq 0\,,\ z \neq 0 \tag{5.31.16}$$

Proof: (5.31.11) and (5.31.12) follow from (3.37.7), and (5.31.13) follows from (5.31.12) by substituting $s = t^q$, where $q = \frac{1}{p}$.

(5.31.14): Let $0 < p < 1$ and $\Re u > 0$. Then by (3.37.7) and Beppo Levi's

theorem (3.32.3) we have

$$\int_0^\infty s^{\alpha-1}\, e^{-us-zs^p}\, ds = \int_0^\infty \sum_{n=0}^\infty \frac{(-z)^n}{n!}\, s^{np+\alpha-1}\, e^{-us}\, ds$$

$$= \sum_{n=0}^\infty \frac{(-z)^n}{n!} \int_0^\infty s^{np+\alpha-1}\, e^{-us}\, ds = \sum_{n=0}^\infty \frac{\Gamma(\alpha+np)}{n!}\,(-z)^n\, u^{-np-\alpha}$$

$$= u^{-\alpha}\,\Gamma(p)\ {}_1G_0\big(\alpha, p, -zu^{-p}\big)$$

Thus, (5.31.14) holds whenever $\Re u > 0$. So suppose that $\Re u = 0$ and $u \neq 0$. Then $\Re z > 0$, and setting $u_\varepsilon := \varepsilon + u$, we have $\Re u_\varepsilon > 0$. Hence, letting $\varepsilon \downarrow 0$ and applying Lebesgue's dominated convergence theorem (3.6), we obtain (5.31.14).

(5.31.15): Suppose that $p > 1$ and $z \neq 0$, and let $q = \frac{1}{p}$. Substituting $s = t^q$ and applying (5.31.14), we get

$$\int_0^\infty s^{\alpha-1} e^{-us-zs^p}\, ds = q \int_0^\infty t^{\alpha q-1}\, e^{-ut^q-zt}\, dt$$

$$= z^{-\alpha q}\, q\,\Gamma(\alpha q)\ {}_1G_0\big(\alpha q, q, -uz^{-\alpha q}\big)$$

Thus, (5.31.15) is proved.

(5.31.16): Follows directly from (5.31.15) and (5.31.9). □

5.32 Stable Densities

Let $f(x)$ *denote the density function of* $Stab(\mu, \sigma, p, \gamma)$ *where* $\mu \in \mathbf{R}$, $0 < \sigma < \infty$, $0 < p \leq 2$ *and* $-1 \leq \gamma \leq 1$. *If* $p \neq 1$, *we define*

$$c = 1 + i\gamma \tan\tfrac{p\pi}{2} \quad , \quad \rho = |c| = \sqrt{1 + \gamma^2 \tan^2 \tfrac{p\pi}{2}}$$

$$q = \tfrac{1}{p} \quad , \quad \theta = \operatorname{Arg}(c) = \arctan\big(\gamma \tan\tfrac{p\pi}{2}\big)$$

Then $|\theta| \leq \frac{p\pi}{2}$ *if* $p < 1$, *and* $|\theta| \leq (2-p)\frac{\pi}{2}$ *if* $p > 1$. *Moreover, we have*

$$f(\sigma x + \mu) = \begin{cases} \dfrac{\Im\Big\{ {}_1G_0\big(1, p, -c\,(ix)^{-p}\big) \Big\}}{\pi\sigma x} & \text{if } p < 1\,,\ x \neq 0 \\[2ex] \dfrac{\Gamma\big(1+q\big)\,\cos\big(q\theta\big)}{\sigma\pi\rho^q} & \text{if } p < 1\,,\ x = 0 \end{cases} \tag{5.32.1}$$

$$(5.32.2)\qquad f(\sigma x+\mu)=\begin{cases}\dfrac{\Im\{\,{}_1G_0(1,q,i(\bar{c})^{-q}x)\,\}}{\pi\sigma x} & \text{if } p>1\,,\ x\neq 0\\[2ex] \dfrac{\Gamma(1+q)\,\cos(q\theta)}{\sigma\pi\rho^q} & \text{if } p>1\,,\ x=0\end{cases}$$

$$(5.32.3)\qquad f(\sigma x+\mu)=\frac{e^{-\frac{1}{4}x^2}}{2\sigma\sqrt{\pi}}\quad \text{if } p=2\,,\ x\in\mathbf{R}$$

$$(5.32.4)\qquad f(\sigma x+\mu)=\frac{1}{\sigma\pi(1+x^2)}\quad \text{if } p=1\,,\ \gamma=0\,,\ x\in\mathbf{R}$$

$$(5.32.5)\qquad f(\sigma x+\mu)=0\quad \text{if } p<1\,,\ \gamma=\pm 1\,,\ \gamma x\geq 0$$

$$(5.32.6)\qquad f(\sigma x+\mu)=\begin{cases}\dfrac{\exp\left(-\frac{1}{2|x|}\right)}{\sigma\sqrt{2\pi|x|^3}} & \text{if } p=\frac{1}{2}\,,\ \gamma=\pm 1\,,\ \gamma x<0\\[2ex] 0 & \text{if } p=\frac{1}{2}\,,\ \gamma=\pm 1\,,\ \gamma x\geq 0\end{cases}$$

Suppose that $1<p\leq 2$ *and let* $g(x)$ *be the density function of* $stab(\nu,\tau,q,\xi)$ *, where* $\nu\in\mathbf{R}$ *,* $\tau>0$ *,* $q=\frac{1}{p}$ *and*

$$\vartheta=q\theta+\tfrac{\pi}{2}(q-1)\quad,\quad r=\sqrt{1+\tan^2\vartheta}\quad,\quad \xi=\frac{\tan\vartheta}{\tan\frac{q\pi}{2}}$$

[Note that $|\theta|\leq(2-p)\frac{\pi}{2}$ *and so* ξ *as defined above belongs to the interval* $[-1,1]$ *.] Then we have*

$$(5.32.7)\qquad g(\tau\rho x+\nu)=\tfrac{\sigma r}{\tau\rho}\,x^{-q-1}\,f(\sigma r x^{-q}+\mu)\qquad \forall x>0$$

Remark: Formulas (5.32.1)–(5.32.6) express the density function for $Stab(\mu,\sigma,p,\gamma)$ in terms of known functions in all cases, except the case: $p=1$, $\gamma\neq 0$. In the exceptional case, the density is given by the second integral in (5.31.1), but I have not been able to compute that integral. Note that by (5.32.5) we have that $f(x)=0$ if either $0<p<1$, $\gamma=1$, $x\geq\mu$, or if $0<p<1$, $\gamma=-1$, $x\leq\mu$. It can be shown that $f(x)>0$ in all other cases.

Proof: (5.32.1+2): If $p\neq 1$, then $f(\sigma x+\mu)=\frac{1}{\pi}\Re\{M(1,p,i\sigma x,\sigma^p c)\}$ by (5.31.1). Hence, (5.32.1) follows by inserting in formulas (5.31.13) and (5.31.14). Similarly, (5.32.2) follows by inserting in formulas (5.31.13) and (5.31.16) and applying (5.31.9).

(5.32.3+4): In the first case, we have $\varphi(t)=e^{it\mu-\sigma^2t^2}$. In the second case, we have $\varphi(t)=e^{it\mu-|\sigma t|}$. Hence, (5.32.3+4) follows from (4.22.3), (5.31.1) and (5.31.11).

(5.32.5): Suppose that $p < 1$ and $\gamma = 1$. Then $\theta = \frac{p\pi}{2}$ and $i^{-p} = e^{-i\theta}$. Hence, we get

$$-c\,(ix)^{-p} = -|c|\,e^{i\theta}\,i^{-p}\,x^{-p} = -|c|\,x^{-p} \in \mathbf{R} \qquad \forall x > 0$$

Similarly, if $p < 1$ and $\gamma = -1$, then we have $-c\,(ix)^{-p} \in \mathbf{R}$ for all $x < 0$. Thus, (5.32.5) follows from (5.31.1+2) and (5.31.7).

(5.32.7): Let $x > 0$ be given and let $d = 1 + i\xi \tan\frac{q\pi}{2} = r\,e^{i\vartheta}$. Since $\vartheta - \frac{q\pi}{2} = q\theta - \frac{\pi}{2}$, we get

$$-d(i\rho x)^{-q} = -r(\rho x)^{-q}\,e^{i\left(\vartheta - \frac{q\pi}{2}\right)} = -r(\rho x)^{-q}\,e^{i\left(q\theta - \frac{\pi}{2}\right)} = i(\bar{c})^{-q}\,r x^{-q}$$

Thus, (5.32.7) follows from (5.32.1+2).

(5.32.6): Setting $p = 2$, $\nu = \mu$ and $\tau = \sigma$ in (5.32.7), we get $q = \frac{1}{2}$, $\vartheta = -\frac{\pi}{4}$, $\xi = -1$, $\rho = 1$ and $r = \sqrt{2}$. If $\gamma = -1$, then (5.32.6) follows from (5.32.1+2) and (5.32.5). If $\gamma = 1$, then (5.32.6) follows from (5.18.8) and the case $\gamma = -1$. □

5.33 Normal and Poisson Approximation of $Bi(n,p)$

Let $X_{n1}, \ldots, X_{nn}$ be independent random variables all having a binomial $Bi(1,p_n)$ -distribution where $0 < p_n < 1$ for all $n \geq 1$. Let

$$S_n = \sum_{j=1}^{n} X_{nj} \quad , \quad s_n = \sqrt{n\,p_n(1-p_n)} \quad , \quad U_n = \frac{1}{s_n}\sum_{j=1}^{n}(X_{nj} - p_n)$$

Note that $p_n = EX_{nj}$, $p_n(1-p_n) = \operatorname{var} X_{nj}$ and $P(|X_{nj} - p_n| > \varepsilon s_n) = 0$ whenever $s_n \geq \varepsilon^{-1}$. Hence, if $s_n \to \infty$, then Lindeberg's condition (5.22.1) holds. Thus, by (5.22) we have

(5.33.1) If $\lim\limits_{n\to\infty} s_n = \infty$, then $U_n \xrightarrow{\sim} N(0,1)$

Since $P(|X_{nj}| > \varepsilon) \leq p_n$, we see that $\{X_{nj}\}$ satisfies the uan condition (5.16.2) whenever $p_n \to 0$. Moreover, we have

$$\sum_{j=1}^{n}\{P(X_{nj} > 1 + x) + P(X_{nj} < -x)\} = 0 \qquad \forall x > 0$$

$$\sum_{j=1}^{n} P(X_{nj} > x) = n\,p_n \qquad \forall 0 < x < 1$$

$$\sum_{j=1}^{n}\int_{\{|X_{nj}|<\delta\}} |X_{nj}|dP = 0 \qquad \forall 0 < \delta < 1$$

Let $0 < \lambda < \infty$ be given. Then by (5.24) we find

(5.33.2) If $\lim_{n\to\infty} np_n = \lambda$, then $S_n \xrightarrow{\sim} Po(\lambda)$

Note that $S_n = s_n U_n + np_n \sim Bi(n, p_n)$. Thus, by (5.33.1) we have that $Bi(n,p)$ is approximately equal to $N(np, np(1-p))$ if $np(1-p)$ is large. Similarly, by (5.33.2) we have that $Bi(n,p)$ is approximately equal to $Po(np)$ if np is neither too large nor too small.

The normal approximation $N(np, np(1-p))$ of $Bi(n,p)$ is good when $np(1-p)$ is not too small, that is, if n is large and p is not close to 0 or 1. The Poisson approximation $Po(np)$ of $Bi(n,p)$ is good when np is not too large, that is, if n or p is small. For this reason (5.33.2) goes under the name *the law of small numbers.*

5.34 Normal Approximation of $Po(\lambda)$

Let $X_{n1}, \ldots, X_{nn}$ be independent random variables that all have a Poisson $Po(\lambda_n)$-distribution. Let $s_n = \sqrt{n\lambda_n}$ and set

$$S_n = \sum_{j=1}^{n} X_{nj} \qquad , \qquad U_n = s_n^{-1} \sum_{j=1}^{n} (X_{nj} - \lambda_n)$$

Note that $EX_{nj} = \operatorname{var} X_{nj} = \lambda_n$. Since the fourth central moment of $Po(\lambda)$ equals $\lambda + 3\lambda^2$, we have

$$s_n^{-4} \sum_{j=1}^{n} E|X_{nj} - \lambda_n|^4 = (n\lambda_n)^{-1} + 3n^{-1}$$

Hence, if $n\lambda_n \to \infty$, then Lyapounov's condition (5.22.3) holds and so by (5.22) we have

(5.34.1) If $\lim_{n\to\infty} n\lambda_n = \infty$, then $U_n \xrightarrow{\sim} N(0,1)$

Since $S_n = s_n U_n + n\lambda_n \sim Po(n\lambda_n)$, we see that the Poisson distribution $Po(\lambda)$ can be approximated by the normal distribution $N(\lambda, \lambda)$ provided that λ is large.

5.35 The Triangular Distribution

The triangular distribution $Tri(\mu, \lambda)$, where $\mu \in \mathbf{R}$ and $\lambda > 0$, is given by the density

$$(5.35.1) \qquad f(x) = \frac{(\lambda - |x-\mu|)^+}{\lambda^2} \qquad \forall x \in \mathbf{R}$$

Note that $f(x) = 0$ if $|x - \mu| \geq \lambda$. The characteristic function of $Tri(\mu, \lambda)$ is easily computed:

$$\begin{aligned}\varphi(t) &= \int_{-\infty}^{\infty} e^{itx} f(x)dx = \frac{1}{\lambda}\int_{-\infty}^{\infty} \left(1 - \left|\frac{x-\mu}{\lambda}\right|\right)^+ e^{itx}\, dx \\ &= 2e^{it\mu}\int_0^1 (1-y)\cos(\lambda t y)dy = 2e^{it\mu}\,\frac{1-\cos\lambda t}{\lambda^2 t^2}\end{aligned}$$

Hence, we see that φ is Lebesgue integrable on $\mathbf{R}$ so by (5.27.5) we have

$$(5.35.2) \qquad \varphi(t) = 2e^{it\mu}\,\frac{1-\cos\lambda t}{\lambda^2 t^2} \quad , \quad f(x) = \int_{-\infty}^{\infty} e^{it(x-\mu)}\,\frac{1-\cos\lambda t}{\pi\lambda^2 t^2}\, dt$$

for all $x, t \in \mathbf{R}$. Let X be a random variable with characteristic function φ_X . Setting $\mu = 0$ and $x = X$ in (5.35.2) and taking the expectation, we get

$$\begin{aligned}E(\lambda - |X|)^+ &= \frac{1}{\pi} E\left\{\int_{-\infty}^{\infty} e^{itX}\,\frac{1-\cos\lambda t}{t^2}\, dt\right\} \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \varphi_X(t)\,\frac{1-\cos\lambda t}{t^2}\, dt\end{aligned}$$

where the interchanging of the integrals is justified by Fubini's theorem (3.21.2). Since the left-hand side is real and $(\lambda - |X|)^+ = \lambda - \lambda \wedge |x|$, we find

$$(5.35.3) \qquad E(\lambda - |X|)^+ = \frac{2}{\pi}\int_0^{\infty} \frac{1-\cos\lambda t}{t^2}\, \Re\varphi_X(t)dt$$

$$(5.35.4) \qquad E(\lambda \wedge |X|) = \frac{2}{\pi}\int_0^{\infty} \frac{1-\cos\lambda t}{t^2}(1 - \Re\varphi_X(t))dt$$

Moreover, since $(\lambda - |X|)^+ \leq \lambda\, 1_{\{|X|\leq\lambda\}} \leq (2\lambda - |X|)^+$, we have

$$(5.35.5) \qquad P(|X| \leq \lambda) \leq \frac{2}{\lambda\pi}\int_0^{\infty} \frac{1-\cos(2\lambda t)}{t^2}\, \Re\varphi_X(t)dt$$

$$(5.35.6) \qquad P(|X| \leq \lambda) \geq \frac{2}{\lambda\pi}\int_0^{\infty} \frac{1-\cos(\lambda t)}{t^2}\, \Re\varphi_X(t)dt$$

(5.35.7) $$P(|X| > \lambda) \le \frac{2}{\lambda\pi}\int_0^\infty \frac{1-\cos(\lambda t)}{t^2}\,(1-\Re\varphi_X(t))dt$$

(5.35.8) $$P(|X| > \lambda) \ge \frac{2}{\lambda\pi}\int_0^\infty \frac{1-\cos(2\lambda t)}{t^2}\,(1-\Re\varphi_X(t))dt$$

Setting $x = \mu$ and $\lambda = 1$ in (5.35.2), we see that $g(t) = (1-\cos t)/(\pi t^2)$ is a density function with characteristic function $\psi(t) = (1-|t|)^+$. Let $k \in \mathbf{Z}$ be an integer, then by (5.35.2) we have:

$$p_k := \int_{-1}^{1} (1-|t|)\, e^{-itk\pi}\, dt = \begin{cases} \frac{1}{2} & \text{if } k = 0 \\ \dfrac{1}{\pi^2 k^2} & \text{if } k \text{ is odd} \\ 0 & \text{if } k \text{ is even and } k \ne 0 \end{cases}$$

Note that $p_k \ge 0$ and $\sum_{-\infty}^{\infty} p_k = 1$. Hence, there exists a discrete random variable Y such that $P(Y = k\pi) = p_k$ for all $k \in \mathbf{Z}$. By (5.27) we have that φ_Y is periodic with period 2 and $\varphi_Y(t) = (1-|t|)$ for all $|t| \le 1$. Thus, we here have an example of two very different distributions [one is absolutely continuous, the other is discrete] whose characteristic functions coincide on the interval $[-1,1]$.

5.36 Fractiles

Let $F : \mathbf{R} \to [0,1]$ be a distribution function. If $0 \le p \le 1$, we say that $x \in \mathbf{R}$ is a *p-fractile* of F if $F(x-) \le p \le F(x)$. The set of *all p-fractiles* of F is denoted $F^{(-1)}(p)$, that is,

(5.36.1) $$F^{(-1)}(p) = \{x \in \mathbf{R} \mid F(x-) \le p \le F(x)\}$$

Let us define $\hat{F}$ and $\check{F}$ as follows [see (3.29)]:

(5.36.2) $$\hat{F}(p) = \inf\{x \in \mathbf{R} \mid F(x) \ge p\} \qquad \forall 0 \le p \le 1$$

(5.36.3) $$\check{F}(p) = \inf\{x \in \mathbf{R} \mid F(x) > p\} \qquad \forall 0 \le p \le 1$$

with the usual convention: $\inf\emptyset = \infty$. Then $\hat{F}$ and $\check{F}$ are increasing functions from $[0,1]$ into $\overline{\mathbf{R}}$ such that $\hat{F}(0) = -\infty$, $\check{F}(1) = \infty$. Moreover, it is easily verified that we have [see for instance, (3.29.1)–(3.29.3)]

(5.36.4) $$F^{(-1)}(p) = [\hat{F}(p), \check{F}(p)]$$

(5.36.5) $$-\infty < \hat{F}(p) \le \check{F}(p) \le \hat{F}(q) < \infty \qquad \text{if} \qquad 0 < p < q < 1$$

(5.36.6) $\quad p \le F(x) \iff \hat{F}(p) \le x$

(5.36.7) $\quad F(x-) \le p \iff x \le \check{F}(p)$

Thus, $\hat{F}(p)$ is *the smallest* and $\check{F}(p)$ is *the largest p-fractile* of F .

Let $\chi(p)$ be a p-fractile of F for all $p \in]0,1[$. By (5.36.4) we have that $\hat{F}(p) \le \chi(p) \le \check{F}(p)$. Hence, by (5.36.5) we have that χ is increasing. Let Y have a rectangular distribution $R(0,1)$ [see (2.24)] and let $X = \chi(Y)$. Then by (5.36.6) and (5.36.7) we have

$$\begin{aligned} F(a-) &= P(Y < F(a-)) \le P(\chi(Y) < a) = F_X(a-) \\ &\le F_X(a) = P(\chi(Y) \le a) \le P(Y \le F(a)) = F(a) \end{aligned}$$

In particular, we see that $F(a) = F_X(a)$ whenever a is a continuity point of F . Hence, by right continuity of F and F_X we have that $F_X = F$. Thus, we obtain the following extension of (3.29.3):

(5.36.8) Let $\chi(p)$ be a p-fractile of F for all $0 < p < 1$, and let Y be a random variable with $Y \sim R(0,1)$. Then $\chi(Y) \sim F$ and for every Borel function $h : \mathbf{R} \to \mathbf{R}$ we have

$$\int_{-\infty}^{\infty} h(x)\, F(dx) = \int_0^1 h(\chi(p))\, dp$$

in the usual sense: If one of the integrals exists, then so does the other, and they are equal

5.37 Convergence of Fractiles

Let $\{F_n\}$ be a sequence of distribution functions such that $F_n \xrightarrow{\sim} F$, and let $\chi_n(p)$ be a p-fractile of F_n for all $n \ge 1$ and all $0 \le p \le 1$. If $p_n \in [0,1]$ and $p_n \to p$, then we have

(5.37.1) $\quad \hat{F}(p) \le \liminf_{n\to\infty} \chi_n(p_n) \le \limsup_{n\to\infty} \chi_n(p_n) \le \check{F}(p)$

In particular, if F has a unique p-fractile $\chi(p)$, that is, if $\hat{F}(p) = \check{F}(p) = \chi(p)$, then

(5.37.2) $\quad \lim_{n\to\infty} \chi_n(p_n) = \chi(p)$

Proof: Let $a < \hat{F}(p)$. Then $F(a) < p$ by (5.36.6). So by (5.9.2) there exists $k \in \mathbf{N}$ such that $F_n(a) < p_n$ for all $n \geq k$. Hence, by (5.36.6) we have $a < \hat{F}_n(p_n) \leq \chi_n(p_n)$ for all $n \geq k$ and so $a \leq \liminf \chi_n(p_n)$ for all $a < \hat{F}(p)$. But then $\hat{F}(p) \leq \liminf \chi_n(p_n)$. The other inequality is proved similarly. □

5.38 Convergence of Empirical Fractiles

Let $\{X_n\}$ be a sequence of independent identically distributed random variables with distribution function F . Let $X_{1n}, \ldots, X_{nn}$ be the order statistics of $X_1, \ldots, X_n$ [see (2.23)] and set $X_{0n} = -\infty$ and $X_{n+1,n} = \infty$. Suppose that $\sigma(1), \sigma(2), \ldots$ is a sequence of integers such that $0 \leq \sigma(n) \leq n+1$ for all $n = 1, 2, \ldots$ and $\lim_{n\to\infty} \frac{\sigma(n)}{n} = p$ for some $p \in [0,1]$. Then we have

$$\hat{F}(p) \leq \liminf_{n\to\infty} X_{\sigma(n),n} \leq \limsup_{n\to\infty} X_{\sigma(n),n} \leq \check{F}(p) \quad a.s. \tag{5.38.1}$$

In particular, if F has a unique p-fractile χ , then

$$\lim_{n\to\infty} X_{\sigma(n),n} = \chi \quad a.s. \tag{5.38.2}$$

Suppose that $p = F(\chi)$ and that F is differentiable at χ with derivative $\lambda := F'(\chi) > 0$. If $\lim_{n\to\infty} \sqrt{n}\,(\frac{\sigma(n)}{n} - p) = c$ for some $c \in \mathbf{R}$, then we have

$$\sqrt{n}\,(X_{\sigma(n),n} - \chi) \xrightarrow{\sim} N\left(\frac{c}{\lambda}, \frac{p(1-p)}{\lambda^2}\right) \tag{5.38.3}$$

Proof: (5.38.1+2): Let $p_n = 1 \wedge \frac{\sigma(n)}{n}$ and let $F_n(t,\omega)$ be the empirical distribution function of $X_1, \ldots, X_n$. Then $F_n(X_{\sigma(n),n}(\omega), \omega) = p_n$, and so $X_{\sigma(n),n}(\omega)$ is a p_n-fractile of $F_n(\cdot, \omega)$ for all ω . By the law of large numbers (4.12) there exists a P-null set N such that $F_n(x,\omega) \to F(x)$ for all $\omega \in \Omega \setminus N$ and all rational numbers x . Hence, by (5.9) we have that $F_n(\cdot,\omega) \xrightarrow{\sim} F$ for a.a. ω. Thus, (5.38.1) follows from (5.37.1) because $p_n \to p$.

(5.38.3): Suppose that $\sqrt{n}\,(p_n - p) \to c$, $F(\chi) = p$ and that F is differentiable at χ with $\lambda := F'(\chi) > 0$. By strict positivity of $F'(\chi)$, we have that $0 < p < 1$. Thus, there exists $k \in \mathbf{N}$ such that $1 < \sigma(n) < n$ for all $n \geq k$. Throughout the rest of this proof let $n \geq k$ and let $a \neq 0$ be a given number. Set

$$\chi_n = \sqrt{n}\,(X_{\sigma(n),n} - \chi) \quad , \quad a_n = \tfrac{a}{\sqrt{n}} \quad , \quad c_n = \sqrt{n}\,(p_n - p)$$
$$\beta_n = F(\chi + a_n) - p \quad , \quad V_n = \sqrt{n}\,(F_n(\chi) - p)$$
$$U_n = \sqrt{n}\,(F_n(\chi + a_n) - F_n(\chi)) \quad , \quad W_n = \sqrt{n}(F_n(\chi + a_n) - p_n)$$

Then $\sqrt{n}\,U_n \sim Bi(n,\beta_n)$ if $a > 0$, and $-\sqrt{n}\,U_n \sim Bi(n,-\beta_n)$ if $a < 0$ [see (2.23.2)]. Thus, we have

$$EU_n = \sqrt{n}\,\beta_n = \frac{a(F(\chi + a_n) - F(\chi))}{a_n} \quad , \quad \operatorname{var} U_n = |\beta_n|(1 - |\beta_n|)$$

Since $a_n \to 0$ and F is differentiable and, thus, continuous at χ , we have that $EU_n \to \lambda a$ and $\operatorname{var} U_n \to 0$. Hence, we see that $U_n \to a\lambda$ in 2-mean and in probability. Note that $F_n(t,\omega) \geq \frac{j}{n}$ if and only if $X_{jn}(\omega) \leq t$. Hence, we find

$$\begin{aligned} P(\chi_n \leq a) &= P(X_{\sigma(n),n} \leq \chi + a_n) = P(F_n(\chi + a_n) \geq p_n) \\ &= P(W_n \geq 0) \end{aligned}$$

Recall that $nF_n(\chi) \sim Bi(n,p)$ [see (2.23.2)]. Hence, if $V \sim N(0,p(1-p))$, then $V_n \xrightarrow{\sim} V$ by (5.33.1). By assumption, we have $c_n \to c$, and since $W_n = U_n - c_n + V_n$, then by (5.4.5) we have $W_n \xrightarrow{\sim} \lambda a - c + V$. Moreover, by (4.22.16) and symmetry of V we get

$$\lim_{n\to\infty} P(\chi_n \leq a) = P(\lambda a - c - V \geq 0) = P\left(\frac{V+c}{\lambda} \leq a\right)$$
$$\frac{V+c}{\lambda} \sim N\left(\frac{c}{\lambda}, \frac{p(1-p)}{\lambda^2}\right)$$

Thus, (5.38.3) is proved. □

Remark: If $x \in \mathbf{R}$, we let $i(x)$ and $j(x)$ denote the smallest integer k with $k > x$ respectively the smallest integer with $k \geq x$. That is, $i(x) = 1 + int(x)$ for all $x \in \mathbf{R}$, and $j(x) = i(x)$ for $x \notin \mathbf{Z}$ and $j(x) = x = int(x)$ for $x \in \mathbf{Z}$. Let $\hat{F}_n(p,\omega)$ and $\check{F}_n(p,\omega)$ be the minimal respectively maximal p-fractile of the empirical distribution $F_n(\cdot,\omega)$. Then by (2.23.1) we have

(5.38.4) $$\hat{F}_n(p,\omega) = X_{j(np),n}(\omega) \quad , \quad \check{F}_n(p,\omega) = X_{i(np),n}(\omega))$$

A random variable χ_n is called an *empirical p-fractile* of $X_1,\ldots,X_n$ if $\chi_n(\omega)$ is an empirical p-fractile of the empirical distribution function $F_n(\cdot,\omega)$ for all $\omega \in \Omega$, or, equivalently if

(5.38.5) $$X_{j(np),n}(\omega) \leq \chi_n(\omega) \leq X_{i(np),n}(\omega) \qquad \forall \omega \in \Omega$$

Let χ_n be an empirical p-fractile of $X_1, \ldots, X_n$ for all $n \geq 1$. Note that $|\frac{i(np)}{n} - p| \leq \frac{1}{n}$ and $|\frac{j(np)}{n} - p| \leq \frac{1}{n}$ for all $0 \leq p \leq 1$ and all $n \geq 1$. Hence, if $j(np) \leq \sigma(n) \leq i(np)$, then $\lim_{n\to\infty} \sqrt{n}\,(\frac{\sigma(n)}{n} - p) = 0$. Thus, by (5.38.1) and (5.38.3) we get

$$\hat{F}(p) \leq \liminf_{n\to\infty} \chi_n \leq \limsup_{n\to\infty} \chi_n \leq \check{F}(p) \qquad P \text{ a.s.} \tag{5.38.6}$$

$$F'(\chi) > 0 \ , \ p = F(\chi) \quad \Rightarrow \quad \sqrt{n}\,(\chi_n - F(\chi)) \xrightarrow{\sim} N\Big(0, \tfrac{p(1-p)}{F'(\chi)^2}\Big) \tag{5.38.7}$$

5.39 References to Chapter 5

Many textbooks uses the terminology *weak convergence* for convergence in law. The concept of convergence in law in general metric spaces was introduced by Alexandrov in a series of papers (see [1]). In [2] you will find a modern approach to convergence in law of probability measures on metric spaces. The implication: "(5.22.1) $\Rightarrow$ (5.22.2)" was found by Lindeberg (see [5]). It can be shown that the converse implication holds whenever $\max(\sigma_{n1}, \ldots, \sigma_{nn}) \to 0$ as $n \to \infty$; that is, the Lindeberg condition (5.22.1) is in essence equivalent to the convergence of the normalized sums in (5.22.2). Gnedenko and Kolmogorov's book [4] is a classical textbook on the central limit theorem for real random variables. Petrov's book [6] is an updated version of [4] and contains a wealth of information about convergence in law of normalized sums of random variables. Evidently, we may extend the notion of a.s. convergence to non-measurable random functions, and from the definition of convergence in *pr* given in Section 5.1, we see that also this notion can be applied to non-measurable random functions. However, if $X, X_1, X_2, \ldots$ is a sequence of arbitrary functions from the probability space $(\Omega, \mathcal{F}, P)$ into the metric space S, then definition (5.1.3) can no longer be applied to define convergence in law of $\{X_n\}$ to X. If the metric space S is non-separable, then measurability of the random functions $X_1, X_2, \ldots$ turns out to very restrictive, and even very simple function may not be measurable. In [3] it is shown that convergence in law may be extended to arbitrary functions such that the inversion theorem (4.15) and the continuity theorem (5.7) for multiplicative transforms remains valid.

[1] Alexandrov, A. D. (1940, 1941, 1943): *Additive Set Functions in Abstract Spaces*, Mat. Sbornik (n.s.) 8, pp. 307–348; Mat. Sbornik (n.s.) 9, pp. 563–628; Mat. Sbornik (n.s.) 13, pp. 169–238

[2] Billingsly, P. (1968): *Convergence of Probability Measures*, John Wiley & Sons, Inc., New York and London

[3] Hoffmann-Jørgensen, J. (1991): *Stochastic Processes on Polish Spaces*, Mat. Inst., Aarhus University, Var. Publ. Ser. 39, pp. 1–278

[4] Gnedenko, B. V. and Kolmogorov, A. N. (1968): *Limit Distributions for Sums of Independent Random Variables*, (second ed.) Addison-Wesley, Reading, Mass.

[5] Lindeberg, J. W. (1922): *Eine Neue Herleitung des Exponetialgestzes in der Wahrscheinlichkeitsrechnung*, Math. Zeitsschr. 15, pp. 211–225

[6] Petrov, V. V. (1975): *Sums of Independent Random Variables*, Springer Verlag, Berlin and New York

5.40 Exercises to Chapter 5

Exc. 5.1: Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with common distribution function $F(x)$. Suppose that $F(0) = 0$ and that the right-hand derivative $\alpha = \lim_{x \downarrow 0} \frac{F(x)}{x}$ exists and is finite and strictly positive. Let $X_{1n}, \ldots, X_{nn}$ be the order statistics of $X_1, \ldots, X_n$, and let $k \in \mathrm{N}$ be a fixed integer. Show that $n X_{kn} \overset{\sim}{\to} \Gamma\big(k, \frac{1}{\alpha}\big)$ where $\Gamma(a, b)$ denotes the gamma distribution. [Apply (2.23).]

Exc. 5.2: Let $X, X_1, X_2, \ldots$ be real random variables such that $X_n \overset{\sim}{\to} X$ and $|X_n| \le c$ a.s. for all $n \ge 1$ where c is some positive finite constant. Show that $\lim\limits_{n \to \infty} EX_n^k = EX^k$ for all $k = 1, 2, \ldots$.

Exc. 5.3: Let μ be a Borel measure on the metric spaces (S, d) and let $X_0, X_1, X_2, \ldots$ be S-valued random variables such that X_n has μ-density f_n for all $n = 0, 1, 2, \ldots$. Suppose that $\liminf\limits_{n \to \infty} f_n = f_0$ μ-a.s. and show that $X_n \overset{\sim}{\to} X_0$. [Apply Exc.3.5.]

Exc. 5.4: Let X and Y be independent random variables and let $U = X - Y$ and $V = XY$. Let φ, ψ and ζ be the characteristic functions of X, U and V, respectively. Show that:

(i) $X \sim Y \quad \Rightarrow \quad \psi(t) = |\varphi(t)|^2 \qquad \forall t \in \mathbf{R}$

(ii) $\Re\varphi(t) = \frac{1}{2}\left(\varphi(t) + \varphi(-t)\right) = E(\cos(tX))$

(iii) $P(Y = 1) = P(Y = -1) = \frac{1}{2} \quad \Rightarrow \quad \zeta(t) = \Re\varphi(t) \qquad \forall t \in \mathbf{R}$

Exc. 5.5: Let X be a random variable with characteristic function φ and suppose that

$$(*) \qquad m = \liminf_{t \to 0} \frac{2 - \varphi(t) - \varphi(-t)}{t^2} < \infty$$

Show that $E X^2 < \infty$ and that φ is twice continuous differentiable on $\mathbf{R}$ with $\varphi''(0) = -E X^2 = -m$

Hint: Show that [apply l'Hospital's rule to prove the second equality]

$$\frac{2 - \varphi(t) - \varphi(-t)}{t^2} = 2\,E\Big(\frac{1 - \cos(t\,X)}{t^2}\Big) \quad , \quad \lim_{t \to 0} \frac{1 - \cos(t\,X)}{t^2} = \frac{X^2}{2}$$

Then apply (*) and Fatou's lemma (3.5) to prove that $EX^2 < \infty$.

Exc. 5.6: Let φ and ψ be a characteristic functions. Recall that $\overline{z}$ denotes the complex conjugate of z . Show the following [apply Exc.5.5 to prove (ii)]:

(i) The functions: $|\varphi(t)|^2$, $\Re\varphi(t)$, $\varphi \cdot \psi$ and $\overline{\varphi(t)}$ are characteristic functions, but $\Im\varphi(t)$ is not a characteristic function

(ii) $\liminf_{t \to 0} \dfrac{1 - \Re\,\varphi(t)}{t^2} = 0 \quad \Rightarrow \quad \varphi(t) = 1 \qquad \forall\, t \in \mathbf{R}$

(iii) The function $e^{-|t|^q}$ is not a characteristic function for any $q > 2$

Remark: If $0 < q \leq 2$, then by (5.18.3) we have that $e^{-|t|^q}$ is the characteristic function of $Stab(0, 1, q, 0)$.

Exc. 5.7: Let X be a random variable with characteristic function φ . Show that

(i) $$E\,|X| = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \Re\,\varphi(t)}{t^2}\, dt$$

Hint: Use the well-known formula [see (3.37.14) and (3.37.16)]

$$\int_{-\infty}^{\infty} \frac{1 - \cos(tx)}{t^2}\, dt = \pi\,|x| \qquad \forall\, x \in \mathbf{R}$$

Exc. 5.8: Let X be a random variable with characteristic function φ . Show that X is symmetric if and only if $\varphi(t)$ is real for all $t \in \mathbf{R}$; and if so then $\varphi(t) = E\,\cos(tX)$.

Exc. 5.9: Let X be a random variable with characteristic function φ and let $\alpha \neq 0$ and $\theta \in \mathbf{R}$ given numbers. Let $\beta = \frac{2\pi}{\alpha}$, and show that the following statements are equivalent:

(a) $X(\omega) \in \{\theta + k\,\alpha \mid k = 0, \pm 1, \pm 2, \dots\}$ for P-a.a. $\omega \in \Omega$

(b) $t \curvearrowright e^{-it\theta}\varphi(t)$ is periodic with period β

(c) $e^{-i\theta\beta}\,\varphi(\beta) = 1$

Suppose that (a) holds and let

$$p_k = P(X = \theta + \alpha k) \quad , \quad q_k = p_k + p_{-k} \quad , \quad r_k = p_k - p_{-k}$$

Show that

(i) $$e^{-it\theta}\,\varphi(t) = \sum_{k=-\infty}^{\infty} p_k\, e^{itk\alpha} = \tfrac{1}{2}\, q_0 + \sum_{k=1}^{\infty}\{q_k \cos(\alpha k t) + i r_k \sin(\alpha k t)\}$$

Hint: The implications: (a) $\Rightarrow$ (b) $\Rightarrow$ (c) are easy. Suppose that (c) holds and show that

$$E\, e^{i\beta(X-\theta)} = 1 \quad , \quad E(1 - \cos(\beta(X-\theta))) = 0$$

Conclude that $\cos(\beta X - a\theta) = 1$ a.s. and recall that $\cos x = 1$ if and only if x is an integer multiple of 2π . Argue that this proves (a).

Exc. 5.10: Let X and Y be independent random variables. Show that

(i) $sp_+(X+Y) = sp_+(X) + sp_+(Y)$

Recall that $sp_+(X)$ is the point spectrum of X [see (2.14)] and that $A + B = \{a + b \mid a \in A\ ,\ b \in B\}$ whenever $A, B \subseteq \mathbf{R}$.

Exc. 5.11: Let X and Y be independent random variables, and suppose that $Z = X + Y$ has a two-point distribution, that is, there exists $a \neq b$ such that

$$P(Z = a) + P(Z = b) = 1$$

Show that either X or Y are degenerated [apply Exc.5.10]. Conclude that the binomial distribution $Bi(1,p)$ is not infinitely divisible for any $0 < p < 1$.

Exc. 5.12: Let φ be an infinitely divisible characteristic function. Show that $\varphi(t) \neq 0$ for all $t \in \mathbf{R}$.

Hint: Let φ_n be a characteristic function such that $\varphi(t) = \varphi_n(t)^n$ for all $t \in \mathbf{R}$ and all $n \geq 1$, and let $\psi_n(t) := |\varphi_n(t)|^2$. Show that ψ_n is a characteristic function such that

$$\lim_{n\to\infty} \psi_n(t) = \psi(t) := \begin{cases} 1 & \text{if } \varphi(t) \neq 0 \\ 0 & \text{if } \varphi(t) = 0 \end{cases}$$

for all $t \in \mathbf{R}$. Show that ψ is continuous at 0 . Then use the continuity theorem (5.10) to conclude that ψ is a characteristic function and, thus, continuous on $\mathbf{R}$. Argue, that this solves the exercise.

Exc. 5.13: Let $f : \mathbf{R} \to \mathbf{C}$ be a continuous complex-valued function satisfying

(a) $\quad f(0) = 1 \quad \text{and} \quad f(t) \neq 0 \qquad \forall t \in \mathbf{R}$

Then using the implicit function theorem, it can be shown that there exists one and only one continuous function $g : \mathbf{R} \to \mathbf{C}$ satisfying

(b) $\quad g(0) = 0 \quad \text{and} \quad f(t) = e^{g(t)} \qquad \forall t \in \mathbf{R}$

The unique continuous function g satisfying (b) is denoted $\operatorname{Log} f := g$ and is called *the main logarithm* of f . If $\alpha \in \mathbf{C}$, we define *the α-main root* of the function f , denoted $f^\alpha(t)$, as follows:

(c) $\quad f^\alpha(t) := e^{\alpha \operatorname{Log} f(t)} \qquad \forall t \in \mathbf{R}\ \forall \alpha \in \mathbf{C}$

Let $f, g : \mathbf{R} \to \mathbf{C}$ be continuous complex-valued functions satisfying (a) and let $\alpha, \beta \in \mathbf{C}$ be given complex numbers. Show the following:

(i) $\quad \operatorname{Log} f(0) = 0 \quad , \quad \Re \operatorname{Log} f(t) = \log(|f(t)|)$

(ii) $\quad \operatorname{Log}(f \cdot g)(t) = \operatorname{Log} f(t) + \operatorname{Log} g(t) \quad , \quad \operatorname{Log} f^\alpha(t) = \alpha \operatorname{Log} f(t)$

(iii) $\quad f^\alpha(0) = 1 \quad , \quad f^{-1}(t) = \dfrac{1}{f(t)} \quad , \quad f^1(t) = f(t) \quad , \quad f^0(t) \equiv 1$

(iv) $\quad (f \cdot g)^\alpha(t) = f^\alpha(t) \cdot g^\alpha(t) \quad , \quad f^{\alpha+\beta}(t) = f^\alpha(t) \cdot f^\beta(t)$

(v) $\quad |f^\alpha(t)| = |f(t)|^\alpha \quad , \quad (f^\alpha)^\beta(t) = f^{\alpha\cdot\beta}(t)$

(vi) $\quad \lim\limits_{\alpha\to 0} \dfrac{f^\alpha(t) - 1}{\alpha} = \operatorname{Log} f(t)$

(vii) $\quad |f^\alpha(t) - 1| \leq |\alpha \operatorname{Log} f(t)|\, e^{|\alpha \operatorname{Log} f(t)|}$

(viii) The function $(\alpha, t) \curvearrowright f^\alpha(t)$ is continuous on $\mathbf{C} \times \mathbf{R}$

Exc. 5.14: Let $\varphi, \psi, \varphi_1, \varphi_2, \ldots$ be infinite divisible characteristic functions. By Exc.5.12, we have that ψ is a continuous complex-valued function satisfying (a) in Exc.5.13, so let ψ^α be defined as in Exc.5.13. As usual we let $\overline{z}$ denotes the complex conjugated of z. Show the following:

(i) $|\varphi(t)|^2$, $\overline{\varphi(t)}$ and $\varphi(t) \cdot \psi(t)$ all are infinite divisible characteristic functions

(ii) $\psi^{1/n}(t)$ is the unique characteristic function satisfying $\psi^{1/n}(t)^n = \psi(t)$ for all $t \in \mathbf{R}$ and all $n = 1, 2, \ldots$

(iii) If the limit $\lim_{n\to\infty} \varphi_n(t) = \zeta(t)$ exists for all $t \in \mathbf{R}$ and ζ is continuous at 0, then ζ is an infinite divisible characteristic function

(iv) ψ^α is an infinite divisible characteristic function for all $\alpha \geq 0$

Hint: (ii): Let ψ_n be a characteristic function such that $\psi_n(t)^n = \psi(t)$. Apply (ii) in Exc.5.13 to show that $\operatorname{Log}\psi_n(t) = \frac{1}{n}\operatorname{Log}\psi(t)$. Conclude that $\psi_n = \psi^{1/n}$. (iii): Apply (viii) in Exc.5.13 and (ii). (iv): Apply (i) and (ii) to show that ψ^α is an infinite divisible characteristic function for all rational numbers $\alpha > 0$. Then apply (viii) in Exc.5.13 and (iii) to prove that ψ^α is an infinite divisible characteristic function for all $\alpha \geq 0$.

Exc. 5.15: A function $f : \mathbf{R} \to \mathbf{C}$ is said to be *n-times differentiable* at a with *derivatives* $c_0, \ldots, c_n$ if we can write f in the form:

(a) $$f(t) = \sum_{j=0}^{n} \frac{c_j}{j!}(t-a)^j + R(t-a) \quad \text{where} \quad \lim_{u\to 0} \frac{R(u)}{u^n} = 0$$

The complex numbers $c_0, \ldots, c_n$ are uniquely determined by (a) and are denoted $f^{(j)}(a) := c_j$. Show the following:

(i) If f is n-times differentiable at a, then f is k-times differentiable at a for all $0 \leq k \leq n$

(ii) f is 0-times differentiable at a if and only if f is continuous at a; and if so, then $f^{(0)}(a) = f(a)$

(iii) f is 1-times differentiable at a if and only if f is ordinary differentiable at a; and if so, then $f^{(1)}(a) = f'(a)$

(iv) If f is $(n-1)$-times ordinary differentiable in a neighborhood of a and the $(n-1)^{\text{th}}$ derivative is ordinary differentiable at a, then f is n-times differentiable at a and $f^{(1)}(a), \ldots, f^{(n)}(a)$ are the ordinary derivatives of f

Let $f, g : \mathbf{R} \to \mathbf{C}$ be n-times differentiable at a and let $h(t) = f(t)\,g(t)$. Show that h is n-times differentiable at a and that we have the following version of *Leibniz' multiplication rule*:

(v) $$h^{(j)}(a) = \sum_{i=0}^{j} \binom{j}{i} f^{(i)}(a)\, g^{(j-i)}(a) \qquad \forall j = 0, \ldots, n$$

Moreover, show that if f, g satisfies

(b) $$f^{(j)}(a) = g^{(j)}(a) = 0 \quad \text{for } j = 0, \ldots, n-1 \quad \text{and} \quad f^{(n)}(a) \neq 0$$

then we have the following version of *l'Hospital's rule*:

(vi) $$\lim_{t \to a} \frac{g(t)}{f(t)} = \frac{g^{(n)}(a)}{f^{(n)}(a)}$$

Exc. 5.16: *(The Lévy-Khinchine representation)* Let F be an infinitely divisible distribution function with characteristic function φ and let F_α be the distribution function having φ^α as its characteristic function for $\alpha \geq 0$ [see Exc.5.14]. Let $q : \mathbf{R} \to \mathbf{R}$ be a bounded continuous function such that q is twice differentiable at 0 with:

(a) $$q(0) = 0 \quad , \quad q'(0) = 1 \quad , \quad q''(0) = 0$$

Show that there exist numbers $\mu \in \mathbf{R}$ and $0 \leq \sigma < \infty$ and a Borel measure ν on $\mathbf{R}$ such that

(i) If $\alpha_n \to \alpha$, then $F_{\alpha_n} \overset{\sim}{\to} F_\alpha$

(ii) (q, ν) is a Lévy-Khinchine pair and $F = Lev(\mu, \sigma, q, \nu)$

[see (5.17.6)–(5.17.9)]. Moreover, show that (μ, σ, ν) are uniquely given by

(iii) $$\mu = \lim_{\alpha \downarrow 0} \left\{ \frac{1}{\alpha} \int_{-\infty}^{\infty} q(x)\, F_\alpha(dx) \right\}$$

(iv) $$\sigma^2 = \lim_{n \to \infty} \left(\lim_{\alpha \downarrow 0} \left\{ \frac{1}{\alpha} \int_{-\infty}^{\infty} x^2\, e^{-nx^2}\, F_\alpha(dx) \right\} \right)$$

(v) $$\lim_{\alpha \downarrow 0} \left\{ \frac{1}{\alpha} \int_{-\infty}^{\infty} (f(x) - f(0))\, F_\alpha(dx) \right\}$$
$$= \mu\, f'(0) + \tfrac{1}{2}\sigma^2 f''(0) + \int_{-\infty}^{\infty} (f(x) - f(0) - f'(0)\, q(x))\, \nu(dx)$$

for every bounded continuous function $f : \mathbf{R} \to \mathbf{C}$ which is twice differentiable at 0 .

Hint: (i): Follows from the continuity theorem (5.10) and Exc.5.13(viii). (ii)–(v): The proof consists of 17 small steps. Let C_0^2 denote the set of all bounded continuous functions $f : \mathbf{R} \to \mathbf{C}$, which are twice differentiable at 0 and set

$$\psi_\alpha(t) = \tfrac{1}{\alpha}\left(\varphi^\alpha(t) - 1\right) \ , \ \psi(t) = \operatorname{Log}\varphi(t) \ , \ \eta(t) = \int_{-1}^{1} \frac{\psi(t) - \psi(t+y)}{2}\, dy$$

$$p(x) = \frac{1}{2}\int_{-1}^{1}\left(1 - e^{ixy}\right) dy = 1 - \frac{\sin x}{x} \ , \ \xi_\alpha(B) = \frac{1}{\alpha}\int_B p(x)\, F_\alpha(dx)$$

Let $t, x \in \mathbf{R}$ and $\alpha > 0$ be given. Show the following:

(1) $\frac{19}{120}\left(x^2 \wedge 1\right) \le p(x) \le 2\left(x^2 \wedge 1\right)$

(2) ξ_α is a finite Borel measure on $\mathbf{R}$

(3) $\hat{\xi}_\alpha(t) = \displaystyle\int_{-1}^{1} \frac{\psi_\alpha(t) - \psi_\alpha(t+y)}{2}\, dy$

(4) $\lim\limits_{\alpha \downarrow 0} \psi_\alpha(t) = \psi(t)$ [see (vi) in Exc.5.13]

(5) $|\psi_\alpha(t)| \le |\psi(t)|\, e^{|\psi(t)|} \quad \forall\, 0 < \alpha \le 1$ [see (vii) in Exc.5.13]

(6) $\lim\limits_{\alpha \downarrow 0} \hat{\xi}_\alpha(t) = \eta(t)$

where $\hat{\xi}_\alpha(t)$ is the Fourier transform of ξ_α . Apply the continuity theorem (5.10) to conclude that there exists a finite Borel measure ξ on $\mathbf{R}$ satisfying

(7) $\hat{\xi}(t) = \eta(t)$ and $\xi_\alpha \overset{\sim}{\to} \xi$ as $\alpha \downarrow 0$

Let $f \in C_0^2$ be given and set

$$\sigma^2 = 6\,\xi(\{0\}) \ , \ g(x) = f(x) - f(0) - f'(0)\, q(x)$$

$$\nu(B) = \int_{B\setminus\{0\}} \frac{1}{p(x)}\, \xi(dx) \ , \ h(x) = \begin{cases} \dfrac{g(x)}{p(x)} & \text{if } x \ne 0 \\ 3\, f''(0) & \text{if } x = 0 \end{cases}$$

$$\mu = i\left\{ -\tfrac{1}{2}\sigma^2 + \int_{-\infty}^{\infty} \left(e^{ix} - 1 - iq(x)\right) \nu(dx) - \psi(1) \right\}$$

Show the following:

(8) ν is a Borel measure on $\mathbf{R}$, and (q, ν) is a Lévy-Khinchine pair

(9) $g \in C_0^2$ and $g(0) = g'(0) = 0$, $g''(0) = f''(0)$

(10) h is bounded and continuous on $\mathbf{R}$ and $g(x) = p(x)\,h(x)$

(11) $$\frac{1}{\alpha}\int_{-\infty}^{\infty} g(x)\,F_\alpha(dx) = \int_{-\infty}^{\infty} h(x)\,\xi_\alpha(dx)$$

(12) $$\int_{-\infty}^{\infty} h(x)\,\xi(dx) = \tfrac{1}{2}\sigma^2 f''(0) + \int_{-\infty}^{\infty} g(x)\,\nu(dx)$$

(13) $$\int_{-\infty}^{\infty} (f(x) - f(0))\,F_\alpha(dx) = f'(0)\int_{-\infty}^{\infty} q(x)\,F_\alpha(dx) + \int_{-\infty}^{\infty} g(x)\,F_\alpha(dx)$$

Let $g_1(x) = e^{ix} - 1 - i\,q(x)$ and apply (13) with $f(x) = e^{ix}$ to show that

(14) $$\frac{1}{\alpha}\int_{-\infty}^{\infty} q(x)\,F_\alpha(dx) = \frac{i}{\alpha}\int_{-\infty}^{\infty} g_1(x)\,F_\alpha(dx) - i\psi_\alpha(1)$$

Apply (7), (10), (11), (12) and (14) to prove (iii). Apply (iii), (11), (7), (12), and (13) to prove (v). Apply (v) with $f(x) = x^2\,e^{-nx^2}$ and $f(x) = e^{itx}$, where $n \in \mathbf{N}$ and $t \in \mathbf{R}$ are fixed numbers, to show the following:

(15) $$\lim_{\alpha\downarrow 0}\frac{1}{\alpha}\int_{-\infty}^{\infty} x^2\,e^{-nx^2}\,F_\alpha(dx) = \sigma^2 + \int_{-\infty}^{\infty} x^2\,e^{-nx^2}\,\nu(dx)$$

(16) $$\lim_{\alpha\downarrow 0}\frac{1}{\alpha}\int_{-\infty}^{\infty} \left(e^{itx} - 1\right)F_\alpha(dx)$$
$$= it\mu - \tfrac{1}{2}t^2\sigma^2 + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - it\,q(x)\right)\nu(dx)$$

(17) $$\frac{1}{\alpha}\int_{-\infty}^{\infty} \left(e^{itx} - 1\right)F_\alpha(dx) = \psi_\alpha(t)$$

Finally, show that (iv) follows from (15) and that (ii) follows (4), (16) and (17).

Exc. 5.17: Let $X, X_1, X_2, \ldots$ be random functions with values in the metric space (S, d) and let $Y, Y_1, Y_2, \ldots$ be random functions with values in the **separable** metric space (T, δ) . Show that $(X_n, Y_n) \xrightarrow{\sim} (X, Y)$ if and only if

(i) $$\liminf_{n\to\infty} E\{f(X_n)\,g(Y_n)\} \geq E\{f(X)\,g(Y)\} \quad \forall f \in \mathrm{Lip}_1^+(S)\ \forall g \in \mathrm{Lip}_1^+(T)$$

Warning: Let (S, d) and (T, δ) be two arbitrary metric spaces. Then it can be shown that (i) and (ii) hold if either $\mu(B) := P(X \in B)$ or $\nu(B) := P(Y \in B)$ is a τ-smooth Borel measure. Recall that it leads to no contradictions in mathematics to assume that all Borel probability measures on all

metric spaces are τ-smooth, neither will it lead to any contradictions in mathematics to assume that there exists a non-τ-smooth Borel probability measure on some metric space. If there exists a metric space (S,d) with a non-τ-smooth Borel probability measure, then it is possible to construct $S \times S$-valued random functions $(X,Y),(X_1,Y_1),(X_2,Y_2),\ldots$ satisfying

(a) $X, Y, X_1, Y_1, X_2, Y_2, \ldots$ are independent

(b) $X_n \overset{\sim}{\to} X$ and $Y_n \overset{\sim}{\to} Y$ but $(X_n, Y_n) \not\overset{\sim}{\to} (X,Y)$

Thus, the propositions "(i) $\Rightarrow$ $(X_n,Y_n) \overset{\sim}{\to} (X,Y)$ " and "(5.4.7) holds for arbitrary metric spaces" are independent of the usual axioms of set theory.

Hint: Let G be an open subset of $S \times T$ and let $\{t_1, t_2, \ldots\}$ be a countable dense subset of T . Set

$$G_{jn} = \{\, s \in S \mid \exists U \in \mathcal{N}_s \;:\; U \times b(t_j, 2^{-n}) \subseteq G \,\}$$

where $\mathcal{N}_s$ denotes the set of all open neighborhoods of s . Show that G_{jn} is an open subset of S and

$$G = \bigcup_{j=1}^{\infty} \bigcup_{n=1}^{\infty} G_{jn} \times b(t_j, 2^{-n})$$

Deduce that $1_G \in con_\sigma(\mathcal{W})$ where $\mathcal{W}$ is the set of all functions $h : S \times T \to \mathbf{R}$ of the form $h(s,t) = f(s)\,g(t)$ for $f \in \mathrm{Lip}_1^+(S)$ and $g \in \mathrm{Lip}_1^+(T)$. Then apply (5.3).

Exc. 5.18: Give an example of two sequences $\{X_n\}$ and $\{Y_n\}$ of random variables satisfying

(i) $X_n \overset{\sim}{\to} X$, $Y_n \overset{\sim}{\to} Y$, $X_n + Y_n \not\overset{\sim}{\to} X + Y$

Exc. 5.19: Let X be random variable with distribution function $F(t) = P(X \le t)$. For $0 < p < 1$, let $\hat{F}(p)$ and $\check{F}(p)$ denote the smallest respectively largest p-fractile of F [see (5.36.2+3)]. Show that $\hat{F}(p)$ and $\check{F}(p)$ are increasing on $]0,1[$ and that

(i) $\check{F}(p) \le \hat{F}(q) \qquad \forall 0 < p < q < 1$

(ii) $\check{F}(p-) \le \hat{F}(p) \le \check{F}(p) \le \hat{F}(p+) \qquad \forall 0 < p < 1$

(iii) $\hat{F}$ is left continuous on $]0,1[$, and $\check{F}$ is right continuous on $]0,1[$

(iv) The set $D := \{\, p \in]0,1[\mid \hat{F}(p) \ne \check{F}(p) \,\}$ is countable

where $f(x+) := \lim_{y \downarrow x} f(y)$ and $f(x-) := \lim_{y \uparrow x} f(y)$ denotes the right- and left-hand limits of f at x .

Exc. 5.20: Let $X_0, X_1, X_2, \ldots$ be sequence of random variables, and let $\chi_n(p)$ be a p-fractile of X_n for all $0 < p < 1$ and all $n = 0, 1, 2, \ldots$. Let U be a random variable which is uniformly distributed on $]0,1[$, and let M be the set of all $0 < p < 1$ such that $\chi_0(p)$ is a unique p-fractile of X_0 . Recall that $\chi_n(U) \sim X_n$ [see (5.36.8)] and show that the following four statements are equivalent:

(i) $X_n \overset{\sim}{\to} X_0$

(ii) $\lim_{n \to \infty} \chi_n(p) = \chi_0(p) \qquad \forall p \in M$

(iii) $\chi_n \to \chi_0$ λ-a.s., where λ is the Lebesgue measure on $]0,1[$

(iv) $\chi_n(U) \to \chi_0(U)$ P-a.s.

Hint: (i) $\Rightarrow$ (ii): Follows from (5.37.2). (ii) $\Rightarrow$ (iii): Follows from (iv) in Exc.5.19. (iii) $\Rightarrow$ (iv): Evident. (iv) $\Rightarrow$ (i): Follows from (5.36.8) and (5.4.4).

Exc. 5.21: Let X be a random variable, and let $f_0, f_1, f_2, \ldots$ be a sequence of **increasing** functions from $\mathbf{R}$ into $\mathbf{R}$. Show that

(i) $f_n(X) \overset{\sim}{\to} f_0(X) \quad \Rightarrow \quad f_n(X) \to f_0(X)$ P-a.s.

Remark: This is the one of the rare cases where convergence in law implies convergence a.s. The increasingness of the f_n 's is absolutely indispensable. For instance, if $X \sim N(0,1)$ and $f_n(x) = (-1)^n x$, then $f_n(X) \overset{\sim}{\to} f_0(X)$, but $f_n(x) \to f_0(x)$ if and only if $x = 0$.

Hint: Let $\chi(p)$ be a p-fractile of X for all $0 < p < 1$ and show that $\chi_n(p) := f_n(\chi(p))$ is a p-fractile of $X_n := f_n(X)$ for all $0 < p < 1$ and all $n = 0, 1, 2, \ldots$. Let U be a random variable which is uniformly distributed on $]0,1[$, and let $D = \{s \in \mathbf{R} \mid f_n(s) \to f_0(s)\}$. Apply Exc.5.20 and (5.37.2) to show that $P(X \in D) = P(\chi(U) \in D) = 1$.

Exc. 5.22: Consider the setting of Exc.2.48 and let $F \in Dis(\mathbf{R})$ be a distribution function. Show that $\Phi_F(a,b)$ is continuous with respect convergence in law, that is, if $a_0, a_1, a_2, \ldots \in \mathbf{R}$ and $b_0, b_1, b_2, \ldots \in]0,\infty[$, then

(i) $(a_n, b_n) \to (a_0, b_0) \quad \Rightarrow \quad \Phi_F(a_n, b_n) \overset{\sim}{\to} \Phi_F(a_0, b_0)$

Moreover, show that the converse implication holds if F is non-degenerated

Hint: Let $Y_n = b_n X + a_n$, where X is a random variable with distribution F . Then $Y_n \sim \Phi_F(a_n, b_n)$. Show that (i) follows from (5.4.4). Note that if $b_n z + a_n \to b_0 z + a_0$ for two different z's, then $a_n \to a_0$ and $b_n \to b_0$. Then show that the converse implication (supposing that F is non-degenerated) follows from Exc.5.21 with $f_n(x) = b_n x + a_n$.

Exc. 5.23: Let $X, X_1, X_2, \ldots$ be k-dimensional random vectors with characteristic functions $\varphi, \varphi_1, \varphi_2, \ldots$. Suppose that $X_n \overset{\sim}{\to} X$ and let $K \subseteq \mathbf{R}^k$ be a bounded set. Show that $\varphi_n(t) \rightrightarrows \varphi(t)$ uniformly for $t \in K$.

Hint: Let $\varepsilon > 0$ be given and set $X_0 = X$ and $\varphi_0 = \varphi$. By (5.5.2) and (5.8.3) there exists $0 < c < \infty$ such that $P(||X_n|| > c) < \varepsilon$ for all $n \geq 0$. Let $A_n = \{ ||X_n|| > c\}$ and show that

$$\left| e^{i(t \bullet X_n)} - e^{i(s \bullet X_n)} \right| \leq c||t-s|| + 2 \cdot 1_{A_n} \quad , \quad |\varphi_n(t) - \varphi_n(s)| \leq c||t-s|| + 2\varepsilon$$

for all $s, t \in \mathbf{R}^k$ and all $n \geq 0$. Let $\delta = \frac{\varepsilon}{c}$ and let the open balls $b(a_1, \delta), \ldots, b(a_m, \delta)$ be chosen such that $K \subseteq b(a_1, \delta) \cup \cdots \cup b(a_m, \delta)$. Then choose an integer $n_0 \geq 1$ such that $|\varphi_n(a_j) - \varphi_0(a_j)| < \varepsilon$ for all $n \geq n_0$ and all $j = 1, \ldots, m$. Finally, show that $|\varphi_n(t) - \varphi_0(t)| < 5\varepsilon$ for all $n \geq n_0$ and all $t \in K$.

Exc. 5.24: Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with characteristic function φ and density function f where

$$f(x) = \begin{cases} \frac{1}{2} x^{-2} & \text{if } |x| \geq 1 \\ 0 & \text{if } |x| < 1 \end{cases}$$

Let $\zeta_2(t)$ denote the function:

$$\zeta_2(t) = \int_t^\infty \frac{1 - \cos x}{x^2} \, dx \qquad \forall t \geq 0$$

Show that

(i) $E|X_n|^q = \frac{1}{1-q} \quad \forall 0 < q < 1$ and $E|X_n| = \infty$

(ii) $\varphi(t) = \varphi(-t) = 1 - |t| \, \zeta_2(|t|) \qquad \forall t \in \mathbf{R}$

(iii) $\frac{1}{n} \sum_{j=1}^{n} X_j \overset{\sim}{\to} C(0, \frac{\pi}{2})$

where $C(a,b)$ denotes the Cauchy distribution.

Hint: Let φ_n be the characteristic function for $U_n := (X_1 + \cdots + X_n)/n$. Show that

$$\varphi_n(t) = \left(1 - \frac{|t|}{n}\,\zeta_2\left(\frac{|t|}{n}\right)\right)^n$$

Recall that $\zeta_2(0) = \frac{\pi}{2}$ and that the Cauchy distribution $C(a,b)$ has characteristic function $\exp(ita - b\,|t|)$.

Exc. 5.25: Let a_{nj} and b_{nj} for $1 \le j \le n < \infty$ be complex numbers satisfying

(a) $$\lim_{n\to\infty} \sum_{j=1}^{n} |1 - a_{nj}|^2 = 0 \quad , \quad \lim_{n\to\infty} \sum_{j=1}^{n} (a_{nj} - 1 - b_{nj}) = a$$

for some complex number $a \in \mathbf{C}$. Show that [see (5.16.6)]

(i) $$\lim_{n\to\infty} \prod_{j=1}^{n} \{\, a_{nj}\, e^{-b_{nj}} \,\} = e^a$$

Hint: Let $\operatorname{Log} z$ denote the main logarithm for $|1 - z| < 1$, that is,

$$\operatorname{Log} z = -\sum_{k=1}^{\infty} \frac{(1-z)^k}{k} \qquad \forall\, z \in \mathbf{C} \quad \text{with} \quad |1 - z| < 1$$

Then $\exp(\operatorname{Log} z) = z$ for all $|1 - z| < 1$. Show that

$$|\operatorname{Log} z - (z-1)| \le 2\,|1-z|^2 \qquad \forall\, z \in \mathbf{C} \quad \text{with} \quad |1 - z| \le \tfrac{1}{2}$$

and choose $m \ge 1$ so large that $|1 - a_{nj}| \le \frac{1}{2}$ for all $n \ge m$ and all $1 \le j \le n$. Let $n \ge m$ be given and show that

$$\prod_{j=1}^{n} a_{nj} e^{-b_{nj}} = \exp\left\{\sum_{j=1}^{n} (\operatorname{Log} a_{nj} - b_{nj})\right\} = \exp\left\{r_n + \sum_{j=1}^{n} (a_{nj} - 1 - b_{nj})\right\}$$

where

$$r_n = \sum_{j=1}^{n} (\operatorname{Log} a_{nj} - (a_{nj} - 1)) \quad , \quad |r_n| \le 2 \sum_{j=1}^{n} |1 - a_{nj}|^2$$

Show that $r_n \to 0$ and argue that this shows (i).

Exc. 5.26: Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables with common characteristic function $\varphi(t)$ and common distribution $\nu(B) = P(\xi_n \in B)$, and let N be a random variable taking values in $\mathbf{N}_0 = \{0, 1, 2, \ldots\}$ such that $N, \xi_1, \xi_2, \ldots$ are independent. Let $S_n = \xi_1 + \cdots + \xi_n$ for $n = 1, 2, \ldots$ and $S_0 = 0$. Let $q(t)$ be the generating function of N and set $S(\omega) = S_{N(\omega)}$ for all $\omega \in \Omega$,. Let $\varphi_S(t)$ be the characteristic function of S . Show that

(i) $S = \sum_{n=0}^{\infty} S_n \, 1_{\{N=n\}}$

(ii) $e^{itS} = \sum_{n=0}^{\infty} e^{itS_n} \, 1_{\{N=n\}} \qquad \forall t \in \mathbf{R}$

(iii) $\varphi_S(t) = q(\varphi(t)) \qquad \forall t \in \mathbf{R}$

(iv) $N \sim Po(\lambda) \;\Rightarrow\; \varphi_S(t) = e^{\lambda(\varphi(t)-1)} = \exp\left\{\lambda \int_{-\infty}^{\infty} \left(e^{itx} - 1\right) \nu(dx)\right\}$

Exc. 5.27: Let X be a non-degenerated random variable with characteristic function $\varphi(t)$. Let $a > 0$ and $q > 1$ be given and show

(i) $|\varphi(at)| \le |\varphi(t)| \quad \forall t \in \mathbf{R} \quad \Rightarrow \quad a \ge 1$

(ii) $|\varphi(at)| \le |\varphi(t)|^q \quad \forall t \in \mathbf{R} \quad \Rightarrow \quad a > 1$

Suppose that X is non-degenerated and stable in the sense of (5.29.1) and let $a_n > 0$ and $b_n \in \mathbf{R}$ be chosen so that (5.29.1)–(5.29.3) hold. Show that

(iii) X is infinitely divisible

(iv) $a_{n \cdot k} = a_n \cdot a_k \qquad \forall n, k \ge 1$

(v) $1 = a_1 < a_2 < \cdots < a_n < \cdots$

(vi) $b_1 = 0$ and $b_{n \cdot k} = a_k \cdot b_n + n \cdot b_k \qquad \forall n, k \ge 1$

(vii) $\exists 0 < p \le 2 \;:\; a_n = n^{1/p} \qquad \forall n \ge 1$

Hint: (i)+(ii): Apply Exc.5.44 to show that there exists a number $t_0 \in \mathbf{R}$ with $0 < |\varphi(t_0)| < 1$. Show that $|\varphi(a^n t_0)| \le |\varphi(t_0)|^{\gamma^n}$ for all $n \ge 1$ where $\gamma = 1$ in case (i) and $\gamma = q$ in case (ii). Then let $n \to \infty$. (iv)–(vi): Let $X_1, X_2, \ldots$ be a sequence of independent random variables having the same distribution as X . Let

$$Y_j^n = \sum_{(j-1)n \cdot < i \le j \cdot n} X_i \quad \forall n, j \ge 1$$

Show that for $n \geq 1$ fixed we have that $Y_1^n, Y_2^n, \ldots$ are independent and identically distributed with $Y_j^n \sim a_n X_j + b_n$ for all $n, j \geq 1$. Note that

$$a_{n \cdot k} X + b_{n \cdot k} \sim \sum_{i=1}^{n \cdot k} X_i = \sum_{j=1}^{k} Y_j^n$$

$$\sim k \cdot b_n + a_n \sum_{j=1}^{k} X_j$$

$$\sim k \cdot b_n + a_n \cdot b_k + a_n \cdot a_k X$$

Apply Exc.2.48.(vi) to prove (iv), (vi) and that $a_1 = 1$. Apply (ii) and (5.29.3) to prove (v). (vii): Let $p = \log 2 / \log a_2$. Show that $a_k = k^{1/p}$ whenever $k = 2^m$ for some $m = 0, 1, 2, \ldots$. Let $n \geq 3$ be given, and set $m_{nq} = int(\frac{q \log n}{\log 2})$ for $q = 1, 2, \ldots$, where $int(x)$ as usual denotes the integer part of x . Show that

$$2^{m_{nq}} \leq n^q < 2^{1+m_{nq}} \qquad \forall q \geq 1$$

$$\frac{m_{nq}}{p} \leq \frac{\log a_{n^q}}{\log 2} = \frac{q \log a_n}{\log 2} \leq \frac{1 + m_{nq}}{p} \qquad \forall q \geq 1$$

$$\frac{\log n}{\log 2} = \lim_{q \to \infty} \frac{m_{nq}}{q} = \frac{\log a_n}{\log a_2}$$

and argue that this shows that $a_n = n^{1/p}$. Note that $|\varphi(n^{-1/p})| = |\varphi(1)|^{1/n}$ and show that

$$p > 2 \quad \Rightarrow \quad \lim_{n \to \infty} \frac{1 - \left| \varphi(n^{-1/p}) \right|^2}{n^{-2/p}} = 0$$

Apply Exc.5.6.(ii) to conclude that $p \leq 2$.

Exc. 5.28: Let X, Y and Z be independent identically distributed random variables with finite and positive variance $\sigma^2 > 0$. Suppose that there exists a positive number $a > 0$ such that $X + Y \sim aZ$. Show that $a = \sqrt{2}$ and that the common distribution of X , Y and Z is the normal distribution $N(0, \sigma^2)$.

Hint: Apply (4.4.18) and the linearity of the mean to show that $EX = 0$ and $a = \sqrt{2}$. Let $\varphi(t)$ be the common characteristic function for X , Y and Z and show that

$$\varphi(t) = \varphi\left(\tfrac{t}{\sqrt{2}}\right)^2 = \cdots = \left\{ 1 + \frac{k\left(\varphi\left(\frac{t}{\sqrt{k}}\right) - 1\right)}{k} \right\}^k \qquad \forall k = 2, 4, 8, 16, \ldots$$

Apply l'Hosptial's rule to show that

$$\lim_{k \to \infty} k \left(\varphi\left(\tfrac{t}{\sqrt{k}}\right) - 1\right) = -\tfrac{1}{2} \sigma^2 t^2$$

and conclude that $X \sim N(0, \sigma^2)$

Exc. 5.29: Let $\{X_n\}$ be a sequence of independent, identically distributed random variables with distribution function $F(t) = P(X_n \le t)$. Let

$$F_n(\omega, t) = \tfrac{1}{n} \sum_{j=1}^{n} 1_{\{X_j \le t\}}(\omega) \qquad \forall n \ge 1 \, \forall \omega \in \Omega \, \forall t \in \mathbf{R}$$

be the empirical distribution functions. Show that for every fixed $t \in \mathbf{R}$ we have $\sqrt{n}\,\big(F_n(t) - F(t)\big) \xrightarrow{\sim} N(0, \xi(t))$, where $\xi(t) = F(t)(1 - F(t))$ [apply (5.16.7) with $X_j := 1_{\{X_j \le t\}}$].

Exc. 5.30: Let $\{X_n\}$ be a sequence of independent, identically distributed random variables with a geometric distribution $Ge(p)$ where $0 < p < 1$. Let $k \ge 1$ be a given integer and set

$$U_n = X_n \wedge k \quad , \quad V_n = 1_{\{X_n < k\}} \quad , \quad \mathsf{U}_n = \sum_{j=1}^{n} U_j$$

$$\mathsf{V}_n = \sum_{j=1}^{n} V_j \quad , \quad \mathsf{W}_n = \frac{\mathsf{U}_n}{\mathsf{V}_n + \mathsf{U}_n} \quad , \quad Z_j = (1-p)U_j - pV_j$$

[see Exc.4.61]. Show that $E\,Z_j = 0$ and

(i) $\quad \frac{1}{n}\left(\mathsf{V}_n + \mathsf{U}_n\right) \to \frac{1-p^k}{1-p}$ a.s.

(ii) $\quad \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j \xrightarrow{\sim} N\big(0, \sigma^2\big)$ where $\sigma^2 = \mathrm{var}(Z_1)$

(iii) $\quad \sqrt{n}\,(\mathsf{W}_n - p) \xrightarrow{\sim} N\big(0, \tau^2\big)$ where $\tau = \frac{1-p}{1-p^k}\,\sigma$

Exc. 5.31: Let $\{X_n\} \subseteq L^3(P)$ be a sequence of independent random variables with means $\mu_n = E\,X_n$ and variances $\sigma_n^2 = \mathrm{var}(X_n)$. Let $s_n = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$ and let $0 < a_1 \le \cdots \le a_n \cdots$ be an increasing sequence of positive numbers such that $E|X_j|^3 \le a_j\,\sigma_j^2$ for all $j \ge 1$ and $\lim_{n\to\infty} a_n\,s_n^{-1} = 0$. Show that $E|X_j - \mu_j|^3 \le 8 \cdot a_j\,\sigma_j^2$ for all $j \ge 1$ [apply Exc.4.57]. Then show that

(i) $$\frac{1}{s_n} \sum_{j=1}^{n} (X_j - \mu_j) \xrightarrow{\sim} N(0,1)$$

Exc. 5.32: Let $\{X_n\}$ be a sequence of independent random variables such that X_n only assumes the values $\sqrt{n}$, $-\sqrt{n}$ and 0 with the following probabilities:

$$P(X_n = \sqrt{n}) = P(X_n = -\sqrt{n}) = \tfrac{1}{2}\, n^{-\beta} \quad , \quad P(X_n = 0) = 1 - n^{-\beta}$$

for all $n \geq 1$ where $0 < \beta < 1$ is a given number [see Exc.4.64]. Show that

(i) $$\frac{1}{s_n} \sum_{j=1}^{n} X_j \overset{\sim}{\to} N(0,1) \quad \text{where} \quad s_n = \sqrt{1^{1-\beta} + 2^{1-\beta} + \cdots + n^{1-\beta}}$$

Hint: Apply the following inequalities:

(a) $$\frac{n^{p+1}}{p+1} \leq \sum_{j=1}^{n} j^p \leq n^{p+1} \qquad \forall p > 0 \, \forall \, n \geq 1$$

to verify Lyapounov's condition.

Exc. 5.33: Let $\{X_n\} \subseteq L^4(P)$ be a sequence of independent random variables with means $\mu_n = E\,X_n$ and variances $\text{var}(X_n) = 1$ for all $n \geq 1$. Let $a_1, a_2, \ldots$ be real numbers, and set

$$T_n = \frac{1}{s_n} \sum_{j=1}^{n} a_j\,(X_j - \mu_j) \quad \text{where} \quad s_n = \sqrt{a_1^2 + a_2^2 + \cdots a_n^2}$$

Suppose that there exists a constant $0 < C < \infty$ such that $E\,X_n^4 \leq C$ for all $n \geq 1$. Show that $E|X_j - \mu_j|^4 \leq 16 \cdot a_j\,\sigma_j^2$ for all $j \geq 1$ [apply Exc.4.57]. Then show that

(i) $$\lim_{n\to\infty} \left\{ s_n^{-4} \sum_{j=1}^{n} a_j^4 \right\} = 0 \qquad \Rightarrow \qquad T_n \overset{\sim}{\to} N(0,1)$$

Exc. 5.34: Let $\{X_n\} \subseteq L^4(P)$ be a sequence of independent random variables with $E\,X_n = 0$, $\text{var}(X_n) = \sigma^2 > 0$ and $E|X_n|^4 \leq c\sqrt{n}$ for all $n \geq 1$ [see Exc.4.66] where $0 < c < \infty$ is a given constant. Let

$$S_n = \sum_{j=1}^{n} X_j \quad , \quad V_n = \sum_{j=1}^{n} X_j^2$$

Show that

(i) $$\frac{S_n}{\sqrt{n}} \overset{\sim}{\to} N(0,\sigma^2) \quad \text{and} \quad \frac{S_n}{\sqrt{V_n}} \overset{\sim}{\to} N(0,1)$$

Exc. 5.35: Let $\{X_n\}$ be a sequence of independent random variables such that X_n is a uniformly distributed on the interval $[-n^\beta, n^\beta]$ for all $n \geq 1$ where $\beta > 0$ is a given positive number [see Exc.4.67]. Let

$$s_n = \sqrt{\tfrac{1}{3}\left(1^{2\beta} + 2^{2\beta} + \cdots + n^{2\beta}\right)} \qquad \forall n = 1, 2, \ldots$$

and show that [apply Exc.5.32.(a) to verify Lyapounov's condition]

(i) $$\frac{1}{s_n} \sum_{j=1}^{n} X_j \overset{\sim}{\to} N(0,1)$$

Exc. 5.36: Let $\{X_n\}$ be independent, identically distributed random variables with density function given by

$$f(x) = \begin{cases} \frac{a}{2}\,|x|^{a-1} & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1 \end{cases}$$

where $a > 0$ is a given positive number. Let $b_n > 0$ be positive numbers such that

$$\sum_{j=1}^{\infty} b_j^2 = \infty \quad , \quad \sum_{j=1}^{\infty} b_j^q < \infty$$

for some $q > 2$ [see Exc.4.68]. Show that

(i) $$\frac{1}{s_n} \sum_{j=1}^{n} b_j X_j \overset{\sim}{\to} N\left(0, \tfrac{a}{a+2}\right) \quad \text{where} \quad s_n = \sqrt{b_1^2 + b_2^2 + \cdots + b_n^2}$$

Exc. 5.37: Let $X_1, X_2, \ldots$ be independent identically distributed random variables with common density function $f(x)$ satisfying

$$\int_{-\infty}^{\infty} x^2 f(x)\, dx = \infty \quad , \quad f(x) = f(-x) \leq C\,|x|^{-3} \qquad \forall x \neq 0$$

where $0 < C < \infty$ is a given constant. Let

(i) $$\beta_n = \int_{-\sqrt{n}}^{\sqrt{n}} x^2 f(x)\, dx \quad , \quad X_{nj} = \begin{cases} X_j & \text{if } |X_j| \leq \sqrt{n} \\ 0 & \text{if } |X_j| > \sqrt{n} \end{cases}$$

(ii) $$U_n = \tfrac{1}{n} \sum_{j=1}^{n} X_j \quad , \quad V_n = \tfrac{1}{\sqrt{n\beta_n}} \sum_{j=1}^{n} X_{nj}$$

Show that

(i) $EX_n = 0$ and $EX_n^2 = \infty$ for all $n \geq 1$

(ii) $EX_{nj} = 0$, $EX_{nj}^2 = \beta_n$ and $E|X_{nj}|^3 \leq 2C\sqrt{n}$ for all $n, j \geq 1$

(iii) $U_n \to 0$ a.s. , $V_n \overset{\sim}{\to} N(0,1)$ and $\dfrac{V_n}{1+U_n} \overset{\sim}{\to} N(0,1)$

Hint: Show that $\{X_{nj} \mid 1 \leq j \leq n\}$ is an independent, triangular scheme and that $\lim\limits_{n\to\infty} \beta_n = \infty$.

Exc. 5.38: Let $\varepsilon_1, \varepsilon_2, \ldots$ be independent identically distributed random variables with $P(\varepsilon_n = 1) = P(\varepsilon_n = -1) = \frac{1}{2}$ for all $n \geq 1$. Show that [apply Exc.5.32.(a) to verify Lyapounov's condition]

(i) $0 < q < \frac{1}{2} \quad \Rightarrow \quad \frac{1}{n} \sum\limits_{j=1}^{n} j^q \varepsilon_j \to 0$ a.s.

(ii) $q > \frac{1}{2} \quad \Rightarrow \quad \sum\limits_{j=1}^{\infty} j^{-q} \varepsilon_j$ is summable a.s.

(iii) $q > 0 \quad \Rightarrow \quad s_n^{-1} \sum\limits_{j=1}^{n} j^q \varepsilon_j \overset{\sim}{\to} N(0,1)$ where $s_n = \sqrt{1^{2q} + \cdots + n^{2q}}$

Exc. 5.39: Let $X_1, X_2, \ldots$ be independent identically random variables with common density function

$$f(x) = \begin{cases} \frac{1}{2}\,|x|^{-2} & \text{if } |x| \geq 1 \\ 0 & \text{if } |x| < 1 \end{cases}$$

Let $\{c_n\}$ be a sequence of positive numbers such that $c_n \geq 2$ for all $n \geq 1$ and $\lim\limits_{n\to\infty} \frac{c_n}{n} = 0$. Set

$$X_{nj} = \begin{cases} X_j & \text{if } |X_j| \leq c_n \\ 0 & \text{if } |X_j| > c_n \end{cases}$$

and show that

(i) $E|X_n| = \infty$, $EX_{nj} = 0$, $EX_{nj}^2 = c_n - 1$, $E|X_{nj}|^3 = \frac{1}{2}\left(c_n^2 - 1\right)$

(ii) $\dfrac{1}{\sqrt{n\,(c_n - 1)}} \sum\limits_{j=1}^{n} X_{nj} \overset{\sim}{\to} N(0,1)$

Exc. 5.40: Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with common density function

$$f(x) = \frac{g(|x|)}{x^2} \qquad \forall x \neq 0$$

where $g : [0, \infty[\to [0, \infty[$ is a **bounded** Borel function satisfying

$$\int_0^\infty \frac{g(x)}{x^2}\, dx = \frac{1}{2} \quad \text{and} \quad \lim_{x \to \infty} g(x) = 0$$

Show that $U_n = (X_1 + \cdots + X_n)/n \to 0$ in *pr.*

Hint: Let $\varphi(t)$ be the common characteristic function for $X_1, X_2, \ldots$, and let $\varphi_n(t)$ be the characteristic function for U_n. Set

$$\psi_n(t) = 2 \int_0^\infty \frac{g(nx)\,(1 - \cos(tx))}{x^2}\, dx \qquad \forall t \in \mathbf{R}\ \forall n = 1, 2, \ldots$$

Apply Lebesgue's dominated convergence theorem to show that $\psi_n(t) \to 0$ as $n \to \infty$ for all $t \in \mathbf{R}$ and note that

$$\varphi_n(t) = \varphi\left(\tfrac{t}{n}\right)^n = \left(1 - \frac{\psi_n(t)}{n}\right)^n \to 1 \quad \text{as } n \to \infty$$

Exc. 5.41: Let $X_1, X_2, \ldots$ be independent identically distributed random variables with common density function

$$f(x) = \begin{cases} \frac{q}{2}\,|x|^{-q-1} & \text{if} \quad |x| \geq 1 \\ 0 & \text{if} \quad |x| < 1 \end{cases}$$

where $0 < q < 2$. Let φ be the common characteristic function for $X_1, X_2, \ldots$ and let φ_n be the characteristic function for the random variable $U_n = n^{-1/q}(X_1 + \cdots + X_n)$. Show that

(i) $$E|X_n|^\alpha = \begin{cases} \frac{q}{q-\alpha} & \text{if} \quad 0 < \alpha < q \\ \infty & \text{if} \quad \alpha \geq q \end{cases}$$

(ii) $$\varphi(t) = q \int_1^\infty \frac{\cos(tx)}{x^{q+1}}\, dx = 1 - q|t|^q \int_{|t|}^\infty \frac{1 - \cos x}{x^{q+1}}\, dx$$

(iii) $$\varphi_n(t) = \left(1 - \frac{q\,|t|^q}{n} \int_{|t|\,n^{-1/q}}^\infty \frac{1 - \cos x}{x^{q+1}}\, dx\right)^n$$

(iv) $$U_n \xrightarrow{\sim} stab(0, \sigma, q, 0) \quad \text{where} \quad \sigma = \left\{ q \int_0^\infty \frac{1 - \cos x}{x^{q+1}}\, dx \right\}^{1/q}$$

Exc. 5.42: Let $X_1, X_2, \ldots$ be independent random variables which only takes the values 0 and 1 with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$ for all $n \geq 1$, where $0 < p_n < 1$ for all $n \geq 1$ [see Exc.4.78]. Show that

(i) $$\sum_{j=1}^{\infty} p_j(1-p_j) = \infty \quad \Rightarrow \quad s_n^{-1} \sum_{j=1}^{n} (X_j - p_j) \xrightarrow{\sim} N(0,1)$$

where [see Exc.4.78]

$$s_n = \sqrt{\sum_{j=1}^{n} p_j(1-p_j)} \quad \forall n \geq 1$$

Exc. 5.43: Let $X_1, X_2, \ldots$ be independent identically distributed random variables with common density function f satisfying $f(x) = f(-x)$ for all $x \in \mathbf{R}$. Let

$$\alpha_n = 2\int_0^{r_n} x^2 f(x)\,dx \quad , \quad \beta_n = 2\int_0^{r_n} x^3 f(x)\,dx$$

$$X_{nj} = \begin{cases} X_j & \text{if } |X_j| \leq r_n \\ 0 & \text{if } |X_j| \leq r_n \end{cases} \quad , \quad q(x) = \begin{cases} x & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| \leq 1 \end{cases}$$

for all $n, j \geq 1$, where $r_1, r_2, \ldots$ is a given sequence of positive numbers such that $\alpha_n > 0$ for all $n \geq 1$. Show that

(i) $$EX_{nj} = 0 \quad , \quad EX_{nj}^2 = \alpha_n \quad , \quad E|X_{nj}|^3 = \beta_n \quad \forall j, n \geq 1$$

(ii) $$Eq\left(\frac{X_j}{r_n}\right) = 0 \quad , \quad Eq\left(\frac{X_j}{r_n}\right)^2 = \frac{\alpha_n}{r_n^2} \quad \forall j, n \geq 1$$

(iii) $$\lim_{n\to\infty} \frac{\beta_n^2}{n\,\alpha_n^3} = 0 \quad \Rightarrow \quad \frac{1}{\sqrt{n\,\alpha_n}} \sum_{j=1}^{n} X_{nj} \xrightarrow{\sim} N(0,1)$$

Suppose that the following condition holds:

(i) $$\lim_{n\to\infty} \frac{n\,\alpha_n}{r_n^2} = 1 \quad \text{and} \quad \lim_{n\to\infty} \int_{\varepsilon r_n}^{\infty} \left(1 + q\left(\tfrac{x}{r_n}\right)^2\right) f(x)\,dx = 0 \quad \forall \varepsilon > 0$$

Show that

(iv) $$\frac{1}{r_n} \sum_{j=1}^{n} X_j \xrightarrow{\sim} N(0,1)$$

Exc. 5.44: Let X a random variable and suppose that there exists nonzero numbers a, b such that $|\varphi_X(a)| = |\varphi_X(b)| = 1$ and $\frac{a}{b}$ is irrational. Show that X is degenerated.

Hint: Note that by irrationality of $\frac{a}{b}$, we have that the two sets

$$A = \left\{ \theta + \tfrac{2k\pi}{a} \;\middle|\; k \in \mathbf{Z} \right\} \quad , \quad B = \left\{ \vartheta + \tfrac{2k\pi}{b} \;\middle|\; k \in \mathbf{Z} \right\}$$

have at most one point in common. Then apply the equivalence of (a) and (c) in Exc.5.9.

Chapter 6

CONDITIONAL EXPECTATIONS

6.1 Information and σ-Algebras

We have chosen Kolmogorov's axiomatic system $(\Omega, \mathcal{F}, P)$ [see (2.1.1) – (2.1.3)] as our model of probabilities. In the Kolmogorov model, the set Ω is interpreted as the set of all possible outcomes, the σ-algebra $\mathcal{F}$ is interpreted as the set of all "observable" events, and the probability $P(F)$ is interpreted as the degree of conviction we have that the event F occurs. In many situations we have a certain information about the outcomes, which allow us to decide if some (but usually not all) events have occurred or not. For instance, suppose that we meet our good friend Agatha at a horse race between the fourth and fifth run, and consider the following two events:

$$A = \{ \text{ Lady Hannover wins the fourth run} \}$$

$$B = \{\text{Cognac Olsen wins the fifth run}\}$$

Agatha can tell us if A has occurred or not, but she cannot yet tell us if B will occur or not. Suppose that 10 horses are starting in the fifth run, a totally ignorant person would then assign the probability $\frac{1}{10}$ to the event B . But Agatha is not ignorant; she tells us that Cognac Olsen is a fast running stallion in a superb condition, but he has a bad record of being disqualified for an irregular trot. Comparing this with the information that Agatha tells about the other nine horses, we may decide to reevaluate our naive probability $\frac{1}{10}$ and assign the probability $\frac{1}{3}$ to the event B .

In this chapter, I shall formalize the notion of information and probabilities or expectations based on information. Well, what is information? Information is the

ability of deciding if G has occurred or not, for all G in a certain subset of $\mathcal{F}$. Thus, information may be described as a paving $\mathcal{G} \subseteq \mathcal{F}$. If we can decide if G has occurred or not, then clearly we can decide if the complement G^c has occurred or not. Similarly, if we can decide if G_n has occurred or not for all $n = 1, 2, \cdots$, then it is natural to claim that we also can decide if one of the G_n's (that is, $G = \cup_1^\infty G_n$) has occurred or not. Thus, it is natural to describe information as a $(\cup c, C)$-stable subpaving of $\mathcal{F}$, that is, as a sub-σ-algebra of $\mathcal{F}$

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then, inspired by the remarks above, we define *information* to be a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$. Then $\mathcal{G}$ is interpreted as the set of all events which are known to us. Suppose that X is a random function with values in the measurable space $(M, \mathcal{B})$. If the events $\{X \in B\}$ belongs to $\mathcal{G}$ for all $B \in \mathcal{B}$ (that is, if X is $\mathcal{G}$-measurable), then the information $\mathcal{G}$ tells exactly what X is. We say that X is *$\mathcal{G}$-observable* whenever X is $\mathcal{G}$-measurable.

Information can also be described as a net on Ω such that two outcomes in the same mesh cannot be distinguished by the information available, but outcomes in two different meshes can. You may think of such nets as a map. On a world map, it is not possible to distinguish the Empire State Building and the United Nation Building in New York City, but a city map of New York contains enough information to distinguish the two sites. In the next section, we shall see how this aspect is built into our definition of information.

6.2 Atoms and Information

Let $\mathcal{G}$ be a σ-algebra on Ω. Then we define *the equivalence relation induced by* $\mathcal{G}$ in the following way:

$$\omega' \equiv \omega (\text{mod } \mathcal{G}) \quad \Longleftrightarrow \quad 1_G(\omega') = 1_G(\omega) \qquad \forall G \in \mathcal{G} \tag{6.2.1}$$

The equivalence classes [see (P.7)] of this equivalence relation are called *the $\mathcal{G}$-atoms*. The equivalence class containing ω_0 is called *the $\mathcal{G}$-atom containing* ω_0 and is denoted

$$\mathcal{G}(\omega_0) = \{\omega \in \Omega \mid \omega \equiv \omega_0 (\text{mod } \mathcal{G})\} \tag{6.2.2}$$

By (P.7.5)–(P.7.7), we know that the $\mathcal{G}$-atoms form a disjoint partition of Ω. Hence, $\mathcal{G}$ induces a net on Ω with the $\mathcal{G}$-atoms as meshes.

If $\omega_0 \not\equiv \omega_1 \, (\text{mod } \mathcal{G})$, then there exists $G \in \mathcal{G}$ so that $\omega_0 \in G$ and $\omega_1 \notin G$. Hence, we can distinguish between ω_0 and ω_1 on the basis of the information $\mathcal{G}$. But if $\omega_0 \equiv \omega_1 \, (\text{mod } G)$ and $G \in \mathcal{G}$, then $\omega_0 \in G$ if and only if $\omega_1 \in G$. So based on the information $\mathcal{G}$, we cannot distinguish between ω_0 and ω_1. This shows that the atomic $\mathcal{G}$-net has the property we were looking for:

Two points in the same atom are indistinguishable, but points in different atoms are distinguishable.

If $\mathcal{G}(\omega) = \{\omega\}$ for all $\omega \in \Omega$; that is, if $1_G(\omega_0) = 1_G(\omega_1)$ for all $G \in \mathcal{G}$ implies $\omega_0 = \omega_1$, then we say that $\mathcal{G}$ *separates points in* Ω . In this case, $\mathcal{G}$ gives complete information about the outcomes ω , in the sense that we can distinguish any two different outcomes $\omega_0 \neq \omega_1$ based on the information $\mathcal{G}$.

6.3 The Calculus of Atoms

Let $\mathcal{H}$ be a paving on Ω , let $\mathcal{G} = \sigma(\mathcal{H})$ be the σ-algebra generated by $\mathcal{H}$, and let $\omega_0, \omega_1 \in \Omega$ be given points in Ω . Then we have

(6.3.1) $$\omega_0 \equiv \omega_1 (\text{mod } \mathcal{G}) \iff 1_H(\omega_0) = 1_H(\omega_1) \quad \forall H \in \mathcal{H}$$

(6.3.2) $$A \subseteq \bigcup_{\omega \in A} \mathcal{G}(\omega) \quad \forall A \subseteq \Omega \quad , \quad G = \bigcup_{\omega \in G} \mathcal{G}(\omega) \quad \forall G \in \mathcal{G}$$

Let $\{\mathcal{G}_j \mid j \in J\}$ be a family of σ-algebras and let $\mathcal{G}^*$ be the smallest σ-algebra containing all the $\mathcal{G}_j$'s. Then we have

(6.3.3) $$\omega_0 \equiv \omega_1 (\text{mod } \mathcal{G}^*) \iff \omega_0 \equiv \omega_1 (\text{mod } \mathcal{G}_j) \quad \forall j \in J$$

(6.3.4) $$\mathcal{G}^*(\omega_0) = \bigcap_{j \in J} \mathcal{G}_j(\omega_0)$$

Let $\{\mathcal{F}_n \mid n \in \mathbf{N}\}$ be a decreasing sequence of σ-algebras and let $\mathcal{F}_*$ be their intersection. Then we have

(6.3.5) $$\omega_0 \equiv \omega_1 (\text{mod } \mathcal{F}_*) \iff \exists n \geq 1 \text{ so that } \omega_0 \equiv \omega_1 (\text{mod } \mathcal{F}_n)$$

(6.3.6) $$\mathcal{F}_*(\omega_0) = \bigcup_{n=1}^{\infty} \mathcal{F}_n(\omega_0)$$

Let X and X_j be a measurable functions from $(\Omega, \mathcal{G})$ into $(M, \mathcal{B})$ respectively $(M_j, \mathcal{B}_j)$ for all $j \in J$, and let $\mathcal{X} = \sigma(X_j \mid j \in J)$. Then we have the following:

(6.3.7) $$\omega_0 \equiv \omega_1 (\text{mod } \mathcal{G}) \quad \Rightarrow \quad X(\omega_0) \equiv X(\omega_1) (\text{mod } \mathcal{B})$$

(6.3.8) *If $\mathcal{B}$ separates points in M , then X is constant on every $\mathcal{G}$-atom*

(6.3.9) $$\omega_0 \equiv \omega_1 (\text{mod } \mathcal{X}) \iff X_j(\omega_0) \equiv X_j(\omega_1) (\text{mod } \mathcal{B}_j) \quad \forall j \in J$$

(6.3.10) $$\mathcal{X}(\omega_0) = \bigcap_{j \in J} X_j^{-1}(\mathcal{B}_j(X_j(\omega_0)))$$

In particular, if $\mathcal{B}_j$ *separates points in* M_j *for all* $j \in J$ *, then we have*

(6.3.11) $\omega_0 \equiv \omega_1 (\text{mod } \mathcal{X}) \iff X_j(\omega_0) = X_j(\omega_1) \quad \forall j \in J$

(6.3.12) $\mathcal{X}(\omega_0) = \{\omega \in \Omega \mid X_j(\omega) = X_j(\omega_0) \quad \forall j \in J\}$

Remark: Note that (6.3.2) states that every set in $\mathcal{G}$ is a (possible uncountable) union of $\mathcal{G}$-atoms. In Section (10.5), we shall see that if the σ-algebra $\mathcal{G}$ and the set G are nice enough, then the converse holds.

Proof: (6.3.1): Let $\mathcal{L} = \{L \subseteq \Omega \mid 1_L(\omega_0) = 1_L(\omega_1)\}$. Then it is easily verified that $\mathcal{L}$ is a σ-algebra. But then (6.3.1) follows from the usual "smallest σ-algebra" argument.

(6.3.2): Note that $\omega \in \mathcal{G}(\omega)$ for all $\omega \in \Omega$. Hence the first inclusion follows. If $\omega \in G \in \mathcal{G}$ and $\omega^* \in \mathcal{G}(\omega)$, then $1_G(\omega^*) = 1_G(\omega) = 1$ and so $\omega^* \in G$. Thus, $\mathcal{G}(\omega) \subseteq G$ for all $\omega \in G$ and so the second equality follows.

(6.3.3+4): Follows from (6.3.1) with $\mathcal{H} = \cup_{j \in J} \mathcal{G}_j$.

(6.3.5+6): If $\omega_0 \equiv \omega_1 (\text{mod } \mathcal{F}_n)$ for some n , then $\omega_0 \equiv \omega_1 (\text{mod } \mathcal{F}_*)$ because $\mathcal{F}_* \subseteq \mathcal{F}_n$. If $\omega_0 \not\equiv \omega_1 (\text{mod } \mathcal{F}_n)$ for all $n \geq 1$, then there exists $F_n \in \mathcal{F}_n$ such that $\omega_0 \in F_n$ and $\omega_1 \notin F_n$ for all $n \geq 1$. Since $F_j \in \mathcal{F}_j \subseteq \mathcal{F}_n$ for all $j \geq n$, we have $F = \limsup F_n \in \mathcal{F}_*$ and since $\omega_0 \in F$ and $\omega_1 \notin F$, we conclude that $\omega_0 \not\equiv \omega_1 \, (\text{mod } \mathcal{F}_*)$. Thus, we have proved (6.3.5) and (6.3.6) follows easily from (6.3.5).

(6.3.7+8): Suppose that $\omega_0 \equiv \omega_1 \, (\text{mod } \mathcal{G})$ and set $v_0 = X(\omega_0)$ and $v_1 = X(\omega_1)$. If $B \in \mathcal{B}$, then $G = X^{-1}(B)$ belongs to $\mathcal{G}$ and so $1_B(v_0) = 1_G(\omega_0) = 1_G(\omega_1) = 1_B(v_1)$. Hence, we have proved (6.3.7), and (6.3.8) follows from (6.3.7).

(6.3.9)–(6.3.12): (6.3.9) is proved in exactly the manner as (6.3.3) and (6.3.10)–(6.3.12) follow easily from (6.3.9). □

6.4 The Factorization Theorem

Let X *and* Y *be two functions from the set* Ω *into the measurable spaces* $(M, \mathcal{B})$ *respectively* $(L, \mathcal{A})$ *. If* φ *is a measurable function from* $(L, \mathcal{A})$ *into* $(M, \mathcal{B})$ *such that* $X(\omega) = \varphi(Y(\omega))$ *for all* $\omega \in \Omega$ *, then* X *is* $\sigma(Y)$*-measurable.*

Conversely, if $X : \Omega \to (M, \mathcal{B})$ *and* $Y : \Omega \to (L, \mathcal{A})$ *are two functions such that* X *is* $\sigma(Y)$*-measurable, then there exists a measurable function* $\varphi : (L, \mathcal{A}) \to (M, \mathcal{B})$ *such that* $X(\omega) = \varphi(Y(\omega))$ *for all* $\omega \in \Omega$ *, provided that just one of the following two conditions holds:*

(6.4.1) M *is a complete separable metric space and* $\mathcal{B} = \mathcal{B}(M)$

(6.4.2) *$\mathcal{B}$ separates points in M and $Y(\Omega) \in \mathcal{A}$*

Remark: Note that $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ satisfies (6.4.1). Hence, every $\sigma(Y)$-measurable real-valued random variable is a measurable function of Y.

Proof: The first part of the theorem is a trivial consequence of (1.38.4).

Suppose that X is $\sigma(Y)$-measurable and that (6.4.1) holds. Let $\{a_j \mid j \geq 1\}$ be a countable dense subset of M and let us define $B_{jn} = b(a_j, 2^{-n})$. Then $X^{-1}(B_{jn}) \in \sigma(Y)$, so there exists $A_{jn} \in \mathcal{A}$ such that $X^{-1}(B_{jn}) = Y^{-1}(A_{jn})$. Let C_{jn} be defined as follows:

$$C_{0n} = L \setminus \bigcup_{j=1}^{\infty} A_{jn} \quad , \quad C_{1n} = A_{1n} \quad , \quad C_{jn} = A_{jn} \setminus \bigcup_{i=1}^{j-1} A_{in} \qquad \forall j \geq 2$$

Then $C_{0n}, C_{1n}, C_{2n}, \ldots$ are disjoint and $L = C_{0n} \cup C_{1n} \cup \cdots$ for all $n \geq 1$ Setting $a_0 = a_1$ we may define $\varphi_n : L \to M$ as follows:

$$\varphi_n(x) = a_j \quad \text{if} \quad x \in C_{jn} \quad \text{and} \quad j \geq 0$$

Note that $C_{jn} \in \mathcal{A}$. Hence, φ_n is measurable and if we then define $\varphi : L \to M$ by

$$\varphi(x) = \begin{cases} \lim\limits_{n\to\infty} \varphi_n(x) & \text{if the limit exists} \\ a_0 & \text{otherwise} \end{cases}$$

then by Exc.1.74 we have that φ is measurable from $(L, \mathcal{A})$ into $(M, \mathcal{B})$. Let $\omega \in \Omega$ and $n \geq 1$ be given, and set $u = X(\omega)$ and $v = Y(\omega)$. Since $\{a_j\}$ is dense in M, we have that $M = \cup_{j=1}^{\infty} B_{jn}$; so by (1.2.13) and (P.3.1) we get

$$\Omega = \bigcup_{j=1}^{\infty} X^{-1}(B_{jn}) = \bigcup_{j=1}^{\infty} Y^{-1}(A_{jn}) = \bigcup_{j=1}^{\infty} Y^{-1}(C_{jn})$$

Hence, for some $j \geq 1$ we have that $\omega \in Y^{-1}(C_{jn}) \subseteq X^{-1}(B_{jn})$, and so $v = Y(\omega) \in C_{jn}$ and $u = X(\omega) \in B_{jn}$. Hence, $\varphi_n(v) = a_j$ and since $u \in B_{jn}$, we have that $d(\varphi_n(v), u) = d(a_j, u) \leq 2^{-n}$. Thus, we see that $\varphi_n(v) \to u$, and so we conclude that $X(\omega) = u = \varphi(v) = \varphi(Y(\omega))$ for all $\omega \in \Omega$.

Suppose that X is $\sigma(Y)$-measurable, and that (6.4.2) holds. Let $L_0 = Y(\Omega)$. Then $L_0 \in \mathcal{A}$ by (6.4.2). Choose a point $a_0 \in M$ and define

$$\varphi(x) = \begin{cases} X(\omega) & \text{if } x = Y(\omega) \text{ for some } \omega \in \Omega \\ a_0 & \text{if } x \in L \setminus L_0 \end{cases}$$

If $Y(\omega_0) = Y(\omega_1)$, then we have that $\omega_0 \equiv \omega_1 (\text{mod } \sigma(Y))$ by (6.3.11), and so $X(\omega_0) = X(\omega_1)$ by (6.3.8) because X is $\sigma(Y)$-measurable and $\mathcal{B}$ separates points in M . Thus, φ is a well-defined function from L into M . Let $B \in \mathcal{B}$. Then $X^{-1}(B) \in \sigma(Y)$ and so there exists $A \in \mathcal{A}$, with $X^{-1}(B) = Y^{-1}(A)$. Hence, by definition of φ we have

$$\varphi^{-1}(B) = \begin{cases} A \cap L_0 & \text{if } a_0 \notin B \\ A \cup (L \setminus L_0) & \text{if } a_0 \in B \end{cases}$$

and since A , L_0 and $L \setminus L_0$ all belong to $\mathcal{A}$, we see that φ is measurable from $(L, \mathcal{A})$ into $(M, \mathcal{B})$. Moreover, by definition of φ we have that $X(\omega) = \varphi(Y(\omega))$ for all $\omega \in \Omega$. □

6.5 Separable and Blackwell σ-Algebras

In (6.3.2), we saw that every set in $\mathcal{G}$ is a union of $\mathcal{G}$-atoms and it is natural to ask if or rather when the converse holds. Similarly, (6.3.7) naturally raises the question when (6.3.7) implies $\mathcal{G}$-measurability of X . To give an answer to this question we need some new concepts concerning σ-algebras.

Let (S, d) be a metric space and let $A \subseteq S$ be a subset of S . Then we say that A is *analytic* if there exists a separable, complete, metric space (T, δ) and a continuous function $f : T \to S$ such that $A = f(T)$. Let $(\Omega, \mathcal{G})$ be measurable space. Then we say that $\mathcal{G}$ is a *Blackwell σ-algebra* if $f(\Omega)$ is analytic for every measurable function f from $(\Omega, \mathcal{G})$ into $\mathbf{R}$, and we say that $\mathcal{G}$ is *separable* if there exist a countable paving $\mathcal{H}$ such that $\mathcal{G} = \sigma(\mathcal{H})$. By Exc.1.82, we have that $\mathcal{G}$ is separable if and only if $\mathcal{G} = \sigma(f)$ for some function $f : \Omega \to \mathbf{R}$. It can be shown that we have the following:

(6.5.1) If (S, d) is analytic, then $\mathcal{B}(S)$ is a separable Blackwell σ-algebra, and every Borel subset of S is analytic

(6.5.2) If $\mathcal{G}$ is a Blackwell σ-algebra, then so is every sub-σ-algebra of $\mathcal{G}$

(6.5.3) The product σ-algebra of Blackwell σ-algebras is a Blackwell σ-algebra

(6.5.4) If $\mathcal{G}$ is a Blackwell σ-algebra on Ω and f is a measurable surjection of $(\Omega, \mathcal{G})$ onto $(M, \mathcal{B})$, then $\mathcal{B}$ is a Blackwell σ-algebra on M

(6.5.5) The product σ-algebra $\mathcal{B}^T$ on $\mathbf{R}^T$ is a Blackwell σ-algebra for any set T

(6.5.6) If (S, d) is a separable metric space, then $\mathcal{B}(S)$ is a separable σ-algebra

(6.5.7) $\mathcal{B}(\mathbf{R}^k)$ is a separable Blackwell σ-algebra for all $k \in \mathbf{N}$

(6.5.8) The product σ-algebra $\mathcal{B}^T$ on $\mathbf{R}^T$ is separable whenever T is finite or countable

(6.5.9) If $f_j : \Omega \to \mathbf{R}$ is a function for all $j \in J$, and if J is finite or countable, then $\sigma(f_j \mid j \in J)$ is a separable σ-algebra

[See (P.5.15)-(P.5.19), Exc.1.13, Exc.1.14 and Exc.1.93 – 1.102.] This means that virtually all σ-algebras are Blackwell σ-algebras. However, there do exist non-Blackwell σ-algebras, but for most practical purposes we may assume that the σ-algebras which we are working with are Blackwell σ-algebras. With these concepts at hand, we can give the answer to the question raised in the beginning of this section.

6.6 Atomic Determination

Let $\mathcal{F}$ be a Blackwell σ-algebra on Ω, let $\{\mathcal{G}_j \mid j \in J\}$ be a set of separable sub-σ-algebras of $\mathcal{F}$, and let $\mathcal{G}$ be the σ-algebra

$$\mathcal{G} = \bigcap_{j \in J} \mathcal{G}_j = \{G \subseteq \Omega \mid G \in \mathcal{G}_j \quad \forall j \in J\}$$

Then $A \in \mathcal{G}$ if and only if $A \in \mathcal{F}$ and A is a union of $\mathcal{G}$-atoms. Moreover, if X is a function from Ω into the measurable space $(M, \mathcal{B})$, then X is $\mathcal{G}$-measurable if and only if X is $\mathcal{F}$-measurable and

(6.6.1) $$\omega_0 \equiv \omega_1 (\text{mod } \mathcal{G}) \quad \Rightarrow \quad X(\omega_0) \equiv X(\omega_1) (\text{mod } \mathcal{B})$$

Remark: Note that if X is constant on all $\mathcal{G}$-atoms then (6.6.1) holds. Moreover, if $\mathcal{B}$ separates points in M, then (6.6.1) holds if and only if X is constant on all $\mathcal{G}$-atoms.

Proof: Suppose that $A \in \mathcal{G}$. Then clearly $A \in \mathcal{F}$ and A is a union of $\mathcal{G}$-atoms by (6.3.2). The converse is very difficult and belongs to the deep theory of analytic sets, so I shall skip the proof here (in reference [3] to this chapter, you will find a complete proof).

Suppose that X is $\mathcal{G}$-measurable. Since $\mathcal{F} \supseteq \mathcal{G}$, we have that X is $\mathcal{F}$-measurable and (6.6.1) follows from (6.3.7). Suppose that X is $\mathcal{F}$-measurable

and that (6.6.1) holds. Let $B \in \mathcal{B}$ and set $A = \{X \in B\}$. If $\omega_0 \in A$ and $\omega \in \mathcal{G}(\omega_0)$, then $\omega \equiv \omega_0 (\text{mod } \mathcal{G})$. Hence, by (6.6.1) we have $X(\omega) \equiv X(\omega_0) (\text{mod } \mathcal{B})$ and since $X(\omega_0) \in B \in \mathcal{B}$, we see that $X(\omega) \in B$. Thus, $\omega \in A$ and $\mathcal{G}(\omega_0) \subseteq A$ for all $\omega_0 \in A$. Hence, by (6.3.2) we see that A is a union of $\mathcal{G}$-atoms and since $A \in \mathcal{F}$, then by the first part of the theorem we have that $A = X^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{B}$. □

6.7 Conditional Expectations

Let X be a real random variable defined on the probability space $(\Omega, \mathcal{F}, P)$, then EX is interpreted as that real number which is "the best possible" estimate of X. Suppose that we have some information available given by a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, we may then seek "the best possible" estimate of X based on all our information $\mathcal{G}$. Since the $\mathcal{G}$-measurable random variables are exactly those random variables we can observe when we have the information $\mathcal{G}$, this means that we are seeking that $\mathcal{G}$-measurable random variable Y which is "nearest possible" to X, and we shall call such a random variable Y a *conditional expectation of* X *given* $\mathcal{G}$. But how are we going to define Y in a precise manner? Well, we know from (6.3.8) that Y should be constant on all $\mathcal{G}$-atoms. So let A be a $\mathcal{G}$-atom. Then we shall replace the many values that X may assume on A by one single value, and it is natural to let that value be the relative mean of X over A, that is,

$$Y(\omega) = \frac{1}{P(A)} \int_A X dP \qquad \forall \omega \in A$$

However, if $P(A) = 0$, then the expression on the right-hand side equals $\frac{0}{0}$, which is not defined. So this definition is not possible in general. However, the idea is a sound starting point for the definition. Let $G \in \mathcal{G}$; then by (6.3.2) we have that $G = \cup A_i$ where $\{A_i\}$ are disjoint $\mathcal{G}$-atoms, so if Y is given as above, then a formal (but in general unjustified) computation gives

$$\begin{aligned}\int_G Y\, dP &= \sum_i \int_{A_i} Y\, dP = \sum_i \frac{1}{P(A_i)} \int_{A_i} X\, dP \cdot P(A_i) \\ &= \sum_i \int_{A_i} X\, dP = \int_G X\, dP\end{aligned}$$

In this formal computation only the first and last term are really meaningful, but it leads us to the following definition:

Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a σ-algebra with $\mathcal{G} \subseteq \mathcal{F}$, and let $X \in L(P)$. Then we define *the conditional expectation of* X *given* $\mathcal{G}$ to be that (those) random variable(s) $Y \in L(P)$, which satisfies

(6.7.1) $\quad Y$ is $\mathcal{G}$-measurable

(6.7.2) $$\int_G Y dP = \int_G X dP \quad \forall G \in \mathcal{G}$$

We use the notation $Y := E(X \mid \mathcal{G})$ for any random variable Y satisfying (6.7.1) and (6.7.2). Note that (6.7.1) says that $E(X \mid \mathcal{G})$ is constant on all $\mathcal{G}$-atoms and that (6.7.2) says that the average of $E(X \mid \mathcal{G})$ over a collection of $\mathcal{G}$-atoms equals the average of X over the same collection of $\mathcal{G}$-atoms. Let T be a random function with values in the measurable space $(M, \mathcal{B})$ and let $X \in L(P)$. Then a conditional expectation of X given the σ-algebra $\sigma(T)$ is called a *conditional expectation of* X *given* T, and we write

(6.7.3) $$E(X \mid T) := E(X \mid \sigma(T))$$

The definitions above poses at least two questions: Does there always exist a conditional expectation? And if so, how many? The answer to these questions is given by the following existence and uniqueness theorem:

(6.7.4) If $X \in L(P)$, then there exists a conditional expectation $E(X \mid \mathcal{G})$ of X given $\mathcal{G}$, and any two conditional expectations of X given $\mathcal{G}$ coincides P-a.s.

(6.7.5) If $X \in L(P)$ and $T : (\Omega, \mathcal{F}) \to (M, \mathcal{B})$ is measurable, then there exists a conditional expectation $E(X \mid T)$ of X given T, and any two conditional expectations of X given T coincides P-a.s.

To see this, let us first suppose that $X \geq 0$ and let $\nu(G) = \int_G X\, dP$ for $G \in \mathcal{G}$. Let P_0 be the restriction of P to the σ-algebra $\mathcal{G}$. Then P_0 and ν are measures on $(\Omega, \mathcal{G})$ such that $P_0(N) = 0$ implies $\nu(N) = 0$. Hence, by (3.18) there exists a $\mathcal{G}$-measurable random variable Y satisfying (6.7.2), and so Y is a conditional expectation of X given $\mathcal{G}$. If $X \in L(P)$ is arbitrary, then by the argument above, we know that $E(X^+ \mid \mathcal{G})$ and $E(X^- \mid \mathcal{G})$ exists and evidently $E(X^+ \mid \mathcal{G}) - E(X^- \mid \mathcal{G})$ is a conditional expectation of X given $\mathcal{G}$. Thus, there exist at least one conditional expectation of X given $\mathcal{G}$. Suppose that Y_1 and Y_2 are two conditional expectations of X given $\mathcal{G}$, that is, Y_1 and Y_2 are $\mathcal{G}$-measurable and

$$\int_G Y_1\, dP = \int_G X\, dP = \int_G Y_2\, dP \qquad \forall G \in \mathcal{G}$$

Then by (3.17) we have that $Y_1 = Y_2$ P-a.s. Thus, (6.7.4) is proved and (6.7.5) follows trivially from (6.7.4).

By the factorization theorem (6.4), we know that $E(X \mid T) = \varphi(T)$ for some measurable function $\varphi : (M, \mathcal{B}) \to \mathbf{R}$ and we shall now identify this function.

So let $T : (\Omega, \mathcal{F}) \to (M, \mathcal{B})$ be a random function and let $X \in L(P)$. Then define *the conditional expectation of* X *given* $T = t$ to be that (those) measurable function(s) $\varphi : (M, \mathcal{B}) \to \overline{\mathbf{R}}$ satisfying

$$\int_B \varphi(t) P_T(dt) = \int_{T^{-1}(B)} X dP \qquad \forall B \in \mathcal{B} \tag{6.7.6}$$

where P_T is the distribution law of T . We shall use the notation $\varphi(t) := E(X \mid T = t)$ for any measurable solution φ to equation (6.7.6). Then we have

(6.7.7) If $X \in L(P)$ and $T : (\Omega, \mathcal{F}) \to (M, \mathcal{B})$ is measurable, then there exists a conditional expectation $E(X \mid T = t)$ of X given $T = t$, and any two conditional expectations of X given $T = t$ coincides for P_T-a.a. $t \in M$. Moreover, if $\varphi(t)$ is a conditional expectation of X given $T = t$, then $\varphi(T)$ is a conditional expectation of X given T , that is,

$$E(X \mid T) = \varphi(T) \quad \text{if} \quad \varphi(t) = E(X \mid T = t) \qquad \forall t \in M$$

The existence and uniqueness of $E(X \mid T = t)$ is proved exactly as above. So suppose that $\varphi(t) = E(X \mid T = t)$ is a conditional expectation of X given $T = t$. By (1.38.4) we have that $Y(\omega) = \varphi(T(\omega))$ is $\sigma(T)$-measurable and by (6.7.6) and the first transformation theorem (3.15), we get

$$\begin{aligned}\int_{T^{-1}(B)} X\, dP &= \int_M 1_B(t)\, \varphi(t)\, P_T(dt) \\ &= \int_\Omega 1_B(T(\omega))\, \varphi(T(\omega))\, P(d\omega) \\ &= \int_{T^{-1}(B)} Y\, dP\end{aligned}$$

for all $B \in \mathcal{B}$. Recall that $\sigma(T) = \{T^{-1}(B) \mid B \in \mathcal{B}\}$. So by definition, we have that $\varphi(T) = E(X \mid T)$. Thus, (6.7.7) is proved.

Warning: The conditional expectation $E(X \mid T)$ of X given T should not be confused with the conditional expectation $E(X \mid T = t)$ of X given $T = t$. The two conditional expectations are **not** the same, but they are related by the formula in (6.7.7).

If T_j is a random function with values in $(M_j, \mathcal{B}_j)$ for $j = 1, \ldots, n$, then we may consider the "vector" function

$$T(\omega) = (T_1(\omega), \ldots, T_n(\omega))$$

which maps Ω into the product space $(M_1 \times \cdots \times M_n, \mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_n)$. We define

$$E(X \mid T_1, \ldots, T_n) := E(X \mid T)$$
$$E(X \mid T_1 = t_1, \ldots, T_n = t_n) = E(X \mid T = (t_1, \ldots, t_n))$$

Let $F \in \mathcal{F}$, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let T be a random function with values in $(M, \mathcal{B})$. Then we define *the conditional probability of* F *given* $\mathcal{G}$ *or* T as follows:

$$P(F \mid \mathcal{G}) := E(1_F \mid \mathcal{G}) \quad , \quad P(F \mid T) := E(1_F \mid T)$$
$$P(F \mid T = t) := E(1_F \mid T = t) \qquad \forall t \in M$$

Thus, $P(F \mid \mathcal{G})$ is that $\mathcal{G}$-measurable random variable which satisfies

$$P(F \cap G) = \int_G P(F \mid \mathcal{G})\, dP \qquad \forall G \in \mathcal{G} \tag{6.7.8}$$

Similarly, $P(F \mid T = t)$ is that measurable function from $(M, \mathcal{B})$ into $\overline{\mathbf{R}}$ which satisfies

$$P(F \cap \{T \in B\}) = \int_B P(F \mid T = t)\, P_T(dt) \qquad \forall B \in \mathcal{B} \tag{6.7.9}$$

Let S and T be random functions with values in $(L, \mathcal{A})$ and $(M, \mathcal{B})$ respectively. Then *the conditional distribution of* S *given* $T = t$ is simply the conditional probability $P(S \in A \mid T = t)$ for $A \in \mathcal{A}$ and $t \in T$. Thus, $P(S \in A \mid T = t)$ is that measurable function from $(M, \mathcal{B})$ into $\overline{\mathbf{R}}$ which satisfies

$$P(S \in A, T \in B) = \int_B P(S \in A \,|\, T = t) P_T(dt) \quad \forall A \in \mathcal{A}\ \forall B \in \mathcal{B} \tag{6.7.10}$$

Let $X \in L^1(P)$. Then we define *the conditional variance of* X *given* $\mathcal{G}$, denoted $\sigma^2(X \mid \mathcal{G})$, by

$$\sigma^2(X \mid \mathcal{G}) = E((X - E(X \mid \mathcal{G}))^2 \mid \mathcal{G})$$

and we define *the conditional variance* $\sigma^2(X \mid T)$ *of* X *given* T similarly.

Let $\mathcal{C}$ and $\mathcal{D}$ be two pavings contained in $\mathcal{F}$, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then we say that $\mathcal{C}$ *and* $\mathcal{D}$ *are conditionally independent given* $\mathcal{G}$ if

$$P(C \cap D \mid \mathcal{G}) = P(C \mid \mathcal{G}) P(D \mid \mathcal{G}) \qquad \forall C \in \mathcal{C}\ \forall D \in \mathcal{D} \tag{6.7.11}$$

Similarly, if S , T and U are random functions, then we say that S *and* T *are conditionally independent given* U if $\sigma(S)$ and $\sigma(T)$ are conditionally independent given $\sigma(U)$.

Note that the definition of the conditional expectation $E(X \mid \mathcal{G})$ is descriptive – not constructive. It doesn't tell us how to compute $E(X \mid \mathcal{G})$, but it tells that $E(X \mid \mathcal{G})$ is the (essentially unique) $\mathcal{G}$-measurable solution to the (possibly infinite) system of equations (6.7.2) (one equation for each $G \in \mathcal{G}$) . Observe that if $X, Y \in L^1(P)$, then we have

$$\int_G X\,dP = \int_G Y\,dP \iff \int_G (X^+ + Y^-)\,dP = \int_G (X^- + Y^+)\,dP$$

Hence, by the uniqueness theorem (1.7) we get the following proposition:

(6.7.12) Let $X, Y \in L^1(P)$ be random variables such that $EX = EY$ and Y is $\mathcal{G}$-measurable where $\mathcal{G}$ is a given sub-σ-algebra of $\mathcal{F}$. If $\mathcal{H}$ is a $(\cap f)$-stable paving satisfying

$$\sigma(\mathcal{H}) = \mathcal{G} \quad \text{and} \quad \int_H Y\,dP = \int_H X\,dP \quad \forall H \in \mathcal{H}$$

then $Y = E(X \mid \mathcal{G})$

In Sections (6.8)–(6.11), you will find the basic rules for conditional expectations. They are formulated in mainly terms of $E(X \mid \mathcal{G})$, but, of course, similar rules for $E(X \mid T)$ and $E(X \mid T = t)$ are valid, whenever they are applicable. Moreover, you will find that with the rules of (6.8)–(6.11) it is not only possible to compute conditional expectations in concrete cases; it is also, in general, easy to do so.

6.8 The Calculus of Conditional Expectations

Let $(\Omega, \mathcal{F}, P)$ *be a probability space, let* $\mathcal{H} \subseteq \mathcal{G}$ *be two sub-*σ*-algebras of* $\mathcal{F}$ *, let* X, Y *and* U *be random variables such that* $X \in L(P)$ *,* $Y \in L(P)$ *and* U *is* $\mathcal{G}$*-measurable, and let* $S, T : \Omega \to (M, \mathcal{B})$ *and* $R : \Omega \to (L, \mathcal{L})$ *be random functions. Then we have the following:*

(6.8.1) $E(U \cdot X) = E(U \cdot E(X \mid \mathcal{G}))$ *if* $U \cdot X \in L(P)$

(6.8.2) $E(U \cdot X \mid \mathcal{G}) = U \cdot E(X \mid \mathcal{G})$ *if* $U \cdot X \in L(P)$

(6.8.3) $E(U \mid \mathcal{G}) = U$ *a.s. if* $U \in L(P)$

(6.8.4) $E(E(X \mid \mathcal{G})) = EX$

(6.8.5) $E(X \mid \mathcal{G}) = EX$ *a.s. if* X *and* $\mathcal{G}$ *are independent*

(6.8.6) $E(X \mid \mathcal{G}) \leq E(Y \mid \mathcal{G})$ *a.s.* $\Leftrightarrow \int_G X\,dP \leq \int_G Y\,dP \;\; \forall G \in \mathcal{G}$

(6.8.7) *Let* $a, b \in \mathbf{R}$ *be real numbers such that* $aX(\omega) + bY(\omega)$ *and* $aE(X \mid \mathcal{G})(\omega) + bE(Y \mid \mathcal{G})(\omega)$ *are well-defined (that is, not of the form* $\infty - \infty$ *) for a.a.* $\omega \in \Omega$. *If* $aX + bY \in L(P)$ *, then*

$$E(aX + bY \mid \mathcal{G}) = a\,E(X \mid \mathcal{G}) + b\,E(Y \mid \mathcal{G})$$

(6.8.8) $E(X \mid \mathcal{H}) = E(E(X \mid \mathcal{H}) \mid \mathcal{G}) = E(E(X \mid \mathcal{G}) \mid \mathcal{H})$

(6.8.9) $E(X \mid \mathcal{H}) = E(X \mid \mathcal{G}) \;\Leftrightarrow\; E(X \mid \mathcal{G})$ *is* $\mathcal{H}$*-measurable*

(6.8.10) $E(X \mid T = x) = E(Y \mid S = x)$ P_T*-a.s. if* $(X, T) \sim (Y, S)$

(6.8.11) $E(X \mid T) = E(Y \mid T)$ *a.s. if* $(X, T) \sim (Y, T)$

(6.8.12) *If* $X \leq Y$ *a.s., then* $E(X \mid \mathcal{G}) \leq E(Y \mid \mathcal{G})$

(6.8.13) *If* $a, b \in \overline{\mathbf{R}}$ *and* $a \leq X \leq b$ *a.s., then* $a \leq E(X \mid \mathcal{G}) \leq b$

(6.8.14) *Let* $f : L \times M \to \overline{\mathbf{R}}$ *be measurable such that* $f(R, S) \in L(P)$ *and* $f(R, x) \in L(P)$ *for* P_S*-a.a.* $x \in M$ *, and let* $\phi(x) = E(f(R, x))$ *. If* R *and* S *are independent, then we have*

$$E(f(R, S) \mid S = x) = \phi(x) \quad \textit{for } P_S\textit{-a.a.} \quad x \in M$$

Moreover, if S *is* $\mathcal{G}$*-measurable and* R *and* $\mathcal{G}$ *are independent, then we have*

$$E(f(R, S) \mid \mathcal{G}) = \phi(S) \qquad P\textit{-a.s.}$$

Remark: (6.8.3) says that if we know U , then U itself is the best estimate of U (of course!!). If we are totally ignorant, then only we know that Ω will occur, and $\emptyset$ will not occur. Thus, total ignorance corresponds to the σ-algebra $\mathcal{G}_0 = \{\emptyset, \Omega\}$, and since $\mathcal{G}_0$ is independent of every random variable, we have by (6.8.5) that $E(X \mid \mathcal{G}_0)$ is identically equal to EX . Observe that (6.8.5) tells us that if our information $\mathcal{G}$ is independent of X , then the best estimate of X is the same as if we were totally ignorant. Thus, if X and $\mathcal{G}$ are independent, then $\mathcal{G}$ carries no information whatsoever about the random variable X . Proposition (6.8.14) is a version of the so-called *useful rule* and, loosely speaking, it says that we may use whatever we know. In (10.3) of Volume II you will find a more general version of the useful rule.

Proof: (6.8.1): Suppose that $E(UX)^+ < \infty$ [the case $E(UX)^- < \infty$ is treated symmetrically] and set $V_0 = E(X^+ \mid \mathcal{G})$, $V_1 = E(X^- \mid \mathcal{G})$. Then evidently we have that $E(X \mid \mathcal{G}) = V_0 - V_1$, and since $(UX)^+ = (UX^+)^+ + (UX^-)^-$, we see that $E(UX^+) < \infty$ and $E(UX^-) > -\infty$. Hence, by (6.7.1+2) and (3.17.2) we have $E(UX^+) = E(UV_0)$ and $E(UX^-) = E(UV_1)$. Thus, by (3.3.11+12) we get

$$E(UX) = E\big(UX^+\big) - E\big(UX^-\big) = E(U(V_0 - V_1)) = E(U \cdot E(X \mid \mathcal{G}))$$

(6.8.2): Let $V = UE(X \mid \mathcal{G})$. Then V is $\mathcal{G}$-measurable, and if $G \in \mathcal{G}$, then applying (6.8.1) with U replaced by $1_G U$, we get

$$\int_G V dP = E(1_G U\, E(X \mid \mathcal{G})) = \int_G U\, X dP$$

Hence. by definition (6.7.1+2) we see that $V = E(UX \mid \mathcal{G})$.

(6.8.3): Trivial consequence of definition (6.7.1+2).

(6.8.4): Set $G = \Omega$ in (6.7.2).

(6.8.5): Set $\mu = EX$ and $V(\omega) = \mu$ for all ω . Then V is $\mathcal{G}$-measurable, and by (4.5.5) we have

$$\int_G X dP = E(1_G X) = \mu P(G) = \int_G V dP \qquad \forall G \in \mathcal{G}$$

because X and 1_G are independent for all $G \in \mathcal{G}$. Hence, by definition (6.7.1+2) we see that $V = E(X \mid \mathcal{G})$.

(6.8.6.): The implication " $\Rightarrow$ " follows directly from definition (6.7.1+2) and the implication " $\Leftarrow$ " follows from (3.17).

(6.8.7): First note that $E(a X \mid \mathcal{G}) = a E(X \mid G)$ by (3.3.11) and definition (6.7.1+2). Hence, we may assume that $a = b = 1$. Let

$$Z = E(X + Y \mid \mathcal{G}) \quad , \quad V = E(X \mid \mathcal{G}) \quad , \quad W = E(Y \mid \mathcal{G})$$

and note that if $EX < \infty$ and $EY < \infty$, then $V + W$ is a $\mathcal{G}$-measurable function which by (3.3.12) satisfies (6.7.2). Thus, $Z = V + W$. In general we set $G_n = \{V \le n \ , \ W \le n\}$, then $G_n \in \mathcal{G}$ and by (6.8.1), we have

$$E(1_{G_n} X) = E(1_{G_n} V) \le n \quad , \quad E(1_{G_n} Y) = E(1_{G_n} W) \le n$$

Hence, by (6.8.2) and the argument above we have $1_{G_n} Z = 1_{G_n}(V+W)$. Setting $C = \{V < \infty \ , \ W < \infty\}$ we see that $G_n \uparrow C$, and so $Z = V + W$ P-a.s

on C . Similarly, if $D = \{V > -\infty\ ,\ W > -\infty\}$, then $Z = V + W$ P-a.s. on D . Thus, $Z = V + W$ P-a.s. on $C \cup D$ and

$$\Omega \setminus (C \cup D) = \{V = \infty,\, W = -\infty\} \cup \{V = -\infty,\, W = \infty\}$$

By assumption we know that $V(\omega) + W(\omega)$ is not of the form $\infty - \infty$ for a.a. ω . Hence, $P(\Omega \setminus (C \cup D)) = 0$ and $Z = V + W$ P-a.s.

(6.8.8): Since $E(X \mid \mathcal{H})$ is $\mathcal{H}$-measurable and $\mathcal{H} \subseteq \mathcal{G}$, then $E(X \mid \mathcal{H})$ is $\mathcal{G}$-measurable and so the first equality follows from (6.8.3). Moreover, if $H \in \mathcal{H}$, then $H \in \mathcal{G}$ and so

$$\int_H E(X \mid \mathcal{H})\, dP = \int_H X\, dP = \int_H E(X \mid \mathcal{G})\, dP$$

Hence, $E(X \mid \mathcal{H}) = E(E(X \mid \mathcal{G}) \mid \mathcal{H})$ by definition (6.7.1+2).

(6.8.9): The implication " $\Rightarrow$ " follows by definition and the implication " $\Leftarrow$ " follows from (6.8.8) and (6.8.3).

(6.8.10): Let $\varphi(t) = E(X \mid T = t)$ and $\psi(t) = E(Y \mid L = t)$. Since $P_{(X,T)} = P_{(Y,L)}$, then by (6.7.6) and (3.15) we have

$$\begin{aligned}\int_B \varphi(t)\, P_T(dt) &= \int_\Omega 1_B(T(\omega))\, X(\omega)\, P(d\omega) = \int_{\mathbf{R}\times M} 1_B(t)\, v\, P_{(X,T)}(dv, dt)\\ &= \int_\Omega 1_B(L(\omega))\, Y(\omega)\, P(d\omega) = \int_B \psi(t)\, P_L(dt)\end{aligned}$$

for all $B \in \mathcal{B}$ and since $P_T = P_L$, then by (3.17) we have $\varphi = \psi$ P_T-a.s.

(6.8.11): Follows from (6.8.10) and (6.7.7).

(6.8.12+13): Follows from (6.8.6).

(6.8.14): Let $B \in \mathcal{B}$ be given and set $g(u,x) = 1_B(x)\, f(u,x)$, then $Eg(R,x) = 1_B(x)\, \phi(x)$. Hence, by (4.5.2) we have

$$\int_{S^{-1}(B)} f(R,S)\, dP = Eg(R,S) = E(1_B(S)\, \phi(S)) = \int_B \phi(x)\, P_S(dx)$$

Thus, $\phi(x) = E(f(R,S) \mid S = x)$ by definition (6.7.6). Suppose that R and $\mathcal{G}$ are independent and let $G \in \mathcal{G}$ be given, then $(1_G, S)$ and R are independent. Hence, if $h(u,x) = u\, \phi(x)$, then by (4.5.2) we find

$$\int_G f(R,S)\, dP = E(1_G\, f(R,S)) = Eh(1_G, S) = \int_G \phi(S)\, dP$$

and since $\phi(S)$ is $\mathcal{G}$-measurable, we see that the second part of (6.8.14) follows from the definition (6.7.1+2). $\square$

6.9 Inequalities for Conditional Expectations

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let X and U be a random variables such that U is $\mathcal{G}$-measurable and non-negative. If $f : \overline{\mathbf{R}}_+ \to \overline{\mathbf{R}}_+$ is an increasing function such that $f(v) > 0$ for all $v > 0$ and $f(U) < \infty$ a.s. on the set $\{U < \infty\}$, then we have

(6.9.1) $$P(|X| > U \mid \mathcal{G}) \leq f(U)^{-1} E(f(|X|) \mid \mathcal{G})$$

(6.9.2) $$P(|X - E(X \mid \mathcal{G})| > U \mid \mathcal{G}) \leq U^{-2} \sigma^2(X \mid \mathcal{G}) \quad \text{if} \quad X \in L^1(P)$$

Moreover, let $\varphi :]a, b[\to \mathbf{R}$ be a convex function and set $\varphi(a) := \lim_{t \downarrow a} \varphi(t)$ and $\varphi(b) := \lim_{t \uparrow b} \varphi(t)$ [the limits exist by convexity of φ, but may possibly be $\pm\infty$; see Exc.3.33]. If $X \in L(P)$ is a random variable such that $P(a \leq X \leq b) = 1$ and $\varphi(X) \in L(P)$, then we have

(6.9.3) $$\varphi(E(X \mid \mathcal{G})) \leq E(\varphi(X) \mid \mathcal{G})$$

In particular, setting $\varphi(t) = |t|^q$, we get

(6.9.4) $$|E(X \mid \mathcal{G})|^q \leq E(|X|^q \mid \mathcal{G}) \quad \text{if } X \in L(P) \text{ and } 1 \leq q < \infty$$

(6.9.5) $$E|E(X \mid \mathcal{G})|^q \leq E|X|^q \quad \text{if } X \in L(P) \text{ and } 1 \leq q < \infty$$

Remark: Note that (6.9.1) is a version of *Markov's inequality* for conditional expectations. Similarly, (6.9.2) and (6.9.3) are versions of *Chebyshev's* and *Jensen's inequalities* for conditional expectations.

Proof: (6.9.1): Note that $1_{\{|X|>U\}} \leq f(U)^{-1} f(|X|)$ and so (6.9.1) follows from (6.8.12) and (6.8.2).

(6.9.2): Apply Markov's inequality (6.9.1) to $X - E(X \mid \mathcal{G})$ with $f(t) = t^2$.

(6.9.3): Let $V = E(X \mid \mathcal{G})$ and $W = E(\varphi(X) \mid \mathcal{G})$. By convexity of φ, we know that the right derivative of φ exists and is increasing on $]a, b[$ [see Exc.3.33]. So let $\varphi'(s)$ denote the right derivative of φ. By (6.8.13), we have that $a \leq V \leq b$. Since a convex function lies above all its tangents, we have

(i) $$\varphi(t) \geq \varphi'(s)(t - s) + \varphi(s) \qquad \forall s \in]a, b[\quad \forall t \in [a, b]$$

Let $\{I_m\}$ be an increasing sequence of compact intervals such that $I_m \uparrow]a, b[$, and set

$$G_m = \{ V \in I_m \} \quad , \quad V_m = 1_{G_m} V \quad , \quad X_m = 1_{G_m} X$$

Then V_m , $1_{G_m}\varphi'(V)$ and $1_{G_m}\varphi(V)$ are bounded $\mathcal{G}$-measurable random variables. Since $G_m \in \mathcal{G}$, we have $EX_m = EV_m$. Hence, $X_m \in L^1(P)$, so by (i) we find

$$1_{G_m}\varphi(X) \geq \varphi'(V)(X_m - V_m) + 1_{G_m}\varphi(V)$$

and since $V = E(X \mid \mathcal{G})$, then by (6.8.7), (6.8.2) and (6.8.3) we get

$$\begin{aligned} 1_{G_m}W &= E(1_{G_m}\varphi(X) \mid \mathcal{G}) \\ &\geq \varphi'(V)(E(X_m \mid \mathcal{G}) - V_m) + 1_{G_m}\varphi(V) \\ &= 1_{G_m}\varphi'(V)(E(X \mid \mathcal{G}) - V) + 1_{G_m}\varphi(V) \\ &= 1_{G_m}\varphi(V) \end{aligned}$$

Thus, we see that $W(\omega) \geq \varphi(V(\omega))$ for a.a. $\omega \in G_m$ and since $I_m \uparrow]a,b[$, we see that $W(\omega) \geq \varphi(V(\omega))$ for P-a.a. $\omega \in \{a < V < b\}$. If φ is decreasing then $\varphi(X) \geq \varphi(b)$ and so $W \geq \varphi(b)$ a.s. Thus, $W(\omega) \geq \varphi(V(\omega))$ for a.a. $\omega \in \{V = b\}$. So suppose that φ is not decreasing; then there exists $s_0 \in]a,b[$ such that $\varphi'(s_0) \geq 0$. Since φ' is increasing, we have $\varphi'(s) \geq 0$ for all $s_0 \leq s < b$. By (i), we have

$$\varphi(X) \geq \varphi'(s)(X - s) + \varphi(s) \qquad \forall s \in]a,b[$$

and so by (6.8.7), we have

$$W(\omega) \geq \varphi'(s)(V(\omega) - s) + \varphi(s) \geq \varphi(s)$$

for all $\omega \in \{V = b\}$ and all $s_0 \leq s < b$. Letting $s \uparrow b$, we see that $W(\omega) \geq \varphi(V(\omega))$ for P-a.a. $\omega \in \{V = b\}$. Similarly, we find that $W(\omega) \geq \varphi(V(\omega))$ for P-a.a. $\omega \in \{V = a\}$. Hence, $W \geq \varphi(V)$ P-a.s., and so (6.9.3) is proved.

(6.9.4): Apply (6.9.3) with $\varphi(t) = |t|^q$.

(6.9.5): Apply (6.9.4) and (6.8.4). □

6.10 Convergence of Conditional Expectations

Let $X, X_1, X_2, \ldots$ be random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a given sub-σ-algebra of $\mathcal{F}$. Let $\Sigma(\mathcal{F})$ be the set of all sub-σ-algebras of $\mathcal{F}$, and suppose that $1 \leq q < \infty$ and that $\mathcal{V} \subseteq L^1(P)$ is a uniformly integrable set of functions. Then we have the following:

(6.10.1) $\mathcal{W} := \{E(X \mid \mathcal{H}) \mid X \in \mathcal{V}\ ,\ \mathcal{H} \in \Sigma(\mathcal{F})\}$ *is uniformly integrable*

(6.10.2) *If* $X_n \to X$ *in q-mean, then* $E(X_n \mid \mathcal{G}) \to E(X \mid \mathcal{G})$ *in q-mean*

(6.10.3) *If* $X_n \uparrow X$ *and* $EX_1 > -\infty$, *then* $E(X_n \mid \mathcal{G}) \uparrow E(X \mid \mathcal{G})$ *a.s.*

(6.10.4) *If* $X_n \downarrow X$ *and* $EX_1 < \infty$, *then* $E(X_n \mid \mathcal{G}) \downarrow E(X \mid \mathcal{G})$ *a.s.*

(6.10.5) *If* $\sum_{n=1}^{\infty} EX_n^+ < \infty$, *then* $E\left(\sum_{n=1}^{\infty} X_n \,\middle|\, \mathcal{G}\right) = \sum_{n=1}^{\infty} E(X_n \mid \mathcal{G})$ *a.s.*

(6.10.6) *If* $\sum_{n=1}^{\infty} EX_n^- < \infty$, *then* $E\left(\sum_{n=1}^{\infty} X_n \,\middle|\, \mathcal{G}\right) = \sum_{n=1}^{\infty} E(X_n \mid \mathcal{G})$ *a.s*

(6.10.7) *If* $X_n \to X$ *a.s. and* $|X_n| \le Y$ *a.s. for all* $n \ge 1$ *for some* $Y \in L^1(P)$, *then* $E(X_n \mid \mathcal{G}) \to E(X \mid \mathcal{G})$ *a.s. and in 1-mean*

(6.10.8) $\liminf_{n\to\infty} E(X_n \mid \mathcal{G}) \ge E\left(\liminf_{n\to\infty} X_n \,\middle|\, \mathcal{G}\right)$ *a.s. if* $\sup_n X_n^- \in L^1(P)$

(6.10.9) $\limsup_{n\to\infty} E(X_n|\mathcal{G}) \le E\left(\limsup_{n\to\infty} X_n \,\middle|\, \mathcal{G}\right)$ *a.s. if* $\sup_n X_n^+ \in L^1(P)$

(6.10.10) *If* $\{X_n^+\}$ *is uniformly integrable and* $\limsup_{n\to\infty} X_n \le U$ *for some* $U \in L(P)$, *then* $\liminf_{n\to\infty} E(X_n \mid \mathcal{G}) \le E(U \mid \mathcal{G})$ *a.s.*

(6.10.11) *If* $\{X_n^-\}$ *is uniformly integrable, and* $\liminf_{n\to\infty} X_n \ge V$ *for some* $V \in L(P)$, *then* $\limsup_{n\to\infty} E(X_n \mid \mathcal{G}) \ge E(V \mid \mathcal{G})$ *a.s.*

Remark: Thus, all the convergence theorems for integrals (monotone, Fatou, Lebesgue, Beppo Levi and Dunford-Pettis) carry over to conditional expectations with one small but noticeable difference: You **cannot** replace $\liminf$ by $\limsup$ in (6.10.10) and you **cannot** replace $\limsup$ by $\liminf$ in (6.10.11).

Proof: (6.10.1): Let $\nu(F) = \sup\{\int_F |X|\, dP \mid X \in \mathcal{V}\}$, then ν is finite and P-continuous by (3.24.2). Let $X \in \mathcal{V}$ and $\mathcal{H} \in \Sigma(\mathcal{F})$ and set $U = E(X \mid \mathcal{H})$. Then by (6.9.3) we have

$$\int_H |U|\, dP \le \int_H E(|X| \mid \mathcal{H})\, dP = \int_H |X|\, dP \le \nu(H) \qquad \forall H \in \mathcal{H}$$

and since U is $\mathcal{H}$-measurable, we have that $\mathcal{H} \supseteq \sigma(U)$. Hence, by (3.24.3) we see that $\mathcal{W}$ is uniformly integrable.

(6.10.2): Follows directly from (6.9.5)

(6.10.3): Let $U_n = E(X_n \mid \mathcal{G})$. Then $U_1 \le U_2 \le \cdots$ by (6.8.12) and so $U_n \uparrow U$ for some $\mathcal{G}$-measurable random variable U . Let $G \in \mathcal{G}$; then by the

monotone convergence theorem (3.4) we have

$$\int_G X\,dP = \lim_{n\to\infty} \int_G X_n\,dP = \lim_{n\to\infty} \int_G U_n\,dP = \int_G U\,dP$$

Hence, $U = E(X \mid \mathcal{G})$ by definition (6.7.1+2).

(6.10.4): Follows from (6.10.3).

(6.10.5): Since $E(Y \mid \mathcal{G})^+ \le E(Y^+ \mid \mathcal{G})$ by (6.9.3), then (6.10.5) follows from Beppo Levi's theorem (3.7) exactly as above.

(6.10.6): Follows from (6.10.5).

(6.10.8): Let $Y_n = \inf_{j\ge n} X_j$ and $X_* = \liminf X_n$, then $Y_n \uparrow X_*$ and $EY_1 < -\infty$. Since $Y_n \le X_n$, then by (6.10.3) we have

$$E(X_* \mid \mathcal{G}) = \lim_{n\to\infty} E(Y_n \mid \mathcal{G}) \le \liminf_{n\to\infty} E(X_n \mid \mathcal{G})$$

(6.10.9): Follows from (6.10.8).

(6.10.7): Follows from (6.10.8) and (6.10.9).

(6.10.10): Let $U_n = E(X_n \mid \mathcal{G})$, $V_n = \inf_{j\ge n} U_j$. Then by Dunford-Pettis' theorem (3.23) we have

$$\int_G V_n\,dP \le \limsup_{j\to\infty} \int_G U_j\,dP \le \limsup_{j\to\infty} \int_G X_j\,dP \le \int_G U\,dP$$

for all $G \in \mathcal{G}$ because $V_n \le U_j$ for all $j \ge n$. Hence, by (6.8.6) we have that $V_n \le E(U \mid \mathcal{G})$, and since $V_n \uparrow \liminf_{j\to\infty} U_j$, we see that (6.10.10) follows.

(6.10.11): Follows from (6.10.10). □

6.11 Conditional Densities

Let $(L, \mathcal{A}, \mu)$ and $(M, \mathcal{B}, \nu)$ be sum-finite measure spaces, and let S and T be random functions with values in $(L, \mathcal{A})$ and $(M, \mathcal{B})$ such that (S, T) is absolutely $\mu \otimes \nu$-continuous with $\mu \otimes \nu$-density $f_{(S,T)}(u, v)$. Let $f_T(v)$ denote the ν-density of T and set

$$f_{S|T}(u|v) = \frac{f_{(S,T)}(u,v)}{f_T(v)} \qquad \forall (u,v) \in S \times T \tag{6.11.1}$$

with the convention: $\frac{a}{0} = 0$ for all $a \ge 0$. Then we have

$$P(S \in A \mid T = v) = \int_A f_{S|T}(u|v)\,\mu(ds) \qquad \forall A \in \mathcal{A} \tag{6.11.2}$$

$$E(\psi(S,T) \mid T = v) = \int_L \psi(u,v)\,f_{S|T}(u|v)\,\mu(du) \tag{6.11.3}$$

for every measurable function $\psi : S \times T \to \overline{\mathbf{R}}$ *with* $\psi(S,T) \in L(P)$.

Conversely, if T *is absolutely* ν*-continuous with* ν*-density* f_T *and (6.11.2) holds for some non-negative* $(\mathcal{A} \otimes \mathcal{B})$*-measurable function* $f_{S|T}(u|v)$ *, then* (S,T) *is absolutely* $\mu \otimes \nu$*-continuous with* $\mu \otimes \nu$*-density* $f_{(S,T)}$ *given by:*

$$f_{(S,T)}(u,v) = f_{S|T}(u|v)\, f_T(v) \qquad \forall\, (u,v) \in S \times T \tag{6.11.4}$$

Proof: (6.11.2+3): Let $\psi : S \times T \to [0,\infty]$ be a non-negative measurable function and let $B \in \mathcal{B}$ be a given set. Since $f_T(v) > 0$ P_T-a.s. there exists a P_T-null set $N \in \mathcal{B}$ such that $f_{(S,T)}(u,v) = f_{S|T}(u|v)\, f_T(v)$ for all $(u,v) \in L \times (M \setminus N)$. Hence, if $\Psi(v)$ denotes the right-hand side of (6.11.3), then by (4.2.2) Tonelli's theorem (3.21.1) we get

$$\begin{aligned}
\int_B \Psi(v)\, P_T(dv) &= \int_B \Psi(v)\, f_T(v)\, \nu(dv) \\
&= \int_L \mu(du) \int_B \psi(u,v)\, f_{S|T}(u|v)\, f_T(v)\, \nu(dv) \\
&= \int_L \mu(du) \int_B \psi(u,v)\, f_{(S,T)}(u,v)\, \nu(dv) \\
&= \int_{T^{-1}(B)} \psi(S,T)\, dP
\end{aligned}$$

Thus, Ψ satisfies (6.7.6) and so (6.11.3) holds whenever ψ is non-negative, but then the general result follows easily by splitting ψ into its positive and negative part. Moreover, setting $\psi(u,v) = 1_A(u)\, 1_B(v)$ in (6.11.3), we see that (6.11.2) holds.

Suppose that T is absolutely ν-continuous with ν-density f_T and that (6.11.2) holds for some non-negative $(\mathcal{A} \otimes \mathcal{B})$-measurable function $f_{S|T}(u|v)$. Let $A \in \mathcal{A}$ and $B \in \mathcal{B}$ be given. Then by (6.7.10) we find

$$\begin{aligned}
P(S \in A\ ,\ T \in B) &= \int_B P_T(dv) \int_A f_{S|T}(u|v)\, \mu(du) \\
&= \int_{A \times B} f_{S|T}(u|v)\, f_T(v)\, (\mu \otimes \nu)(du, dv)
\end{aligned}$$

and since $\{A \times B \mid A \in \mathcal{A}\ ,\ B \in \mathcal{B}\}$ is an $(\cap f)$-stable paving generating the product σ-algebra $\mathcal{A} \otimes \mathcal{B}$, then by the uniqueness theorem (1.7) we conclude that (S,T) is absolutely $\mu \otimes \nu$-continuous with $\mu \otimes \nu$-density $f_{S|T}(u|v)\, f_T(v)$. $\square$

6.12 Conditionally Independent Pavings

Let $(\Omega, \mathcal{F}, P)$ be probability spaces, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $\mathcal{C}$ and $\mathcal{D}$ be two pavings contained in $\mathcal{F}$. Set $\mathcal{C}^* = \sigma(\mathcal{G} \cup \mathcal{C})$ and $\mathcal{D}^* = \sigma(\mathcal{G} \cup \mathcal{D})$. If $\mathcal{D}$ is $(\cap f)$-stable, then the following three statements are equivalent:

(6.12.1) *$\mathcal{C}$ and $\mathcal{D}$ are conditionally independent given $\mathcal{G}$*

(6.12.2) $P(C \mid \mathcal{D}^*) = P(C \mid \mathcal{G})$ *a.s.* $\forall C \in \mathcal{C}$

(6.12.3) *$\mathcal{C}$ and $\mathcal{D}^*$ are conditionally independent given $\mathcal{G}$*

Let $L(P, \mathcal{H})$ denote the set of all $\mathcal{H}$-measurable functions $Z \in L(P)$ whenever $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{F}$. If $\mathcal{C}$ and $\mathcal{D}$ both are $(\cap f)$-stable, then the following five statements are equivalent:

(6.12.4) *$\mathcal{C}$ and $\mathcal{D}$ are conditionally independent given $\mathcal{G}$*

(6.12.5) *$\mathcal{C}^*$ and $\mathcal{D}^*$ are conditionally independent given $\mathcal{G}$*

(6.12.6) $E(X \mid \mathcal{D}^*) = E(X \mid \mathcal{G})$ *a.s.* $\forall X \in L(P, \mathcal{C}^*)$

(6.12.7) $E(Y \mid \mathcal{C}^*) = E(Y \mid \mathcal{G})$ *a.s.* $\forall Y \in L(P, \mathcal{D}^*)$

(6.12.8) $E(XY \mid \mathcal{G}) = E(X \mid \mathcal{G})\, E(Y \mid \mathcal{G})$ *a.s. whenever* $X \in L(P, \mathcal{C}^*)$, $Y \in L(P, \mathcal{D}^*)$ *and* $X \cdot Y \in L(P)$

Proof: (6.12.1) $\Rightarrow$ (6.12.2): Let $\mathcal{H} = \{D \cap G \mid D \in \mathcal{D}_0\ ,\ G \in \mathcal{G}\}$ where $\mathcal{D}_0 = \mathcal{D} \cup \{\Omega\}$. Note that $\mathcal{C}$ and $\mathcal{D}_0$ are conditionally independent given $\mathcal{G}$ and that $\mathcal{H}$ is a $(\cap f)$-stable paving with $\sigma(\mathcal{H}) = \mathcal{D}^*$. Let $C \in \mathcal{C}$ and $H \in \mathcal{H}$ be given. Then $H = D \cap G$ for some $D \in \mathcal{D}_0$ and some $G \in \mathcal{G}$ and by (6.8) and (6.7.11), we have

$$\begin{aligned} P(C \cap H) &= E(P(C \cap D \cap G \mid \mathcal{G})) = E(1_G\, P(C \mid \mathcal{G})\, P(D \mid \mathcal{G})) \\ &= E(E(P(C \mid \mathcal{G})\, 1_{D \cap G} \mid \mathcal{G})) \\ &= \int_{D \cap G} P(C \mid \mathcal{G})\, dP = \int_H P(C \mid \mathcal{G})\, dP \end{aligned}$$

and since $P(C) = EP(C \mid \mathcal{G})$, we see that (6.12.2) follows from (6.7.12).

(6.12.2) $\Rightarrow$ (6.12.3): Let $C \in C$ and $A \in \mathcal{D}^*$ be given. Then by assumption we have that $P(C \mid \mathcal{G}) = P(C \mid \mathcal{D}^*)$ a.s. Hence, by (6.8) we get

$$\begin{aligned} P(C \cap A \mid \mathcal{G}) &= E(E(1_C\, 1_A \mid \mathcal{D}^*) \mid \mathcal{G}) = E(1_A\, E(1_C \mid \mathcal{D}^*) \mid \mathcal{G}) \\ &= E(1_A\, P(C \mid \mathcal{G}) \mid \mathcal{G}) = P(C \mid \mathcal{G}) P(A \mid \mathcal{G}) \end{aligned}$$

Thus, (6.12.3) holds.

(6.12.3) $\Rightarrow$ (6.12.1): Evident!

(6.12.4) $\Rightarrow$ (6.12.5): Follows from the equivalence of (6.12.1) and (6.12.3).

(6.12.5) $\Rightarrow$ (6.12.6): Let $\mathcal{V}$ be the set of all non-negative random variables Z such that $E(Z \mid \mathcal{D}^*) = E(Z \mid \mathcal{G})$. Then by (6.8.7) and (6.10.3) we have that $\mathcal{V}$ is a $(\uparrow)$-stable convex cone, and by (6.12.2) and assumption we have that $1_A \in V$ for all $A \in \mathcal{C}^*$. Hence, by (1.46.2) we have that $\mathcal{V}$ contains all non-negative $\mathcal{C}^*$-measurable functions. But then (6.12.6) follows easily by splitting X into its positive and negative parts.

(6.12.6) $\Rightarrow$ (6.12.7): Let $Y \in L(P)$ be a $\mathcal{D}^*$-measurable random variable and let $A \in \mathcal{C}^*$ be given. Then by (6.8.1), (6.8.4) and assumption we get

$$\begin{aligned}\int_A E(Y \mid \mathcal{G})\, dP &= E(E(Y \mid \mathcal{G})\, E(1_A \mid \mathcal{G})) = E(Y\, E(1_A \mid \mathcal{G})) \\ &= E(Y\, E(1_A \mid \mathcal{D}^*)) = E(E(Y\, 1_A \mid \mathcal{D}^*)) \\ &= \int_A Y\, dP\end{aligned}$$

and since $E(Y \mid \mathcal{G})$ is $\mathcal{C}^*$-measurable, we see that (6.12.7) follows.

(6.12.7) $\Rightarrow$ (6.12.8): Let $X, Y \in L(P)$ satisfy the conditions of (6.12.8). Then by assumption and (6.8) we have

$$\begin{aligned}E(XY \mid \mathcal{G}) &= E(E(XY \mid \mathcal{C}^*) \mid \mathcal{G}) = E(XE(Y \mid \mathcal{C}^*) \mid \mathcal{G}) \\ &= E(XE(Y \mid \mathcal{G}) \mid \mathcal{G}) = E(X \mid \mathcal{G})E(Y \mid \mathcal{G})\end{aligned}$$

Thus, (6.12.8) holds.

(6.12.8) $\Rightarrow$ (6.12.4): Evident! □

6.13 The Weak Markov Property

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{F}_n$ and $\mathcal{G}_n$ be sub-σ-algebras of $\mathcal{F}$ for each $n = 0, 1, 2 \ldots$. Then we say that $\{\mathcal{F}_n, \mathcal{G}_n\}$ has *the weak Markov property* if

(6.13.1) $\quad \mathcal{G}_n \subseteq \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \qquad \forall n \geq 0$

(6.13.2) $\quad P(G \mid \mathcal{F}_n) = P(G \mid \mathcal{G}_n) \qquad \forall G \in \mathcal{G}_{n+1} \quad$ a.s. $\quad \forall n \geq 0$

This has the following interpretation: Let $n = 0, 1, \ldots$ denote the time, let $\mathcal{G}_n$ be *the present information* (that is, the information based on observation of day number n) and let $\mathcal{F}_n$ be *the past information* (that is, the information based on observations of the days $0, 1, \ldots, n$). Then the weak Markov property expresses that if we know the present $\mathcal{G}_n$, then the past $\mathcal{F}_n$ does not give any further information about the immediate future $\mathcal{G}_{n+1}$. Note that if (6.13.1) holds, then by (6.12) we have that (6.13.2) holds if and only if $\mathcal{F}_n$ and $\mathcal{G}_{n+1}$ are conditionally independent given $\mathcal{G}_n$, that is, if and only if the past and the immediate future are independent given the present.

Let $\{\mathcal{F}_n, \mathcal{G}_n \mid n = 0, 1, \ldots\}$ be a given collection of sub-σ-algebras of $\mathcal{F}$ satisfying (6.13.1). Then we define *the future information*, denoted $\mathcal{F}^n$, as follows:

$$\mathcal{F}^n = \sigma\left(\bigcup_{j=n}^{\infty} \mathcal{G}_j\right) \qquad \forall n = 0, 1, \ldots \tag{6.13.3}$$

Let $X_0, X_1, X_2, \ldots$ be a sequence of random functions with values in the measurable space $(M, \mathcal{B})$, and let $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$ be a sequence of sub-σ-algebras of $\mathcal{F}$. Then we say that $\{X_n, \mathcal{F}_n \mid n \geq 0\}$ is a *weak Markov chain* if $\{\mathcal{F}_n, \sigma(X_n) \mid n \geq 0\}$ has the weak Markov property. Note that if $\mathcal{G}_n = \sigma(X_n)$, then the future σ-algebra $\mathcal{F}^n$ from (6.13.3) is given by

$$\mathcal{F}^n = \sigma(X_n, X_{n+1}, \ldots) \qquad \forall n \geq 0 \tag{6.13.4}$$

6.14 The Markov Equivalence Theorem

Suppose that $\{\mathcal{F}_n, \mathcal{G}_n \mid n = 0, 1, 2, \ldots\}$ are sub-σ-algebras of $\mathcal{F}$ satisfying (6.13.1), then the following four statements are equivalent:

(6.14.1) $\{\mathcal{F}_n, \mathcal{G}_n\}$ *has the weak Markov property*

(6.14.2) $E(X \mid \mathcal{F}_n) = E(X \mid \mathcal{G}_n)$ *a.s.* $\forall n \geq 0 \; \forall X \in L(P, \mathcal{F}^n)$

(6.14.3) $\mathcal{F}_n$ *and* $\mathcal{G}_{n+1}$ *are conditionally independent given* $\mathcal{G}_n$ $\forall n \geq 0$

(6.14.4) $\mathcal{F}_n$ *and* $\mathcal{F}^n$ *are conditionally independent given* $\mathcal{G}_n$ $\forall n \geq 0$

where $\mathcal{F}^n$ is the future σ-algebra from (6.13.3) and $L(P, \mathcal{F}^n)$ is the set of all $\mathcal{F}^n$-measurable functions $X \in L(P)$.

Proof: (6.14.1) $\Rightarrow$ (6.14.2): Suppose that X and Y are non-negative random variables such that X is $\mathcal{G}_{n+1}$-measurable and Y is $\mathcal{G}_{n+2}$ -measurable.

Since $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ and $\mathcal{G}_{j+1}$, and $\mathcal{F}_j$ are conditionally independent given $\mathcal{G}_j$ for all $j \geq 0$, then by (6.12) and (6.8) we find

$$\begin{aligned} E(XY \mid \mathcal{F}_n) &= E(E(XY \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) = E(XE(Y \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) \\ &= E(XE(Y \mid \mathcal{G}_{n+1}) \mid \mathcal{F}_n) = E(XE(Y \mid \mathcal{G}_{n+1}) \mid \mathcal{G}_n) \end{aligned}$$

Hence, we see that $E(XY \mid \mathcal{F}_n)$ is $\mathcal{G}_n$-measurable and so by (6.8.12), we have $E(XY \mid \mathcal{F}_n) = E(XY \mid \mathcal{G}_n)$. Continuing in this way we find that

$$E\left(\prod_{j=0}^{k} X_j \,\middle|\, \mathcal{F}_n\right) = E\left(\prod_{j=0}^{k} X_j \,\middle|\, \mathcal{G}_n\right)$$

whenever X_j is $\mathcal{G}_{n+j}$-measurable and non-negative for all $j = 0, 1, \ldots, n$. But then it follows easily from the functional Sierpinsky lemma (1.49) that we have $E(X \mid \mathcal{F}_n) = E(X \mid \mathcal{G}_n)$ whenever X is $\mathcal{F}^n$-measurable and non-negative. Thus, (6.14.2) follows by splitting X into its positive and negative part.

(6.14.2) $\Rightarrow$ (6.14.3) $\Rightarrow$ (6.14.4) $\Rightarrow$ (6.14.1): Follows easily from (6.12).□

6.15 Conditionally Independent Random Functions

Let S , T and U be random functions with values in the measurable spaces $(K, \mathcal{K})$, $(L, \mathcal{L})$ and $(M, \mathcal{M})$, respectively, and let $\mu = P_{(T,U)}$ and $\nu = P_U$ be the distributions laws of (T, U) and U , respectively. Then the following four statements are equivalent:

(6.15.1) *S and T are conditionally independent given U*

(6.15.2) *(S, U) and (T, U) are conditionally independent given U*

(6.15.3) *For all $A \in \mathcal{K}$ there exist a measurable function $q_A : M \to [0, 1]$ such that $P(S \in A \mid (T, U) = (t, x)) = q_A(x)$ μ-a.s.*

(6.15.4) $P(S \in A \,|\, (T, U) = (t, x)) = P(S \in A \,|\, U = x)$ *μ-a.s.* $\forall A \in \mathcal{A}$

Let $\varphi : K \times M \to R$ and $\psi : L \times M \to Q$ be measurable functions where $(R, \mathcal{R})$ and $(Q, \mathcal{Q})$ are measurable spaces, and let $f : K \to \overline{\mathbf{R}}$ and $g : L \to \overline{\mathbf{R}}$ be measurable functions such that $F := f(S)$, $G := g(T)$ and FG belong to $L(P)$. If S and T are conditionally independent given U , then we have the following:

(6.15.5) $E(FG \mid U = x) = E(F \mid U = x) \cdot E(G \mid U = x)$ *ν-a.s.*

(6.15.6) *If (S, U) is independent of T , then S and T are conditionally independent given U*

(6.15.7) *If S and T are conditionally independent given U, then so are $\varphi(S,U)$ and $\psi(T,U)$*

Proof: The proof proceeds in the same manner as the proof of (6.12), and I shall leave the details to the reader. □

6.16 Weak Markov Chains

Let $\{X_n \mid n \geq 0\}$ be a sequence of random functions with values in the measurable space $(M,\mathcal{B})$ and let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \ldots \subseteq \mathcal{F}$ be an increasing sequence of σ-algebras such that X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$. If $\mathcal{F}^n = \sigma(X_n, X_{n+1}, \ldots)$ and $L(P,\mathcal{F}^n)$ is the set of all $\mathcal{F}^n$-measurable functions $Y \in L(P)$, then the following six statements are equivalent:

(6.16.1) $\{X_n, \mathcal{F}_n \mid n \geq 0\}$ *is a weak Markov chain*

(6.16.2) X_{n+1} *and* $\mathcal{F}_n$ *are conditionally independent given* X_n $\forall n \geq 0$

(6.16.3) $\mathcal{F}^n$ *and* $\mathcal{F}_n$ *are conditionally independent given* X_n $\forall n \geq 0$

(6.16.4) $E(Y \mid \mathcal{F}_n) = E(Y \mid X_n) \quad \forall n \geq 0 \, \forall Y \in L(P,\mathcal{F}^n)$

(6.16.5) *There exist a function* $Q_n(B,t) : \mathcal{B} \times M \to [0,1]$ *satisfying*

$$P(X_{n+1} \in B \mid \mathcal{F}_n) = Q_n(B, X_n) \quad \forall n \geq 0 \, \forall B \in \mathcal{B}$$

(6.16.6) $P(X_{n+1} \in B \mid \mathcal{F}_n) = P(X_{n+1} \in B \mid X_n) \quad \forall n \geq 0 \, \forall B \in \mathcal{B}$

Proof: The proof proceeds in the same manner as the proof of (6.14), and I shall leave the details to the reader. □

6.17 The Permutation Invariant σ-Algebras

Let $(T,\mathcal{B})$ be a measurable space and let $d \in \mathbf{N} \cup \{\infty\}$. Then T^d denotes the *d-fold product* of T, that is,

$$T^d = \{(t_j)_{j<d+1} \mid t_j \in T \quad \forall 1 \leq j < d+1\}$$

with the usual convention: $\infty + 1 = \infty$. Let $1 \leq j < d+1$. Then we define *the projection*, denoted $\wp_j$, onto the jth coordinate and *the product σ-algebra*, denoted $\mathcal{B}^d$, as in (1.43), that is,

$$\wp_j(t_1, t_2, \ldots) = t_j \quad , \quad \mathcal{B}^d = \sigma(\wp_j \mid 1 \leq j < d+1)$$

If $n \geq 1$ is an integer, then we let $\mathcal{P}_n$ denote the set of all *permutations* of $\{1, \ldots, n\}$, that is, the set of all bijective functions from $\{1, \ldots, n\}$ onto itself. Let f be a function from T^d into the set S and let $1 \leq n < d+1$ be an integer. Then we say that f is *n-symmetric* if

$$(6.17.1) \qquad f(t) = f(t_{\sigma(1)}, \ldots, t_{\sigma(n)}, t_{n+1} \cdots) \quad \forall t = (t_1, t_2 \cdots) \in T^d \ \forall \sigma \in \mathcal{P}_n$$

and we say that f is *symmetric* if f is n-symmetric for all integers $1 \leq n < d+1$. Similarly, if $1 \leq n < d+1$ is an integer, then we say that f is *n-tame* if there exists a function $g : T^n \to S$ such that

$$(6.17.2) \qquad f(t) = g(t_1, \ldots, t_n) \qquad \forall t = (t_1, t_2, \ldots) \in T^d$$

and we say that f is *tame* if f is n-tame for some integer $1 \leq n < d+1$. Moreover, if f is n-tame and n-symmetric for some integer $1 \leq n < d+1$, then we say that f is an *n-tame symmetric function*. For instance, if $T = \mathbf{R}$, $d = \infty$ and $f(t) = t_1 + t_2 + t_3 + t_5$, then f is 5-tame and 1-, 2- and 3-symmetric, but f is not 4-symmetric nor is f 4-tame.

Let $1 \leq n \leq d$ be given. Then we define *the permutation invariant σ-algebra of order* n, denoted $\mathcal{S}^n$, as follows:

$$(6.17.3) \qquad \mathcal{S}^n = \{B \in \mathcal{B}^d \mid 1_B \text{ is } n\text{-symmetric}\}$$

[It is then easy to check that $\mathcal{S}^n$ is a sub-σ-algebra of $\mathcal{B}^d$.] Since every function is 1-symmetric and every n-symmetric function is k-symmetric for all $1 \leq k \leq n$, we have

$$(6.17.4) \qquad \mathcal{B}^d = \mathcal{S}^1 \supseteq \mathcal{S}^2 \supseteq \cdots \supseteq \mathcal{S}^d$$

$$(6.17.5) \qquad \mathcal{S}^\infty = \bigcap_{n=1}^{\infty} \mathcal{S}^n \qquad \text{if} \qquad d = \infty$$

Let $f : T^d \to \overline{\mathbf{R}}$ be a $\mathcal{B}^d$-measurable, n-symmetric function. Then we have that $\{f \leq a\} \in \mathcal{S}^n$ for all $a \in \mathbf{R}$ and so f is $\mathcal{S}^n$-measurable. Conversely, if f is $\mathcal{S}^n$-measurable, then $\{f = a\} \in \mathcal{S}^n$ for all $a \in \mathbf{R}$, but then it follows easily that f is n-symmetric and $\mathcal{B}^d$-measurable. Thus, we have

(6.17.6) $f : T^d \to \overline{\mathbf{R}}$ is $\mathcal{S}^n$-measurable if and only if f is $\mathcal{B}^d$-measurable and n-symmetric

Let $f : T^d \to \overline{\mathbf{R}}$ be a given function which assumes at most one of the values $\pm\infty$. If $1 \le n < d+1$, then we define *the n-symmetrization*, denoted $S_n f(t)$, of f as follows:

$$(6.17.7) \qquad S_n f(t) := \frac{1}{n!} \sum_{\sigma \in \mathcal{P}_n} f(t_{\sigma(1)}, \ldots, t_{\sigma(n)}, t_n, \ldots) \qquad \forall t = (t_i) \in T^d$$

Then $S_n f(t)$ is an n-symmetric function on T^d .

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X = (X_1, X_2, \ldots) : \Omega \to T^d$ be a d-dimensional random function with values in the product space $(T^d, \mathcal{B}^d)$. Then we define *the permutation invariant σ-algebra of order n based on* $X = (X_1, X_2, \ldots)$, denoted $\mathcal{S}_X^n$, as follows:

$$(6.17.8) \qquad \mathcal{S}_X^n := X^{-1}(\mathcal{S}^n) = \{X^{-1}(\mathcal{B}) \mid B \in \mathcal{S}^n\} \qquad \forall 1 \le n \le d$$

Then $\mathcal{S}_X^n$ is a σ-algebra on Ω, and by (6.17.4)–(6.17.6) and (6.4) we have

$$(6.17.9) \qquad \mathcal{F} \supseteq \sigma(X) = \mathcal{S}_X^1 \supseteq \mathcal{S}_X^2 \supseteq \cdots \supseteq \mathcal{S}_X^d$$

$$(6.17.10) \qquad \mathcal{S}_X^\infty = \bigcap_{n=1}^{\infty} \mathcal{S}_X^n \qquad \text{if} \qquad d = \infty$$

(6.17.11) $Y : \Omega \to \overline{\mathbf{R}}$ is $\mathcal{S}_X^n$-measurable if and only if $Y = f(X)$ for some $\mathcal{B}^d$-measurable, n-symmetric function $f : T^d \to \overline{\mathbf{R}}$

Let $\sigma \in \mathcal{P}_n$ be a permutation of $\{1, \ldots, n\}$ where $1 \le n < d+1$. Then we set

$$(6.17.12) \qquad X_\sigma = (X_{\sigma(1)}, \ldots, X_{\sigma(n)}, X_{n+1}, \ldots)$$

and we say that $(X_1, X_2, \ldots)$ are *exchangeable* if $X_\sigma \sim X$ for all $\sigma \in \mathcal{P}_n$ and all $1 \le n < d+1$.

6.18 U-Statistics

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X = (X_1, X_2, \cdots) : \Omega \to T^d$ be an exchangeable random function with values in the product space $(T^d, \mathcal{B}^d)$ where $d \in \mathbf{N} \cup \{\infty\}$. If f is a measurable function from $(T^d, \mathcal{B}^d)$ into $\overline{\mathbf{R}}$ such that $f(X) \in L(P)$, then we have

$$(6.18.1) \qquad E(f(X) \mid \mathcal{S}_X^n) = S_n f(X) \qquad \forall 1 \le n < d+1$$

where $S_n f$ *and* $\mathcal{S}_X^n$ *are given by (6.17.7) and (6.17.8). Let* $\mathcal{P}_{kn}$ *denote the set of all injective functions from* $\{1,\ldots,k\}$ *into* $\{1,\ldots,n\}$ *and suppose that* $f(t) = g(t_1,\ldots,t_k)$ *for some function* $g : T^k \to \mathbf{R}$ *where* $1 \le k < d+1$. *Then for every* $k \le n < d+1$ *and every* $t = (t_1, t_2, \ldots) \in T^d$ *we have*

$$S_n f(t) = \frac{(n-k)!}{n!} \sum_{\lambda \in \mathcal{P}_{kn}} g(t_{\lambda(1)},\ldots,t_{\lambda(k)}) \tag{6.18.2}$$

$$S_n f(t) = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \cdots < i_k \le n} g(t_{i_1},\ldots,t_{i_k}) \quad \textit{if } g \textit{ k-symmetric} \tag{6.18.3}$$

Remarks: (1): If $X_1, X_2, \ldots$ are independent and identically distributed, then, evidently, we have that $(X_1, X_2, \ldots)$ are exchangeable. Similarly, if $X_n = X_1$ for all $n = 1, 2, \ldots$, then $(X_1, X_2, \ldots)$ are exchangeable.

(2): Let $X = (X_1, X_2, \cdots) : \Omega \to T^d$ be an exchangeable, d-dimensional, random function with values in the product space $(T^d, \mathcal{B}^d)$ and let $f : T^d \to \overline{\mathbf{R}}$ be a measurable function. Then $S_n f(X)$ is called *the* n^{th} *U-statistics of* X *associated with* f where $S_n f$ is given by (6.17.7). Many of the usual *estimators* (also called *statistics*) are U-statistics. For instance:

The sample mean: Let $g : T \to \mathbf{R}$ be a measurable function and let $f(t_1, t_2, \ldots) = g(t_1)$. Then by (6.18.1) and (6.18.3) we get

$$S_n f(t_1, t_2, \ldots) = \tfrac{1}{n} \sum_{j=1}^{n} g(t_j) \qquad \forall 1 \le n < d+1 \tag{6.18.4}$$

$$S_n f(X) = \tfrac{1}{n} \sum_{j=1}^{n} g(X_j) = E(g(X_1) \mid \mathcal{S}_X^n) \qquad \forall 1 \le n < d+1 \tag{6.18.5}$$

which is called *the sample mean* of $g(X_1), \ldots, g(X_n)$.

The sample variance with known mean: Let $\mu \in \mathbf{R}$, let $g : T \to \mathbf{R}$ be a measurable function, and let $f(t_1, t_2, \ldots) = (g(t_1) - \mu)^2$. If $1 \le n < d+1$, then by (6.18.1) and (6.18.3) we get

$$S_n f(t_1, t_2, \ldots) = \tfrac{1}{n} \sum_{j=1}^{n} (g(t_j) - \mu)^2 \tag{6.18.6}$$

$$S_n f(X) = \tfrac{1}{n} \sum_{j=1}^{n} (g(X_j) - \mu)^2 = E\Big((g(X_1) - \mu)^2 \;\Big|\; \mathcal{S}_X^n\Big) \tag{6.18.7}$$

which is called *the sample variance of* $g(X_1), \ldots, g(X_n)$ *with known mean* μ .

The sample variance with unknown mean: Let $g : T \to \mathbf{R}$ be a measurable function and let $f(t_1, t_2, \cdots) = \frac{1}{2}(g(t_1) - g(t_2))^2$. If $2 \le n < d+1$ and

$\overline{g}_n(t) := \frac{1}{n}(g(t_1) + \cdots + g(t_n))$ and $\overline{G}_n := \overline{g}_n(X)$ are the sample means, then by (6.18.1) and (6.18.3) we get

$$(6.18.8) \qquad S_n f(t_1, t_2, \ldots) = \frac{1}{n-1} \sum_{j=1}^{n} (g(t_j) - \overline{g}_n(t))^2$$

$$(6.18.9) \qquad S_n f(X) = \frac{1}{n-1} \sum_{j=1}^{n} \left(g(X_j) - \overline{G}_n\right)^2 = \tfrac{1}{2} E\Big((g(X_1) - g(X_2))^2 \;\Big|\; \mathcal{S}_X^n\Big)$$

which is called *the sample variance of* $g(X_1), \ldots, g(X_n)$ *with unknown mean.*

The general sample covariance: Let $g : T \times T \to \mathbf{R}$ be a measurable function and let $f(t_1, t_2, \cdots) = g(t_1, t_2)$. If $2 \le n < d+1$ and $G_{ij} = g(X_i, X_j)$, then by (6.18.1) and (6.18.3) we get

$$(6.18.10) \qquad S_n f(t_1, t_2, \ldots) = \frac{1}{n(n-1)} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} g(t_i, t_j) - \sum_{k=1}^{n} g(t_k, t_k) \right\}$$

$$(6.18.11) \qquad S_n f(X) = \frac{1}{n(n-1)} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ij} - \sum_{k=1}^{n} G_{kk} \right\} = E(G_{12} \mid \mathcal{S}_X^n)$$

which is called *the general sample covariance of* $\{g(X_i, X_j)\}$.

Proof: Let $B \in \mathcal{S}^n$ be given. If $\sigma \in \mathcal{P}_n$, then by n-symmetry of 1_B and exchangeability of X we have $(f(X), 1_B(X)) \sim (f(X_\sigma), 1_B(X_\sigma))$ and $1_B(X_\sigma) = 1_B(X)$. Hence, we find

$$\int_{\{X \in B\}} f(X)\, dP = \int_{\{X \in B\}} f(X_\sigma)\, dP \qquad \forall \sigma \in \mathcal{P}_n$$

Summing over $\sigma \in \mathcal{P}_n$ and dividing by $n! = \#(\mathcal{P}_n)$, we get

$$\int_{\{X \in B\}} f(X)\, dP = \int_{\{X \in B\}} S_n f(X)\, dP \qquad \forall B \in \mathcal{S}^n$$

and since $S_n f(X)$ is $\mathcal{S}_X^n$-measurable, we see that (6.18.1) follows from definition (6.7.1+2). □

6.19 Conditional Partial Sums

Let $X = (X_1, X_2, \ldots) : \Omega \to \mathbf{R}$ be a d-dimensional, exchangeable random vector where $1 \le d \le \infty$. Let $\mathcal{S}_X^n$ be the permutation invariant σ-algebra [see (6.17.8)] and set $S_n = X_1 + \cdots + X_n$ and

$$(6.19.1) \qquad \mathcal{T}_X^k = \sigma\{X_j \mid j \ge k+1\} \quad , \quad \mathcal{T}_S^n = \sigma\{S_j \mid j \ge n\}$$

whenever $0 \le k < d$ and $1 \le n < d+1$. If $d < \infty$ we set $\mathcal{T}_X^d = \{\emptyset, \Omega\}$; if $d = \infty$ we set

$$(6.19.2) \qquad \mathcal{T}_X^\infty = \bigcap_{n=0}^{\infty} \mathcal{T}^n \quad , \quad \mathcal{T}_S^\infty = \bigcap_{n=1}^{\infty} \mathcal{T}_S^n$$

Since $f(t) = t_1 + \cdots + t_j$ is n-symmetric for all $n \le j$ we have that $\mathcal{T}_S^n \subseteq \mathcal{S}_X^n$; and since $X_j = S_j - S_{j-1}$, we have that $\mathcal{T}_X^n \subseteq \mathcal{T}_S^n$. Thus, we find

$$(6.19.3) \qquad \mathcal{T}_X^n \subseteq \mathcal{T}_S^n \subseteq \mathcal{S}_X^n \qquad\qquad \forall 1 \le n \le d$$

By definition, we have

$$(6.19.4) \qquad \sigma(X) = \mathcal{T}_X^0 \supseteq \mathcal{T}_X^1 \supseteq \cdots \supseteq \mathcal{T}_X^d$$

$$(6.19.5) \qquad \sigma(X) = \mathcal{T}_S^1 \supseteq \mathcal{T}_S^2 \supseteq \cdots \supseteq \mathcal{T}_S^d$$

The σ-algebra $\mathcal{T}_X^\infty$ is called *the tail σ-algebra* of $(X_1, X_2 \cdots)$ and $\mathcal{T}_S^\infty$ is called *the tail σ-algebra* of $(S_1, S_2, \cdots)$. If $d = \infty$, by (6.19.3) we have that $\mathcal{T}_X^\infty \subseteq \mathcal{T}_S^\infty \subseteq \mathcal{S}_X^\infty$. In Section (7.27), we shall see that these three σ-algebras essentially coincide. Suppose that $X_1 \in L(P)$. Since $\sigma(S_n) \subseteq \mathcal{T}_S^n \subseteq \mathcal{S}_X^n$, then by (6.18.5) and (6.8.9) we get

$$(6.19.6) \qquad \frac{S_n}{n} = E(X_1 \mid \mathcal{S}_X^n) = E(X_1 \mid \mathcal{T}_S^n) = E(X_1 \mid S_n)$$

Thus, if we only know the sum S_n , then the best estimate of X_1 is $\frac{S_n}{n}$, and knowledge of the later sums, and even knowledge of every n-symmetric function of X , does not give any further information about X_1 .

6.20 The Distribution of the Order Statistics

Let $X = (X_1, \ldots, X_n)$ be a random vector satisfying $P(X_i = X_j) = 0$ for all $i \ne j$ and let $R = (X_{1n}, \ldots, X_{nn})$ be the order statistics of X [see (2.23)]. Set

$$\Delta := \{ t \in \mathbf{R}^n \mid t_1 < \cdots < t_n \} \quad , \quad X_\sigma = \big(X_{\sigma(1)}, \ldots, X_{\sigma(n)}\big) \qquad \forall \sigma \in \mathcal{P}_n$$

Then we have

$$(6.20.1) \qquad P(R \in A) = \sum_{\sigma \in \mathcal{P}_n} P(X_\sigma \in A \cap \Delta) \qquad \forall A \in \mathcal{B}(\mathbf{R}^n)$$

In particular, if $(X_1, \ldots, X_n)$ *are exchangeable, then*

$$P(R \in A) = n!\, P(X \in A \cap \Delta) \qquad \forall A \in \mathcal{B}(\mathbf{R}^n) \tag{6.20.2}$$

If $(X_1, \ldots, X_n)$ *is absolutely continuous with density function* $f(t_1, \ldots, t_n)$ *, then* R *is absolutely continuous with density* g *given by*

$$g(t_1, \ldots, t_n) = \begin{cases} \sum_{\sigma \in \mathcal{P}_n} f\big(t_{\sigma(1)}, \ldots, t_{\sigma(n)}\big) & \text{if } (t_1, \ldots, t_n) \in \Delta \\ 0 & \text{if } (t_1, \ldots, t_n) \notin \Delta \end{cases} \tag{6.20.3}$$

In particular, if $(X_1, \ldots, X_n)$ *are exchangeable and absolutely continuous with density* f *, then* R *is absolutely continuous with density* g *given by*

$$g(t_1, \ldots, t_n) = \begin{cases} n!\, f\big(t_1, \ldots, t_n\big) & \text{if } (t_1, \ldots, t_n) \in \Delta \\ 0 & \text{if } (t_1, \ldots, t_n) \notin \Delta \end{cases} \tag{6.20.4}$$

Remark: Note that if either X is absolutely continuous, or if $X_1, \ldots, X_n$ are independent and diffuse, then $P(X_i = X_j) = 0$ for all $i \neq j$.

Proof: Let $D(\sigma) = \{X_\sigma \in \Delta\}$ for all $\sigma \in \mathcal{P}_n$. Then $D(\sigma') \cap D(\sigma'') = \emptyset$ whenever $\sigma' \neq \sigma''$ and by assumption, we have

$$P\left(\Omega \setminus \bigcup_{\sigma \in \mathcal{P}_n} D(\sigma)\right) = P\left(\bigcup_{1 \le i < j \le n} \{X_i = X_j\}\right) = 0$$

Moreover, since $R(\omega) = X_\sigma(\omega)$ for all $\omega \in D(\sigma)$ and $\#(\mathcal{P}_n) = n!$, we see that (6.20.1) and (6.20.2) hold. Suppose that X is absolutely continuous with density f . Then X_σ has density $f(t_{\tau(1)}, \ldots, t_{\tau(n)})$ where $\tau = \sigma^{-1}$ is the inverse of σ . Hence, (6.20.3) follows from (6.20.1). If, moreover, X is exchangeable, then f is n-symmetric and so (6.20.4) follows from (6.20.3). □

6.21 Conditional Maximal Inequalities

Let $U_1, \ldots, U_n$ *be random variables and let* $M_n = \max\{U_1, \ldots, U_n\}$. *Let* $r \le q$ *be given numbers, and set*

$$F_1 = \{U_1 > r\} \quad , \quad F_j = \{U_1 \le r, \ldots, U_{j-1} \le r, U_j > r\}$$
$$G_1 = \{U_1 > t\} \quad , \quad G_j = \{U_1 \le q, \ldots, U_{j-1} \le q, U_j > q\}$$

for $j = 2, 3, \ldots, n$. *Suppose that* $V, W \in L(P)$ *are random variables and that* $R_1, \ldots, R_n \in L^1(P)$ *are non-negative random variables satisfying*

$$\int_{F_i} R_i \, dP \le \int_{F_i} V \, dP \qquad \forall 1 \le i \le n \tag{6.21.1}$$

$$\int_{F_i \cap G_j} R_i \, dP \ge \lambda \int_{F_i \cap G_j} W \, dP \qquad \forall 1 \le i \le j \le n \tag{6.21.2}$$

for some number $\lambda \ge 0$. *Then we have*

$$\lambda P(M_n > q) \le \int_{\{M_n > r\}} V \, dP + \lambda \int_{\{M_n > q\}} (1 - W) \, dP \tag{6.21.3}$$

Proof: Note that $F_1 \ldots F_n$ are disjoint sets with union $\{M_n > r\}$ and that $G_1 \ldots G_n$ are disjoint sets with union $\{M_n > q\}$. Let $\hat{R} = R_1 1_{F_1} + \cdots R_n 1_{F_n}$. Then $\hat{R}$ is a non-negative integrable random variable and by (6.21.1), we get

$$\begin{aligned} \int_{\{M_n > q\}} \hat{R} \, dP &\le \sum_{i=1}^{n} \int_{F_i} R_i \, dP \le \sum_{i=1}^{n} \int_{F_i} V \, dP \\ &= \int_{\{M_n > r\}} V \, dP \end{aligned}$$

Moreover, since $r \le q$, we have that $F_i \cap G_j = \emptyset$ for all $1 \le j < i \le n$. Hence, by (6.21.2) we find

$$\begin{aligned} \int_{\{M_n > q\}} \hat{R} \, dP &\ge \sum_{i=1}^{n} \sum_{j=i}^{n} \int_{F_i \cap G_j} R_i \, dP \ge \lambda \sum_{i=1}^{n} \sum_{j=i}^{n} \int_{F_i \cap G_j} W \, dP \\ &= \lambda \int_{\{M_n > q\}} W \, dP \end{aligned}$$

Thus, we get

$$\begin{aligned} \lambda P(M_n > q) &= \int_{\{M_n > q\}} \hat{R} \, dP + \int_{\{M_n > q\}} (\lambda - \hat{R}) \, dP \\ &\le \int_{\{M_n > r\}} V \, dP + \lambda \int_{\{M_n > q\}} (1 - W) \, dP \end{aligned}$$

which proves (6.21.3). □

Remark: Inequality (6.21.3) belongs to the family of maximal inequalities [see (2.17+18)]. Setting $r = q = x$, $\lambda = \min_{1 \le j \le n} P(V_j \le y)$, $R_i = 1_{\{V_i \le y\}}$, $V = 1_{\{L_n > x - y\}}$ and $W \equiv 1$, it is easily verified that (6.21.1+2) hold and that (2.18.2) follows from (6.21.3). In spite of the simple proof, (6.21.3) contains a series of classical non-trivial maximal inequalities as special cases which you will find below.

6.22 The Martingale Inequality

Let $U_1, \dots U_n$ *be random variables and suppose that* $U \in L(P)$ *satisfies*

$$U_j \le E(U \mid \mathcal{F}_j) \qquad \forall 1 \le j \le n \tag{6.22.1}$$

where $\mathcal{F}_j$ *is a* σ*-algebra containing* $\sigma(U_1, \dots U_j)$. *If* $M_n = \max\{U_1, \dots U_n\}$ *and* $\varphi : \mathbf{R} \to [0, \infty[$ *is an increasing, non-negative, convex function, then we have*

$$\varphi(t) P(M_n > t) \le \int_{\{M_n > t\}} \varphi(U) dP \qquad \forall t \in \mathbf{R} \tag{6.22.2}$$

Proof: Set $r = q = t$, $R_i = \varphi(U_i)$, $V = \varphi(U)$, $W \equiv 1$ and $\lambda = \varphi(t)$. Then (6.21.1) follows from Jensen's inequality (6.9.3) and (6.21.2) follows from increasingness of φ . Thus, (6.22.2) follows from (6.21.3) □

6.23 Kolmogorov's Inequality

Suppose that $X_1, \dots X_n$ *are independent random variables with finite means* $\mu_j = EX_j$ *and finite variances* $\sigma_j^2 = \operatorname{var} X_j$. *Then we have*

$$P\left(\max_{1 \le j \le n} \left| \sum_{k=1}^{j} (X_j - \mu_j) \right| > t \right) \le t^{-2} \sum_{j=1}^{n} \sigma_j^2 \qquad \forall t > 0 \tag{6.23.1}$$

Note that if $n = 1$, *then (6.23.1) reduces to Chebyshev's inequality.*

Proof: Set $S_j = (X_1 - \mu_1) + \cdots + (X_j - \mu_j)$ and $\mathcal{F}_j = \sigma(X_1, \dots, X_j)$. Then by (6.8.3) and (6.8.5), we have $S_j = E(S_n \mid \mathcal{F}_j)$. So by Jensen's inequality (6.9.3) we have $S_j^2 \le E(S_n^2 \mid \mathcal{F}_j)$. Hence, (6.23.1) follows from (6.22.2) with $U_j = S_j^2$ and $U = S_n^2$. □

6.24 Ottaviani's Inequality

Suppose that $X_1, \dots X_n$ *are independent random variables and let* $M_n = \max\{S_1, \dots S_n\}$ *where* $S_j = X_1 +, \dots + X_j$. *Then we have*

$$P(M_n > s + t) \min_{1 \le j \le n} P(|S_n - S_j| \le s) \le P(|S_n| > t) \quad \forall s, t \ge 0 \tag{6.24.1}$$

Proof: Set $r = q = s + t$, $R_i = 1_{\{|S_n - S_i| \le s\}}$, $V = 1_{\{|S_n| > t\}}$, $W \equiv 1$, $U_j = |S_j|$ and $\lambda = \min_{1 \le j \le n} P(|S_n - S_j| \le s)$. Then it is easily checked that (6.21.1+2) holds. Thus, (6.24.1) follows from (6.21.3). □

6.25 Lévy's Inequality

Let $X_1, \dots X_n$ be real random variables such that for all $1 \le j < n$ we have $(X_1, \dots X_n) \sim (X_1, \dots X_j, -X_{j+1}, \dots - X_n)$. Then

$$P\Big(\max_{1 \le j \le n} |S_j| > t\Big) \le 2P(|S_n| > t) \qquad \forall t \ge 0 \tag{6.25.1}$$

Proof: Set $r = q = t$, $U_j = |S_j|$, , $\lambda = 1$, $R_j = 1_{\{|2S_j - S_n| > t\}}$, $V = 1_{\{|S_n| > t\}}$ and $W = 1_{\{|S_n| \le t\}}$. Then it is easily checked that (6.21.1+2) hold. Thus, (6.25.1) follows from (6.21.3). □

6.26 The Exponential Inequality

Let $X_1, \dots X_n$ be independent random variables. Set $S_n = X_1 + \cdots X_n$ and

$$M_n = \max\{|S_1|, \dots, |S_n|\} \quad , \quad L_n = \max\{|X_1|, \dots, |X_n|\}$$
$$\Upsilon(x) = \max_{1 \le j \le n} P(|S_n - S_j| > x)$$

Then for all $s, t, u, v \ge 0$ we have

$$P(M_n > s + t + u + v) \le P(L_n > s) + \frac{\Upsilon(v) P(M_n > u)}{1 - \Upsilon(t)} \tag{6.26.1}$$

$$\Upsilon(s) P(M_n \le t) \le P(S_n > s + t) \tag{6.26.2}$$

$$\Upsilon(s + t) \le P(|S_n| > s) + \max_{1 \le j \le n} P(|S_j| > t) \tag{6.26.3}$$

$$P(|S_n| > s) \le \max_{1 \le j \le n} P(|S_j| > s) \le P(M_n > s) \tag{6.26.4}$$

Proof: Set $r = u$, $q = s + t + u + v$, $R_i = 1_{\{|S_n - S_i| > v\}}$, $U_j = |S_j|$, $V \equiv \Upsilon(v)$ and $W = 1_{\{L_n > s\}}$ Then it is easily checked that (6.21.1+2) hold. Thus, (6.26.1) follows from (6.21.3). Moreover, (6.26.2)–(6.26.4) follow directly from the definition of $\Upsilon(s)$ and M_n . □

6.27 Fish Auction

Andrew is a fishmonger and every morning he goes to the fish auction to buy the fish that he is going to sell in his shop that day. On the day when we are meeting him, he has decided to buy a box of cod, and on that day there are q boxes of cod for sale at the auction. At a modern fish auction there is a big "clock" on the wall; when a box of fish is going on sale, the clock is started at a certain maximum price and moves backward. All the participants have a button which can stop the clock, and the one who stops the clock first gets the fish at the price showing on the clock. Thus, the longer you wait, the cheaper you get the fish, but somebody else can stop the clock, and then the fish is sold and gone. Fishmonger Andrew wants to find that strategy which maximizes the probability that he gets the cheapest box of cod sold on the auction. In order to find the optimal strategy he has to find the following probability:

Suppose that k boxes have been sold $(1 \leq k \leq q)$. Then given the information available at this time, what is the probability that the kth box will be the cheapest among all q boxes? If Andrew has been inattentive, then he would probably assign the probability $\frac{1}{q}$ to this event (there are q boxes in all, and each of them has presumably the same chance of being the cheapest). But Andrew has not been inattentive, he has followed the bids for the k first boxes closely. Thus, if the kth box is not the cheapest among the k first boxes; he knows that the probability is 0. So suppose that the kth box is the cheapest among the k first boxes; then he may argue like this: There is $q-k$ boxes left and each of these together with the kth box has the possibility of being the cheapest, presumably each of these $q-k+1$ boxes has the same probability of being the cheapest. Thus, the probability must be $\frac{1}{(q-k+1)}$. The argument is appealing, but remember that I have warned you to rely heavily on your intuition when you evaluate conditional probabilities. So let us make a model and compute the probability.

Let us rank the boxes such that the cheapest box gets rank 1, the second cheapest rank 2, etc. Let X_j be the rank of the jth box for $j = 1, 2, \ldots, q$. Thus, if the seventh box is the second cheapest, then $X_7(\omega) = 2$. Then $(X_1, \ldots, X_q)$ is a random permutation of $\{1, \ldots, q\}$, and our a priori hypothesis is that all permutations are equally probable; that is,

$$P(X = \sigma) = \frac{1}{q!} \qquad \forall \sigma \in \mathcal{P}_q \tag{6.27.1}$$

where $X = (X_1, \ldots, X_q)$, and $\mathcal{P}_q$, as usual, is the set of all permutations of $\{1, \ldots, q\}$. Of course, we cannot observe the X_j's before the auction is completed. But right after the jth box has been sold, we can observe the random variable Y_j given by

$$Y_j = \text{the rank of the } j^{\text{th}} \text{ box among the first } j \text{ boxes} \tag{6.27.2}$$

Note that $Y_j = f_j(X)$ where $f_j : \mathcal{P}_q \to \{1, \ldots, j\}$ is given by

(6.27.3) $$f_j(\sigma) = \#\{1 \le k \le j \mid \sigma(k) \le \sigma(j)\} \qquad \forall \sigma \in \mathcal{P}_q \; \forall 1 \le j \le q$$

Let $Y = (Y_1, \ldots, Y_q)$ and $f(\sigma) = (f_1(\sigma), \ldots, f_q(\sigma))$. Then $Y = f(X)$ and f maps $\mathcal{P}_q$ into $\mathcal{L}_q$ where

(6.27.4) $$\mathcal{L}_q = \prod_{j=1}^{q} \{1, \ldots, j\} = \{\beta \in \mathbf{N}^q \mid 1 \le \beta_j \le j \quad \forall j = 1, \ldots, q\}$$

Suppose that $\sigma, \tau \in \mathcal{P}_q$ and $\sigma \ne \tau$. Then there exists $j \le q$ so that $\sigma(j) \ne \tau(j)$ and $\sigma(i) = \tau(i)$ for $i = j+1, \ldots, q$. Hence, the two sets of numbers $\{\sigma(1), \ldots, \sigma(j)\}$ and $\{\tau(1), \ldots, \tau(j)\}$ coincide, but possibly in a different order. Since $\sigma(j) \ne \tau(j)$, then by definition of f_j we have $f_j(\sigma) \ne f_j(\tau)$. Thus, $f(\sigma) \ne f(\tau)$ whenever $\sigma \ne \tau$; that is, f is an injective function from $\mathcal{P}_q$ into $\mathcal{L}_q$. And since the sets $\mathcal{P}_q$ and $\mathcal{L}_q$ have the same number elements, namely $q!$, we conclude from the *pigeonhole principle* that f is bijective. (The *pigeonhole principle* says that if n pigeons are occupying n pigeonholes such that there is at most 1 pigeon in each hole, then there is exactly 1 pigeon in each pigeonhole.) Hence, if $\beta = (\beta_1, \ldots, \beta_q) \in \mathcal{L}_q$, then by (6.27.1) we see that

$$\begin{aligned} P(Y_1 = \beta_1, \ldots, Y_q = \beta_q) &= P(f(X) = \beta) = P(X = f^{-1}(\beta)) \\ &= \tfrac{1}{q} = \prod_{j=1}^{q} \tfrac{1}{j} \end{aligned}$$

Thus, using (4.5.12) with $S_j = \{1, \ldots, j\}$ and $\mu_j = \#$, we see that $Y_1, \ldots, Y_q$ are independent and $P(Y_j = k) = \frac{1}{j}$ for all $k = 1, \ldots, j$. Hence, we have proved the following proposition:

(6.27.5) $Y_1, \ldots, Y_q$ are independent and $P(Y_j = k) = \frac{1}{j} \quad \forall 1 \le k \le j$

Right after the kth box has been sold, we know the values of the random variables $Y_1, \ldots, Y_k$, so our information at this time is described by the σ-algebra

$$\mathcal{F}_k = \sigma\{Y_1, \ldots, Y_k\} \qquad k = 1, 2, \ldots, q$$

Hence, we are chasing the conditional probability: $P(X_k = 1 \mid \mathcal{F}_k)$. So let us compute this probability. To do this, let us set $A = \{Y_k = 1\}$ and $B = \{Y_{k+1} > 1, \ldots, Y_q > 1\}$. Then we have

$$\{X_k = 1\} = \{Y_k = 1, Y_{k+1} > 1, \ldots, Y_q > 1\} = A \cap B$$

and since $A \in \mathcal{F}_k$ and B is independent of $\mathcal{F}_k$, then by (6.8.2) and (6.8.3) we get

$$\begin{aligned} P(X_k = 1 \mid \mathcal{F}_k) &= E(1_{A\cap B} \mid \mathcal{F}_k) = E(1_A 1_B \mid \mathcal{F}_k) \\ &= 1_A E(1_B \mid \mathcal{F}_k) = 1_A P(B) \end{aligned}$$

Moreover, since $B = \{Y_{k+1} > r\} \cap \{Y_{k+2} > 1\} \cap \cdots \cap \{Y_q > 1\}$, then by (6.27.5) we find:

$$\begin{aligned} P(B) &= \prod_{j=k+1}^{q} P(Y_j > 1) = \prod_{j=k+1}^{q} (1 - P(Y_j = 1)) \\ &= \prod_{j=k+1}^{q} \left(1 - \frac{1}{j}\right) = \frac{k}{k+1} \cdot \frac{k+1}{k+2} \cdots \frac{q-1}{q} = \frac{k}{q} \end{aligned}$$

Thus, we have

$$P(X_k = 1 \mid \mathcal{F}_k) = \tfrac{k}{q}\, 1_{\{Y_k=1\}} = \begin{cases} \frac{k}{q} & \text{if } Y_k = 1 \\ 0 & \text{if } Y_k \neq 1 \end{cases} \tag{6.27.6}$$

which is in accordance with our intuitive argument when $Y_k \neq 1$ (that is, when the kth box is not the cheapest among the k first boxes), but inconsistent with our intuitive argument when $Y_k = 1$. The computation above is correct, so here we have yet another example of our poor intuition concerning conditional probabilities. You may try to find an intuitive justification for (6.27.6). I haven't found any which looks convincing to me.

I shall leave fishmonger Andrew here without solving his problem about the most favorable strategy. In Volume II of this book I shall return to him and solve his problem [see (11.20)]. However, the solution requires heavier artillery which I will build up in Chapter 11.

6.28 References to Chapter 6

Propositions (6.5.1)–(6.5.9) can be found in [3], [4] and Reference [4] to Chapter 1, where you will find a lot more about analytic spaces and Blackwell σ-algebras. In particular, the atomic determination theorem 6.6 is due to Falkner (see Proposition 1.13 in [3]). Conditional expectation (given a real-valued random variable) was introduced by Kolmogorov (see Reference [3] to Chapter 1). The notion was extended to arbitrary σ-algebras by Doob in his classical book [2]. Markov chains

is one of the great subjects of probability theory and there is a huge number of textbooks on Markov chains (see, for instance, [1], [5] and [6]).

[1] Chung, K. L. (1967): *Markov Chains With Stationary Transition Probabilities*, (second ed.) Springer Verlag, Berlin and New York

[2] Doob, J. L. (1953): *Stochastic Processes*, John Wiley & Sons, Inc., New York and London

[3] Falkner, N. (1981): *Generalizations of Standard and Analytic Measurable Spaces*, Math. Scand. 49, pp. 283–301

[4] Hoffmann-Jørgensen, J. (1970): *The Theory of Analytic Sets*, Mat. Inst., Aarhus University, Var. Publ. Ser. 10, pp. 1–314

[5] Kemney, J. G., Snell, J. L. and Knapp, A. W. (1966): *Denumerable Markov Chains*, D. Van Nostrand Co., Inc., New York and London

[6] Revuz, D. (1975): *Markov Chains*, North-Holland Publ. Co., Amsterdam; and American Elsevier Publ. Co., Inc., New York

6.29 Exercises to Chapter 6

Exc. 6.1: Let $X_1, X_2, \ldots$ be independent random variables with finite means $\mu_1, \mu_2, \ldots$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and show that

(i) $$E\Big(\sum_{j=1}^{k} X_j \;\Big|\; \mathcal{F}_n\Big) = \sum_{j=1}^{n} X_j + \sum_{j=n+1}^{k} \mu_j \qquad \forall k > n$$

(ii) $$E\Big(\prod_{j=1}^{k} X_j \;\Big|\; \mathcal{F}_n\Big) = \Big(\prod_{j=1}^{n} X_j\Big)\Big(\prod_{j=n+1}^{k} \mu_j\Big) \qquad \forall k > n$$

Suppose that the variances $\sigma_n^2 = \mathrm{var}(X_n)$ exist and are finite for all $n \geq 1$. Show that

(iii) $$E\Big\{\Big(\sum_{j=1}^{k} (X_j - \mu_j)\Big)^2 \;\Big|\; \mathcal{F}_n\Big\} = \Big(\sum_{j=1}^{n} (X_j - \mu_j)\Big)^2 + \sum_{j=n+1}^{k} \sigma_j^2 \qquad \forall k > n$$

Exc. 6.2: Let (X, Y) be a 2-dimensional random vector with distribution functions $F(x, y)$. Let $H(x|y) := P(X \leq x \mid Y = y)$ be *the conditional distribution function* of X given Y . Show that

(i) $$F(x, y) = \int_{-\infty}^{y} H(x|u)\, F_Y(du) \qquad \forall (x, y) \in \mathbf{R}^2$$

where F_Y is the marginal distribution function of Y .

Exc. 6.3: The conditional distribution of X given $Y = t$ is the binomial distribution $Bi(n,t)$ where Y has a beta distribution $B(\alpha,\beta)$ and $n \in \mathbb{N}$ and $\alpha,\beta > 0$ are given numbers. Find the distribution of (X,Y) and the conditional distribution of Y given X . [Apply (6.11.4).]

Exc. 6.4: The conditional distribution of X given $Y = t$ is the exponential distribution $E(t)$ where Y has a gamma distribution $\Gamma(\alpha,\beta)$ and $\alpha,\beta > 0$ are given numbers. Find the distribution of (X,Y) and the conditional distribution of Y given X . [Apply (6.11.4).]

Exc. 6.5: The conditional distribution of X given $Y = n$ is the binomial distribution $Bi(n,p)$ where Y has a Poisson distribution $Po(\lambda)$ and $p \in]0,1[$ and $\lambda > 0$ are given numbers. Find the distribution of (X,Y) and the conditional distribution of Y given X . [Apply (6.11.4).]

Exc. 6.6: Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with common distribution function F . Let $S_n = X_1 + \cdots + X_n$ and $1 \le n < k$. Find the conditional distribution of S_k given $S_n = x$, when $F = Po(\lambda)$, $F = E(\lambda)$ or $F = Bi^-(n,\kappa)$.

Exc. 6.7: Cindy and Arthur have a date between 5 P.M. and 6 P.M. They have agreed not wait more than 10 minutes for each other. Arthur arrives X minutes past 5 and Cindy arrives Y minutes past 5 where X and Y are independent and uniformly distributed in the interval $[0,60]$. Let A be the event that Cindy and Arthur actually meet. Show that

$$\text{(i)} \qquad P(A \mid X = t) = \begin{cases} \frac{10+t}{60} & \text{if} \quad 0 \le t \le 10 \\ \frac{1}{3} & \text{if} \quad 10 \le t \le 50 \\ \frac{70-t}{60} & \text{if} \quad 50 \le t \le 60 \end{cases}$$

Deduce that $P(A) = \frac{11}{36}$.

Exc. 6.8: Let $S = N_1 + \cdots + N_k$ where $N_1, \ldots, N_k$ be independent random variables with $N_j \sim Po(\lambda_j)$ for $j = 1, \ldots, k$. Show that the conditional distribution of $(N_1, \ldots, N_{k-1})$ given S is the multinomial distribution $Mu_{k-1}(n, (p_1, \ldots, p_{k-1}))$ [see Exc.2.46] where

$$p_j = \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_k} \qquad \forall j = 1, \ldots, k-1$$

Exc. 6.9: Let $\tau, X_1, X_2, \ldots$ be independent random variables where τ is non-negative and integer-valued. Let $S_\tau(\omega) = S_{\tau(\omega)}(\omega)$ where $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$. Show that

(i) $\quad P(S_\tau \in A \mid \tau = n) = P(S_n \in A) \quad \forall A \in \mathcal{B}(\mathbf{R})\ \forall n = 0, 1, 2, \ldots$

Let $q(t)$ be the generating function for τ and let $\psi(t)$ be the characteristic function for S_τ . Suppose that $X_1, X_2, \ldots$ are identically distributed with common characteristic function $\varphi(t)$. Show the following:

(ii) $\quad \psi(t) = q(\varphi(t)) \quad \forall t \in \mathbf{R}$

(iii) $\quad \tau \sim Po(\lambda) \quad \Rightarrow \quad \psi(t) = e^{\lambda(\varphi(t)-1)} \quad \forall t \in \mathbf{R}$

(iv) If τ is infinitely divisible and $P(\tau = 0) > 0$, then S_τ is infinitely divisible

Hint: (iv): Let $n \geq 2$ be given and let $\tau_1, \ldots, \tau_n$ be independent identically distributed random variables such that $\tau = \tau_1 + \cdots + \tau_n$. Show that

$$P(\tau_1 < 0)^n \leq P(\tau < 0) = 0 \quad , \quad P(\tau_1 = 0)^n = P(\tau = 0) > 0$$

$$P(\tau_1 \notin \mathbf{N}_0) \cdot P(\tau_1 = 0)^{n-1} \leq P(\tau \notin \mathbf{N}_0) = 0$$

where $\mathbf{N}_0 = \{0, 1, 2, \ldots\}$ is the set of all non-negative integers. Deduce that τ_1 is non-negative and integer-valued and that $q(t) = q_1(t)^n$ where q_1 is the generating function of τ_1 . Let $\xi(t)$ be characteristic function of S_{τ_1} and show that $\psi(t) = \xi(t)^n$.

Exc. 6.10: Let $\mathcal{C}$ be the σ-algebra on $\mathbf{R}$ generated by all finite subsets of $\mathbf{R}$, that is, $C \in \mathcal{C}$ if and only if either C is countable or the complement $\mathbf{R} \setminus C$ is countable [see (1.12)]. Show that $\mathcal{C} \subseteq \mathcal{B}(\mathbf{R})$ and that $\mathcal{C}$ separates points in $\mathbf{R}$. Show that $[0, 1] \in \mathcal{B}(\mathbf{R}) \setminus \mathcal{C}$. Why doesn't this contradicts (6.6)?

Exc. 6.11: Let T be a set and let $\mathbf{R}^T$ be the set of all functions from T into $\mathbf{R}$. Let $\wp_t(f) = f(t)$ for $f \in \mathbf{R}^T$ and $t \in T$ and let $\mathcal{F}^A = \sigma(\wp_t \mid t \in A)$ whenever $A \subseteq T$. Show that

(i) $\quad f \equiv g \left(\operatorname{mod} \mathcal{F}^A\right) \quad \Leftrightarrow \quad f(t) = g(t) \quad \forall t \in A$

(ii) $\quad \mathcal{F}^A(f) = \left\{ g \in \mathbf{R}^T \;\middle|\; g(t) = f(t) \quad \forall t \in A \right\}$

Exc. 6.12: Let $X_1, X_2, \ldots$ be functions from Ω into a measurable space $(M, \mathcal{B})$. Suppose that $\mathcal{B}$ separates points in M and set

$$\mathcal{F}_n = \sigma(X_j \mid j \geq n) \quad , \quad \mathcal{F}_\infty = \bigcap_{n=1}^{\infty} \mathcal{F}_n$$

Show that [apply (6.2.5+6)]

(i) $\omega' \equiv \omega'' \,(\mathrm{mod}\, \mathcal{F}_\infty) \quad \Leftrightarrow \quad \exists\, n \geq 1$ so that $X_j(\omega') = X_j(\omega'') \quad \forall\, j \geq n$

(ii) $\mathcal{F}_\infty(\omega_0) = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} \{\omega \mid X_j(\omega) = X_j(\omega_0)\}$

Exc. 6.13: Let X be a real random variable on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_{t+} = \sigma(A_s \mid s \leq t)$ and $\mathcal{F}_{t-} = \sigma(A_s \mid s < t)$, where $A_s = \{X \leq s\}$. Show that

(i) $\mathcal{F}_{t+} = \sigma(X \wedge t, 1_{A_t}) \quad , \quad \mathcal{F}_{t-} = \sigma(X \wedge t)$

(ii) $$\mathcal{F}_{t+}(\omega_0) = \begin{cases} \{X = a\} & \text{if } a = X(\omega_0) \leq t \\ \{X > t\} & \text{if } a = X(\omega_0) > t \end{cases}$$

(iii) $$\mathcal{F}_{t-}(\omega_0) = \begin{cases} \{X = a\} & \text{if } a = X(\omega_0) < t \\ \{X \geq t\} & \text{if } a = X(\omega_0) \geq t \end{cases}$$

Let $\psi : \mathbf{R} \to \mathbf{R}$ be a Borel function such that $E|\psi(X)| < \infty$ and set

$$\alpha(t) = \frac{1}{P(X > t)} \int_{\{X > t\}} \psi(X)\, dP \quad , \quad \beta(t) = \frac{1}{P(X \geq t)} \int_{\{X \geq t\}} \psi(X)\, dP$$

with the convention $\frac{1}{0} \cdot 0 = 0$. Show that

(iv) $E(\psi(X) \mid \mathcal{F}_{t+}) = \psi(X)\, 1_{\{X \leq t\}} + \alpha(t)\, 1_{\{X > t\}}$

(v) $E(\psi(X) \mid \mathcal{F}_{t-}) = \psi(X)\, 1_{\{X < t\}} + \beta(t)\, 1_{\{X \geq t\}}$

Hint: (i): Note that $A_s = \{X \wedge t \leq s\}$ for all $s < t$ and $A_t = \{1_{A_t} = 1\}$. Argue that this shows that $\mathcal{F}_{t+} \subseteq \sigma(X \wedge t, 1_{A_t})$ and $\mathcal{F}_{t-} \subseteq \sigma(X \wedge t)$. Note that $\{X \wedge t \leq c\} = A_c$ if $c < t$ and $\{X \wedge t \leq c\} = \Omega$ if $c \geq t$. Argue that this implies the converse inclusions. (ii)+(iii): Apply (6.3.12). (iv): Let $U = \psi(X)\, 1_{\{X \leq t\}} + \alpha(t)\, 1_{\{X > t\}}$ and show that U is $\mathcal{F}_{t+}$-measurable and

$$E\psi(X) = EU \quad \text{and} \quad \int_{A_s} \psi(X)\, dP = \int_{A_s} U\, dP \qquad \forall s \leq t$$

Finally apply (6.7.12) with $X := \psi(X)$, $Y := U$ and $\mathcal{H} = \{A_s \mid s \leq t\}$ to prove (iv).

Exc. 6.14: Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X, Y \in L^1(P)$ be random variables with $EX = EY$. Let $\mathcal{H} \subseteq \mathcal{F}$ be a $(\cap f)$-stable paving and set $\mathcal{G} = \sigma(\mathcal{H})$. Suppose that Y is $\mathcal{G}$-measurable and that

$$\int_H Y\, dP = \int_H X\, dP \qquad \forall H \in \mathcal{H}$$

Show that $Y = E(X \mid \mathcal{G})$. [Apply Sierpinsky's stability lemma (1.6).]

Exc. 6.15: Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X, Y \in L^1(P)$ be random variables. Let $\mathcal{L}$ and $\mathcal{M}$ be two sub-σ-algebras of $\mathcal{F}$ such that

$$\int_{L \cap M} Y\, dP = \int_{L \cap M} X\, dP \qquad \forall L \in \mathcal{L}\, \forall M \in \mathcal{M}$$

Show that $E(X \mid \mathcal{G}) = E(Y \mid \mathcal{G})$, where $\mathcal{G} = \sigma(\mathcal{L} \cup \mathcal{M})$. [Apply Exc.6.14.]

Exc. 6.16: Let $X, Y \in L^1(P)$ be random variables such that $E(X|Y) = Y$ and $E(Y|X) = X$. Show that $X = Y$ a.s.

Hint: Let $\psi : \mathbf{R} \to \mathbf{R}$ be a fixed bounded, continuous, strictly increasing function, for instance, $\psi = \arctan$. Show that

$$E(\psi(X)(X - Y) \mid X) = 0 = E(\psi(Y)(Y - X) \mid Y)$$

$$EU = 0 \quad \text{where} \quad U = (\psi(X) - \psi(Y)) \cdot (X - Y)$$

Note that $U(\omega) \geq 0$ for all ω, and $U(\omega) = 0$ if and only if $X(\omega) = Y(\omega)$.

Exc. 6.17: Let $X \in L^2(P)$ be a random variable and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let $U \in L^2(P)$ be a $\mathcal{G}$-measurable random variable and let $f : \mathbf{R} \to \mathbf{R}$ be a *semi-contraction*, that is, $|f(x)-f(y)| \le |x-y|$ for all $x, y \in \mathbf{R}$. Show that

(i) $E\Big((X-U)^2 \,\Big|\, \mathcal{G}\Big) = \sigma^2(X|\mathcal{G}) + (E(X|\mathcal{G}) - U)^2$

(ii) $\sigma^2(X \mid \mathcal{G}) \le E\Big((X-U)^2 \,\Big|\, \mathcal{G}\Big)$

(iii) $E\,\sigma^2(X \mid \mathcal{G}) \le \sigma^2(X)$

(iv) $\sigma^2(f(X) \mid \mathcal{G}) \le \sigma^2(X \mid \mathcal{G})$

Hint: (iii): Set $U \equiv EX$ in (ii) and take means. (iv): Let $V = E(X \mid \mathcal{G})$. Then $(f(X) - f(V))^2 \le (X - V)^2$. Then apply (ii) with $U = f(V)$ and X replaced by $f(X)$.

Exc. 6.18: Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A_1, \ldots, A_n$ be exchangeable events [see Exc.2.26]. Let $A_\emptyset = \Omega$ and $A_\alpha = \cap_{j \in \alpha} A_j$ whenever α is a non-empty subset of $\{1, 2, \ldots, n\}$. Set

$$S_n(\omega) = \sum_{j=1}^{n} 1_{A_j}(\omega) = \#\{j \mid 1 \le j \le n\,,\ \omega \in A_j\}$$

Let α be a non-empty subset of $\{1, \ldots, n\}$ with $\#(\alpha) = j$. Show that

(i) $$P(A_\alpha \mid S_n = k) = \frac{\binom{n-j}{k-j}}{\binom{n}{k}} \qquad \forall\, 1 \le k \le n$$

Exc. 6.19: Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $F^+ = \{\, P(F \mid \mathcal{G}) > 0 \,\}$, where $F \in \mathcal{F}$ and $\mathcal{G}$ is a given sub-σ-algebra of $\mathcal{F}$. Show that F^+ satisfies

(i) $F^+ \in \mathcal{G}$ and $F \subseteq F^+$ a.s.

(ii) $G \in \mathcal{G}$ and $F \subseteq G$ a.s. $\Rightarrow$ $F^+ \subseteq G$ a.s.

A set F^+ satisfying (i) and (ii) is called a $(P \mid \mathcal{G})$-*hull* of F. Show that any two $(P \mid \mathcal{G})$-hulls of F are equal a.s.

Exc. 6.20: Let $W_0, W_1, \ldots$ be real random variables and let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ be an increasing sequence of sub-σ-algebras of $\mathcal{F}$ such that W_n is $\mathcal{F}_n$-measurable and W_{n+1} and $\mathcal{F}_n$ are independent for all $n = 0, 1, 2, \ldots$. Let $\phi : \mathbf{R}^2 \to \mathbf{R}$ be a Borel function and let X_n be defined recursively as follows:

$$X_0 := W_0 \quad , \quad X_n := \phi(W_n, X_{n-1}) \qquad \forall\, n = 1, 2, \ldots$$

Let $Q_n(B,x) = P(\phi(W_{n+1},x) \in B)$ for all $B \in \mathcal{B}(\mathbf{R})$ and all $x \in \mathbf{R}$. Show that

(i) X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$

(ii) $P(X_{n+1} \in B \mid \mathcal{F}_n) = Q_n(B, X_n) \quad \forall n \geq 0 \; \forall B \in \mathcal{B}(\mathbf{R})$

(iii) $\{X_n, \mathcal{F}_n \mid n \geq 0\}$ is a weak Markov chain

Hint: (i): Apply induction in n. (ii): Apply (6.8.14). (iii): Apply (ii) and (6.16.5).

Exc. 6.21: Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, let $X, Y : \Omega \to \overline{\mathbf{R}}$ be $\mathcal{G}$-measurable random variables, and let $Z \in L(P)$ such that $X \leq E(Z \mid \mathcal{G}) \leq Y$ a.s. Show that there exists a P-null set $N \in \mathcal{F}$ satisfying

(i) $\omega \in \Omega \setminus N$ and $Y(\omega) < \infty \quad \Rightarrow \quad Z(\omega) < \infty$

(ii) $\omega \in \Omega \setminus N$ and $X(\omega) > -\infty \quad \Rightarrow \quad Z(\omega) > -\infty$

Hint: Show that $\int_{G_n} Z\,dP \leq n$ and $\int_{H_n} Z\,dP \geq n$ for all integers $n \in \mathbf{Z}$ where $G_n = \{Y \leq n\}$ and $H_n = \{X \geq n\}$. Then apply (3.3.6+7).

Exc. 6.22: Let X be a random variable such that $1_A X \in L(P)$ where $A \in \mathcal{F}$ is a given set. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and let $G \in \mathcal{G}$ such that $A \subseteq G$ a.s. Show that $E(1_A X \mid \mathcal{G})(\omega) = 0$ for a.a. $\omega \in \Omega \setminus G$.

Hint: Note that $1_{\Omega \setminus G}\, 1_A\, |X| = 0$ a.s. and apply (6.8.2) and (6.9.4).

Chapter 7

MARTINGALES

7.1 Filters and Stopping Times

In Chapter 6, I explained how information can be described as a sub-σ-algebra of the basic σ-algebra $\mathcal{F}$ on Ω. I shall now add a dynamic element into this description. Usually or hopefully, we become wiser as time goes by, meaning that our information increases in time. Here, we shall only consider the discrete time case, meaning that we only update our information at times which are integer multiples of a given time unit (such as minutes, hours, days, weeks, etc.). Hence, we shall only consider time sets which are *integer intervals*, that is, sets of the form

$$[[\alpha, \beta]] := \{n \in \mathbf{Z} \mid \alpha \leq n \leq \beta\}$$

where $\alpha \in \mathbf{Z} \cup \{-\infty\}$, $\beta \in \mathbf{Z} \cup \{\infty\}$ and $\alpha \leq \beta$. If $T = [[\alpha, \beta]]$, we set $T^* = T \cup \{\alpha, \beta\}$, that is,

$$T^* = \begin{cases} T & \text{if} \quad -\infty < \alpha \leq \beta < \infty \\ T \cup \{\infty\} & \text{if} \quad -\infty < \alpha < \beta = \infty \\ T \cup \{-\infty\} & \text{if} \quad -\infty = \alpha < \beta < \infty \\ T \cup \{-\infty, \infty\} & \text{if} \quad -\infty = \alpha < \beta = \infty \end{cases}$$

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let T be an integer interval, then a *filter with time set* T is a collection $\{\mathcal{F}_n \mid n \in T\}$ of σ-algebras satisfying

(7.1.1) $\qquad \mathcal{F}_n \subseteq \mathcal{F}_m \subseteq \mathcal{F} \qquad \forall n, m \in T \quad \text{with} \quad n \leq m$

The σ-algebra $\mathcal{F}_n$ will be interpreted as *the information available at* time n. Thus, (7.1.1) just states that our information increases with time. If $T = [[\alpha, \beta]]$ is an integer interval and $\{\mathcal{F}_n \mid n \in T\}$ is a filter, we set

$$(7.1.2) \qquad \mathcal{F}_\alpha = \bigcap_{n \in T} \mathcal{F}_n \text{ if } \alpha = -\infty \quad , \quad \mathcal{F}_\beta = \sigma\left(\bigcup_{n \in T} \mathcal{F}_n\right) \text{ if } \beta = \infty$$

In this way, we have defined $\mathcal{F}_n$ for all $n \in T^*$. Note that $\mathcal{F}_\alpha$ represents our *initial information* and $\mathcal{F}_\beta$ represents our *ultimate information.*

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let T be an integer interval. Then a T-*time* is a random variable τ such that $\tau(\omega) \in T^*$ for all $\omega \in \Omega$. If τ is a T-time with $P(\tau \in T) = 1$, we say that τ is a *proper* T-*time.* Let τ be a T-time. Then $\{|\tau| > n\} \downarrow \{\tau \notin T\} = \{\tau = \pm\infty\}$. Hence, we find

$$(7.1.3) \qquad \tau \text{ is proper} \quad \Leftrightarrow \quad P(\tau = \pm\infty) = 0 \quad \Leftrightarrow \lim_{n \to \infty} P(|\tau| > n) = 0$$

In particular, we have:

(7.1.4) Let τ be a T-time where $T = [[\alpha, \beta]]$ is an integer interval. If T is finite, then τ is proper. If $\alpha > -\infty$, then we have

$$\tau \text{ is proper} \quad \Leftrightarrow \quad P(\tau = \infty) = 0 \quad \Leftrightarrow \quad \lim_{n \to \infty} P(\tau > n) = 0$$

and if $\beta < \infty$, then we have

$$\tau \text{ is proper} \quad \Leftrightarrow \quad P(\tau = -\infty) = 0 \quad \Leftrightarrow \quad \lim_{n \to -\infty} P(\tau < n) = 0$$

One of the most intriguing problems in gambling is the problem of stopping at the right time. Suppose that our good friend Ann is visiting a gambling house. Ann has decided to play at most 100 games, but she wants to stop at the time most favorable to her. Of course, the most favorable time to stop is the first time she has achieved her maximal net profit in the 100 games, but unfortunately she only knows this time after she has played all 100 games, and then it is too late to stop. On the other hand, she may decide to use the following strategy: "Stop at the first time your net profit exceeds \$ 10 if this ever happens; otherwise stop after 100 plays." What is the difference between this strategy and the clearly optimal strategy: "Stop at the first time you have achieved your maximal net profit."? The main difference is that the first strategy may be carried out without clairvoyance, but the second cannot. Let us look at a model.

Let X_n be Ann's net profit (possibly negative) in the first n games and let $\mathcal{F}_n$ be the information available to Ann at time n for $n \in T = \{1, 2, \ldots, 100\}$. It is reasonable to assume that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_{100}$ and that X_n is $\mathcal{F}_n$-measurable

for all $1 \le n \le 100$. Let $M = \max\{X_1, \ldots, X_{100}\}$ be the maximal net profit. The first strategy says: "Stop at time τ ", where τ is the T-time given by

$$\tau(\omega) = \begin{cases} \min\{n \in T \mid X_n(\omega) \ge 10\} & \text{if } M(\omega) \ge 10 \\ 100 & \text{if } M(\omega) < 10 \end{cases}$$

The second strategy says "Stop at time σ ," where

$$\sigma(\omega) = \min\{n \in T \mid X_n(\omega) = M(\omega)\}$$

Now note that $\{\tau > n\} = \{X_1 < 10, \ldots, X_n < 10\}$ for $1 \le n \le 99$ and that $\{\tau > 100\} = \emptyset$. Hence, $\{\tau > n\} \in \mathcal{F}_n$ for all $n \in T$, and so at each time $n \in T$ Ann can decide if she should continue [if $\tau(\omega) > n$] or stop [if $\tau(\omega) \le n$] based on her information at time n . The T-time σ does not have this property since $\{\sigma > n\} = \{X_1 < M. \ldots . X_n < M\}$ involves knowledge of all the 100 random variables $X_1, \ldots, X_{100}$. Inspired by this example, we make the following definition, which will be the cornerstone of this chapter:

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{\mathcal{F}_n \mid n \in T\}$ be a filter where $T = [[\alpha, \beta]]$ is a given integer interval. Then a *stopping time* for the filter $\{\mathcal{F}_n \mid n \in T\}$ is a T-time τ satisfying

(7.1.5) $\{\tau > n\} \in \mathcal{F}_n \qquad \forall n \in T$

A *proper stopping time* τ , is a stopping time which is a proper T-time, that is, a stopping time satisfying $P(\tau \in T) = 1$. Note that (7.1.5) says that at every time $n \in T$ we can decide to stop or continue based on our information $\mathcal{F}_n$ available at time n . If $\beta = \infty$, then a stopping time τ may assume the value $+\infty$, meaning that we never stop playing, and if $\alpha = -\infty$, then a stopping time may assume the value $-\infty$, meaning that we never start playing. Let τ be a T-time and let $k, n \in T$ and $B \subseteq \overline{\mathbf{R}}$ be given. Then we have

$$\{\tau = n\} = \{\tau > n-1\} \setminus \{\tau > n\} \ , \ \{\tau \in B\} = \{\tau \in B^c\}^c = \bigcup_{j \in T^* \cap B} \{\tau = j\}$$

$$\{\tau = \alpha\} = \bigcap_{j \in [[\alpha, k]]} \{\tau > j\}^c \ , \ \{\tau = \beta\} = \bigcap_{j \in [[k, \beta]]} \{\tau > j-1\}$$

Hence, if τ is a T-time, then we have the following:

(7.1.6) τ is a stopping time for $\{\mathcal{F}_n\}$ if and only if $\{\tau = n\} \in \mathcal{F}_n$ for all $n \in T^*$

(7.1.7) If τ is a stopping time for $\{\mathcal{F}_n\}$, then $\{\tau \in B\} \in \mathcal{F}_n$ whenever $n \in T$ and either $B \subseteq [-\infty, n]$ or $B \supseteq]n, \infty]$

Let τ be a stopping time for the filter $\{\mathcal{F}_n \mid n \in T\}$. Then we define *the information available at time* τ to be the paving $\mathcal{F}_\tau$ given by

(7.1.8) $$F \in \mathcal{F}_\tau \iff F \cap \{\tau = n\} \in \mathcal{F}_n \qquad \forall n \in T^*$$

Let $F \in \mathcal{F}_\tau$. Then (7.1.8) says that if $\tau(\omega) = n$, we know if F has occurred or not based on our information $\mathcal{F}_n$ available at time n . Thus, $\mathcal{F}_\tau$ may be interpreted as the information of the random time τ . Observe that

$$\{\tau = n\} \cap F^c = \{\tau = n\} \setminus (F \cap \{\tau = n\})$$
$$\{\tau = n\} \cap \bigcup_{j=1}^{\infty} F_j = \bigcup_{j=1}^{\infty} \{\tau = n\} \cap F_j$$

Hence, if τ is a stopping time, then exactly as above we find the following:

(7.1.9) $\mathcal{F}_\tau$ is a σ-algebra on Ω such that $\mathcal{F}_\tau \subseteq \mathcal{F}$, and $F \in \mathcal{F}_\tau$ if and only if $\{\tau \le n\} \cap F \in \mathcal{F}_n$ for all $n \in T^*$

(7.1.10) $$F \cap \{\tau \in B\} \in \mathcal{F}_n \qquad \forall F \in \mathcal{F}_\tau \ \forall n \in T^* \ \forall B \subseteq [-\infty, n]$$

Note that the stopping time: "The first time Ann's net profit exceeds \$ 10," may be described as the first time that the event $\{X_n \ge 10\}$ occurs. In applications, most stopping times are given in this way, namely, as the first time that a certain event occurs. In the next section we shall study such T-times.

7.2 Occurrence Times

Let $\{\mathcal{F}_n \mid n \in T\}$ *be a filter on the probability space* $(\Omega, \mathcal{F}, P)$ *where* $T = [[\alpha, \beta]]$ *is an integer interval. Let* τ *be a given* T*-time and let* $\Gamma = \{F_n \mid n \in T\} \subseteq \mathcal{F}$ *be a given sequence of events. Then we define*

(7.2.1) $$\Upsilon_\Gamma(\omega) := \inf\{n \in T \mid \omega \in F_n\} \qquad \forall \omega \in \Omega$$

(7.2.2) $$(\Upsilon_\Gamma \nabla \tau)(\omega) := \inf\{n \in T \mid n > \tau(\omega) \text{ and } \omega \in F_n\} \qquad \forall \omega \in \Omega$$

(7.2.3) $$\Upsilon_\Gamma^1 := \Upsilon_\Gamma \quad , \quad \Upsilon_\Gamma^{k+1} := \Upsilon_\Gamma \nabla \Upsilon_\Gamma^k \qquad \forall k \ge 1$$

with the convention: $\inf \emptyset = \beta$. *If* $m \in T$ *such that* $m < \beta$, *then we have the following:*

(7.2.4) *If* $F_n \in \mathcal{F}_n$ *for all* $n \in T$, *then* Υ_Γ *and* Υ_Γ^k *are stopping times for* $\{\mathcal{F}_n\}$ *for all* $k = 1, 2, \ldots$

(7.2.5) *If* $F_n \in \mathcal{F}_n$ *for all* $n \in T$ *and* τ *is a stopping time for* $\{\mathcal{F}_n\}$ *, then* $\Upsilon_\Gamma \nabla \tau$ *is a stopping time for* $\{\mathcal{F}_n\}$

(7.2.6) $$\{\Upsilon_\Gamma > m\} = \bigcap_{j \in [[\alpha, m]]} F_j^c \quad , \quad \{\Upsilon_\Gamma = \beta\} = \bigcap_{j \in [[\alpha, \beta-1]]} F_j^c$$

Remarks (1): Note that Υ_Γ is the first time that the event F_n occurs and the T-time Υ_Γ is called *the first occurrence time* in the sequence Γ . Similarly, $\Upsilon_\Gamma \nabla \tau$ is the first time after τ that the event F_n occurs and the T-time $\Upsilon_\Gamma \nabla \tau$ is called *the first occurrence time after time* τ in the sequence Γ . Finally, by induction we see that Υ_Γ^k is the kth time, that the event F_n occurs and Υ_Γ^k is called *the* k^{th} *occurrence time* in the sequence Γ .

(2): Let τ be any stopping time for $\{\mathcal{F}_n\}$ and set $F_n = \{\tau \le n\}$ for $n \in T$. Then $F_n \in \mathcal{F}_n$ for all $n \in T$. If $\Gamma = \{F_n\}$ then $\tau = \Upsilon_\Gamma$. Hence, every stopping time is the first occurrence time in some sequence $\{F_n\}$ satisfying $F_n \in \mathcal{F}_n$ for all $n \in T$.

(3): Returning to Ann's stopping strategy τ , we have $T = \{1, 2, \ldots, 100\}$ and $\tau = \Upsilon_\Gamma$ where $\Gamma = \{F_n\}$ and $F_n = \{X_n \ge 10\}$ for all $1 \le n \le 100$. Since X_n is $\mathcal{F}_n$-measurable, we have that $F_n \in \mathcal{F}_n$ and so τ is, indeed, a stopping time.

Proof (7.2.6) follows directly from the definition of Υ_Γ . Hence, if $F_n \in \mathcal{F}_n$ for all $n \in T$, then Υ_Γ is a stopping time. Note that if $\Lambda = \{G_n\}$, where $G_n = F_n \cap \{\tau < n\}$, then $\Upsilon_\Gamma \nabla \tau = \Upsilon_\Lambda$. Hence, if $F_n \in \mathcal{F}_n$ for all $n \in T$ and τ is a stopping time, then by the argument above we have that $\Upsilon_\Gamma \nabla \tau$ is a stopping time. Thus, by induction we see that Υ_Γ^k is a stopping time and the theorem is proved. □

7.3 The Calculus of Stopping Times

Let $\{\mathcal{F}_n \mid n \in T\}$ *be a filter on the probability space* $(\Omega, \mathcal{F}, P)$ *, let* σ *and* τ *be stopping times, and let* $k \in T^*$ *. Then we have the following:*

(7.3.1) $\xi(\omega) \equiv k$ *is a stopping time and* $\mathcal{F}_\xi = \mathcal{F}_k$

(7.3.2) *If* $\sigma \le \tau$ *then* $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$

(7.3.3) $\sigma \wedge \tau$ *is a stopping time and* $\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$

(7.3.4) $\sigma \vee \tau$ *is a stopping time and* $\mathcal{F}_{\sigma \vee \tau} = \sigma(\mathcal{F}_\sigma \cup \mathcal{F}_\tau)$

(7.3.5) $\{\sigma \le \tau\}$, $\{\sigma < \tau\}$ *and* $\{\sigma = \tau\}$ *all belong to* $\mathcal{F}_{\sigma \wedge \tau}$

(7.3.6) τ *is* $\mathcal{F}_\tau$ *-measurable*

(7.3.7) *If* ρ *is* $\mathcal{F}_\tau$*-measurable,* $\rho \geq 0$ *and* $\rho + \tau$ *is a* T*-time, then* $\rho + \tau$ *is a stopping time*

(7.3.8) *If* $\sigma \wedge \tau \geq 0$ *and* $\sigma + \tau$ *is a* T*-time, then* $\sigma + \tau$ *is a stopping time*

Proof (7.3.1): Since $\{\xi > n\}$ is either $\emptyset$ or Ω, we see that ξ is a stopping time; and since $\{\xi = n\} \cap F = \emptyset$ for $n \neq k$ and $\{\xi = k\} \cap F = F$, we see that $\mathcal{F}_\xi = \mathcal{F}_k$.

(7.3.2): Let $F \in \mathcal{F}_\sigma$ and $n \in T^*$. Since $\sigma \leq \tau$, we have

$$\{\tau = n\} \cap F = \bigcup_{j \in T^*, j \leq n} \{\sigma = j\} \cap F \cap \{\tau = n\} \in \mathcal{F}_n$$

because $\{\sigma = j\} \cap F \in \mathcal{F}_j \subseteq \mathcal{F}_n$ for $j \leq n$and $\{\tau = n\} \in \mathcal{F}_n$. Thus, $F \in \mathcal{F}_\tau$ and so $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

(7.3.3): Since $\{\sigma \wedge \tau > n\} = \{\sigma > n\} \cap \{\tau > n\}$, we see that $\sigma \wedge \tau$ is a stopping time and by (7.3.2), we have $\mathcal{F}_{\sigma\wedge\tau} \subseteq \mathcal{F}_\sigma \cap \mathcal{F}_\tau$. Let $F \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$ and $n \in T^*$. Then

$$\{\sigma \wedge \tau \leq n\} \cap F = (\{\sigma \leq n\} \cap F) \cup (\{\tau \leq n\} \cap F)$$

so by (7.1.9) we have that $F \in \mathcal{F}_{\sigma\wedge\tau}$. Thus, $\mathcal{F}_{\sigma\wedge\tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.

(7.3.4): Since $\{\sigma \vee \tau > n\} = \{\sigma > n\} \cup \{\tau > n\}$, we see that $\sigma \vee \tau$ is a stopping time and by (7.3.2), we have $\mathcal{F}_{\sigma\vee\tau} \supseteq \sigma(\mathcal{F}_\sigma \cup \mathcal{F}_\sigma)$. Let $F \in \mathcal{F}_{\sigma\vee\tau}$ and set $F_0 = F \cap \{\sigma \leq \tau\}$ and $F_1 = F \cap \{\tau \leq \sigma\}$. Then

$$F_0 \cap \{\tau = n\} = F \cap \{\sigma \vee \tau = n\} \cap \{\tau = n\} \in \mathcal{F}_n$$
$$F_1 \cap \{\sigma = n\} = F \cap \{\sigma \vee \tau = n\} \cap \{\sigma = n\} \in \mathcal{F}_n$$

for all $n \in T^*$. Hence, $F = F_0 \cup F_1$, $F_0 \in \mathcal{F}_\tau$ and $F_1 \in \mathcal{F}_\sigma$. Thus, $F \in \sigma(\mathcal{F}_\sigma \cup \mathcal{F}_\tau)$ and so $\mathcal{F}_{\sigma\vee\tau} = \sigma(\mathcal{F}_\sigma \cup \mathcal{F}_\tau)$.

(7.3.5): Since $\{\sigma \leq \tau\} \cap \{\sigma \wedge \tau = n\} = \{\sigma = n\} \cap \{\tau \geq n\} \in \mathcal{F}_n$, we see that $\{\sigma \leq \tau\} \in \mathcal{F}_{\sigma\wedge\tau}$. In a similar manner, we find that $\{\sigma < \tau\}$ and $\{\sigma = \tau\}$ belongs to $\mathcal{F}_{\sigma\wedge\tau}$.

(7.3.6): Note that $\{\tau \in B\} \cap \{\tau = n\} = \{\tau = n\} \in \mathcal{F}_n$ if $n \in B$, and $\{\tau \in B\} \cap \{\tau = n\} = \emptyset \in \mathcal{F}_n$ if $n \notin B$. Hence, $\{\tau \in B\}$ belongs to $\mathcal{F}_\tau$ for all $B \subseteq \overline{\mathbf{R}}$. Thus, τ is $\mathcal{F}_\tau$-measurable.

(7.3.7): Since $\rho \geq 0$ and ρ is $\mathcal{F}_\tau$-measurable, we have

$$\{\rho + \tau \leq n\} = \bigcup_{j \in T^*, j \leq n} \{\rho \leq n - j\} \cap \{\tau = j\} \in \mathcal{F}_n$$

for all $n \in T$. Hence, $\rho + \tau$ is a stopping time.

(7.3.8): Set $\rho = \sigma \wedge \tau$. Since $\rho + (\sigma \vee \tau) = \sigma + \tau$ and $\rho \geq 0$, we see that (7.3.8) follows from (7.3.7), (7.3.6) and (7.3.2). □

7.4 Composite Random Variables

Let $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $\{\mathcal{F}_n \mid n \in T\}$ *be a filter where* $T = [[\alpha, \beta]]$ *is a given integer interval. Let* $\{Y_n \mid n \in T^*\}$ *be a sequence of random variables, let* τ *be a* T*-time, and let* $Y_\tau(\omega) := Y_{\tau(\omega)}(\omega)$ *for all* $\omega \in \Omega$ *. Then we have the following:*

(7.4.1) *If* τ *is a stopping time and* Y_n *is* $\mathcal{F}_n$*-measurable for all* $n \in T^*$ *, then* Y_τ *is* $\mathcal{F}_\tau$*-measurable*

(7.4.2) $$Y_\tau = \sum_{n \in T^*} 1_{\{\tau = n\}} Y_n$$

(7.4.3) $$EY_\tau = \sum_{n \in T^*} \int_{\{\tau = n\}} Y_n dP \quad \text{if} \quad Y_\tau \in L(P)$$

(7.4.4) $$Y_\tau = \sum_{n \in T} 1_{\{\tau = n\}} Y_n \quad \text{a.s. if } \tau \text{ is proper}$$

(7.4.5) $$EY_\tau = \sum_{n \in T} \int_{\{\tau = n\}} Y_n dP \quad \text{if } \tau \text{ is proper and} \quad Y_\tau \in L(P)$$

Suppose that τ *is a stopping time and that* $Y_\tau \in L(P)$ *and* $Y_n \in L(P)$ *for all* $n \in T^*$ *. If* $U_n := E(Y_n \mid \mathcal{F}_n)$ *for all* $n \in T^*$ *, then we have*

(7.4.6) $$E(Y_\tau \mid \mathcal{F}_\tau) = U_\tau$$

Moreover, if $Y_\tau \in L(P)$ *and* $1_{\{\tau = n\}} Y_n \in L(P)$ *for all* $n \in T^*$ *, then (7.4.6) remains valid if we define* $U_n := E(1_{\{\tau = n\}} Y_n \mid \mathcal{F}_n)$ *for all* $n \in T^*$ *.*

Remarks (1): The random variable Y_τ is called *a composite random variable*. The random variable Y_τ is so to say doubly stochastic: We pick a random variable at random. This means that you should be careful in computation with Y_τ . Equation (7.4.6) is a simple version of *the optional sampling theorem* which we shall meet many times later. Note that (7.4.6) states that a stopping time acts as if it was non-random. Below we shall see that Y_τ in many cases acts as if τ is non-random; however, this is **not** always so, and even when it is so, it is usually far from evident.

(2): It happens often that we only have given Y_n for $n \in T$, but not for $n = \pm\infty$ but we still want to define Y_τ . If $\tau(\omega) \in T$ for all $\omega \in \Omega$, there is no problem: We just define: $Y_\tau(\omega) := Y_{\tau(\omega)}(\omega)$. However, if $\tau(\omega) = \pm\infty$ for some ω we cannot define Y_τ by (7.4.4) without specifying Y_∞ and/or $Y_{-\infty}$. If τ is proper, then the problem is negligible because $P(\tau = \pm\infty) = 0$ and we *define* Y_τ by (7.4.4). If τ is not proper, we have to specify Y_∞ and/or

$Y_{-\infty}$. This can, of course, be done in many ways. ***However, if no other way is explicitly stated I shall always use the standard specifications***:

$$(7.4.7) \qquad Y_\infty(\omega) = \begin{cases} \lim_{n\to\infty} Y_n(\omega) & \text{if the limit exists in } \overline{\mathbf{R}} \\ 0 & \text{otherwise} \end{cases} \qquad \text{if } \beta = \infty$$

$$(7.4.8) \qquad Y_{-\infty}(\omega) = \begin{cases} \lim_{n\to-\infty} Y_n(\omega) & \text{if the limit exists in } \overline{\mathbf{R}} \\ 0 & \text{otherwise} \end{cases} \qquad \text{if } \alpha = -\infty$$

Proof (7.4.1): Note that $\{\tau = n\} \cap \{Y_\tau \in B\} = \{\tau = n\} \cap \{Y_n \in B\} \in \mathcal{F}_n$ for all $n \in T^*$ and all $B \in \mathcal{B}(\overline{\mathbf{R}})$. Thus, Y_τ is $\mathcal{F}_\tau$-measurable.

(7.4.2)–(7.4.5): (7.4.2) follows directly by definition of Y_τ and (7.4.3) follows from (7.4.2) and (3.13.2). If τ is proper, then $P(\tau \notin T) = 0$. Hence, (7.4.4+5) follow from (7.4.2+3).

(7.4.6): Since $U_n = E(Y_n \mid \mathcal{F}_n)$ is $\mathcal{F}_n$ -measurable for all $n \in T^*$, we have that U_τ is $\mathcal{F}_\tau$-measurable by (7.4.1). Let $F \in \mathcal{F}_\tau$ be given. Since $F \cap \{\tau = n\} \in \mathcal{F}_n$, then by (7.4.3) we get

$$\begin{aligned} \int_F Y_\tau dP &= \sum_{n\in T^*} \int_{\{\tau=n\}\cap F} Y_n dP = \sum_{n\in T^*} \int_{\{\tau=n\}\cap F} U_n dP \\ &= \int_F U_\tau dP \end{aligned}$$

and so (7.4.6) holds. The same argument works if Y_τ and $1_{\{\tau=n\}} Y_n$ belongs to $L(P)$ and $U_n := E(1_{\{\tau=n\}} Y_n \mid \mathcal{F}_n)$. □

7.5 Martingales

Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter on the probability space $(\Omega, \mathcal{F}, P)$ where T is a given integer interval. If X_n is a random variable for each $n \in T$, then we say that $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a *submartingale* with *time set* T if

(7.5.1) $\quad X_n$ is $\mathcal{F}_n$-measurable and $EX_n^+ < \infty \quad \forall n \in T$

(7.5.2) $\quad X_n \le E(X_{n+1} \mid \mathcal{F}_n) \quad$ if $n \in T$ and $n+1 \in T$

We say that $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a *supermartingale* with *time set* T if

(7.5.3) $\quad X_n$ is $\mathcal{F}_n$-measurable and $EX_n^- < \infty \quad \forall n \in T$

(7.5.4) $\quad X_n \ge E(X_{n+1} \mid \mathcal{F}_n) \quad$ if $n \in T$ and $n+1 \in T$

Finally, we say that $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a *martingale* with *time set* T if

(7.5.5) $\quad X_n$ is $\mathcal{F}_n$-measurable and $E|X_n| < \infty \quad \forall n \in T$

(7.5.6) $\quad X_n = E(X_{n+1} \mid \mathcal{F}_n) \quad$ if $n \in T$ and $n+1 \in T$

Note that (7.5.2) states that at every time n we expect the future value X_{n+1} to be larger than the present value X_n. Thus, popularly speaking, a submartingale is an optimist who believes that things get better and better. Similarly, you should think of a supermartingale as a pessimist believing that the future will be worse than today; whereas a martingale is an indifferent person, believing that the weather tomorrow will be like the weather today. The word martingale is an ancient word common to all European languages with several different meanings, all referring to a safety device. Webster's Encyclopedic Dictionary gives the following explanation:

> **martingale**, mär'tin·gäl, n.; old kind of breeches; from *Martigal*, an inhabitant of *Martigues*, in Provence; a strap from a horse's head to the girth under his belly and passing between his forelegs, to prevent him from rearing; naut. a short perpendicular spar under the bowsprit.

However, the word has plenty of other meanings. Among gamblers, a martingale is the name of a certain strategy which some optimistic gamblers believe is a sure winner. Suppose that you are playing on a roulette: If you bet on red, you get twice your stake if red comes out, and you lose your stake if not. The martingale strategy goes as follows. You start betting say \$1 on red; if you lose you bet \$2 on red if you lose again you bet \$4 etc., doubling the stake whenever you lose until you finally, win. Then you start all over again betting \$1 on red, etc. Observe that if red comes out, the nth time, then your stake was \$ 2^{n-1} and you win \$ 2^n. Moreover, the sum of your stakes in the n games equals: $1 + 2 + 4 + \cdots + 2^{n-1} = 2^n - 1$. Thus, your net profit is increased by \$1 whenever red comes out. The martingale strategy seems to be a sure way to beat the gambling house – but be careful! It is, as we shall see below, instead a sure way to be ruined. If p is the probability of red [which on an European roulette equals $\frac{18}{37} = 48.6\%$, on an American roulette equals $\frac{18}{38} = 47.4\%$, and on an Mexican roulette equals $\frac{18}{39} = 46.2\%$], then it is fairly easy [see (7.15)] to see that your net profit X_n after the nth game forms a martingale if $p = \frac{1}{2}$, a supermartingale if $p < \frac{1}{2}$, and a submartingale if $p > \frac{1}{2}$.

Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ be a ***submartingale***. Then a simple induction using (6.8.8) and (6.8.4) shows that

(7.5.7) $\quad X_n \le E(X_k \mid \mathcal{F}_n) \qquad \forall n, k \in T \quad$ with $\quad n \le k$

(7.5.8) $\quad EX_n \le EX_k < \infty \qquad \forall n, k \in T \quad$ with $\quad n \le k$

Similarly, if $\{X_n, \mathcal{F}_n\}$ is a ***supermartingale***, then

(7.5.9) $\quad X_n \geq E(X_k \mid \mathcal{F}_n) \qquad \forall n, k \in T \quad$ with $\quad n \leq k$

(7.5.10) $\quad EX_n \geq EX_k > -\infty \qquad \forall n, k \in T \quad$ with $\quad n \leq k$

Finally, if $\{X_n, \mathcal{F}_n\}$ is a ***martingale***, then

(7.5.11) $\quad X_n = E(X_k \mid \mathcal{F}_n) \qquad \forall n, k \in T \quad$ with $\quad n \leq k$

(7.5.12) $\quad -\infty < EX_n = EX_k < \infty \qquad \forall n, k \in T$

Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ be a submartingale and let φ be an increasing convex function from the open interval $]a, b[$ into $\mathbf{R}$. Suppose that $a \leq X_n \leq b$ a.s. and $\varphi(X_n) \in L(P)$ for all $n \in T$, then by Jensen's inequality (6.9.3) we have

$$\varphi(X_n) \leq \varphi(E(X_{n+1} \mid \mathcal{F}_n)) \leq E(\varphi(X_{n+1}) \mid \mathcal{F}_n)$$

and the converse inequality holds if $\{X_n, \mathcal{F}_n\}$ is a supermartingale and φ is increasing and concave. Moreover, note that this still holds if $\{X_n, \mathcal{F}_n\}$ is a martingale and φ is convex, but not necessarily increasing. Thus, we have the following:

(7.5.13) If $\{X_n, \mathcal{F}_n\}$ is submartingale and $\varphi :]a, b[\to \mathbf{R}$ is an increasing convex function such that $E\varphi(X_n) < \infty$ and $a \leq X_n \leq b$ a.s. for all $n \in T$, then $\{\varphi(X_n), \mathcal{F}_n\}$ is a submartingale

(7.5.14) If $\{X_n, \mathcal{F}_n\}$ is supermartingale and $\varphi :]a, b[\to \rho$ is a concave increasing function such that $E\varphi(X_n) > -\infty$ and $a \leq X_n \leq b$ a.s. for all $n \in T$, then $\{\varphi(X_n), \mathcal{F}_n\}$ is a supermartingale

(7.5.15) If $\{X_n, \mathcal{F}_n\}$ is martingale and $\varphi :]a, b[\to \mathbf{R}$ is a convex [respectively concave] function such that $E\varphi(X_n) < \infty$ [respectively $E\varphi(X_n) > -\infty$] and $a \leq X_n \leq b$ a.s. for all $n \in T$, then $\{\varphi(X_n), \mathcal{F}_n\}$ is a submartingale [respectively a supermartingale]

For instance, if $\{X_n, \mathcal{F}_n\}$ is a martingale such that $EX_n^2 < \infty$ for all n, then $\{X_n^2, \mathcal{F}_n\}$ is a submartingale because $\varphi(t) = t^2$ is convex. If $\{X_n, \mathcal{F}_n\}$ is a submartingale and $a \in \mathbf{R}$, then $\{(X_n - a)^+, \mathcal{F}_n\}$ is a submartingale because $\varphi(t) = (t-a)^+$ is increasing and convex and $E(X_n - a)^+ \leq |a| + EX_n^+ < \infty$. Similarly, if $\{X_n, \mathcal{F}_n\}$ is a non-negative supermartingale, then $\{\sqrt{X_n}, \mathcal{F}_n\}$ is a supermartingale because $\varphi(t) = \sqrt{t}$ is increasing and concave on $[0, \infty[$ and $E\sqrt{X_n} \geq 0 > -\infty$. Finally, observe the following trivial, but very useful facts:

(7.5.16) $\{X_n\}$ is a supermartingale if and only if $\{-X_n\}$ is a submartingale

(7.5.17) $\{X_n\}$ is a martingale if and only if $\{X_n\}$ is both a submartingale and a supermartingale

This means that any result for submartingales may be converted into analogous results for supermartingales and martingales. Thus, most of the results in this chapter comes in three versions: a submartingale, a supermartingale and a martingale version. Being an optimist, I shall do the proofs for submartingales, and for the most important results write down the supermartingale and martingale version, and for the rest I shall leave the formulation of the supermartingale and martingale version to the reader.

A submartingale or supermartingale or martingale with time set $T = [[-\infty, \beta]]$ for some $\beta \in \mathbf{Z}$ is called a *reversed submartingale or supermartingale or martingale.*

Note: In the definition of a submartingale we only assumed that $E X_n^+ < \infty$. This means that $E X_n < \infty$ and $X_n < \infty$ a.s. Hence, we may have submartingales $\{X_n\}$ with $EX_n = -\infty$ and even with $P(X_n = -\infty) > 0$. This is not a very essential point, but it is convenient, and it means that we have to be a little careful with the arithmetic in order to avoid $\infty - \infty$.

7.6 Examples of Martingales

Let me give a brief list of examples of submartingales or supermartingales or martingales. Most of the examples below will be treated in more details later.

Lévy martingales. Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter, and let X be a given random variable such that $X \in L(P)$. Set

(7.6.1) $$X_n = E(X \mid \mathcal{F}_n) \qquad \forall n \in T$$

Then by (6.8.8) we have

$$E(X_{n+1} \mid \mathcal{F}_n) = E(E(X \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n) = X_n$$

Hence, $\{X_n\}$ is a martingale if $E|X| < \infty$, a submartingale if $EX^+ < \infty$, and a supermartingale if $EX^- < \infty$. This may seem to be a rather artificial way to construct martingales – but it is not. In Section (7.12), we shall see that any uniformly integrable martingale is of the form (7.6.1) for some $X \in L^1(P)$.

Sums of independent random variables. Let $\{X_n \mid n \geq 1\}$ be a sequence of independent random variables with finite means $\mu_n = EX_n$. Set

(7.6.2) $$S_n = \sum_{j=1}^{n} X_j \ , \quad \hat{S}_n = \sum_{j=1}^{n} (X_j - \mu_j) \ , \quad \mathcal{F}_n = \sigma\{X_1, \ldots, X_n\}$$

Then $\{\mathcal{F}_n \mid 1 \le n < \infty\}$ is a filter, and by (6.8) we have

$$E(S_{n+1} \mid \mathcal{F}_n) = S_n + E(X_{n+1} \mid \mathcal{F}_n) = S_n + \mu_{n+1}$$

Hence, $\{S_n\}$ is a martingale if $\mu_n = 0$ for all n, a submartingale if $\mu_n \ge 0$ for all n, and a supermartingale if $\mu_n \le 0$ for all n. Moreover, $\{\hat{S}_n\}$ is always a martingale. If $E|X_n|^2 < \infty$, then we define

$$(7.6.3) \qquad SSD_n = \sum_{j=1}^{n} (X_j - \mu_j)^2 \quad , \quad \sigma_j^2 = \mathrm{var} X_j \quad , \quad s_n^{\,2} = \sum_{j=1}^{n} \sigma_j^2$$

Since $\sigma_j^2 = E(X_j - \mu_j)^2$, then, as above, we have that $\{SSD_n\}$ is a submartingale and $\{SSD_n - s_n^2\}$ is a martingale. Moreover, by (7.5.15) we have that $\{\hat{S}_n^2\}$ is a submartingale, and since $\hat{S}_n^2$ is $\mathcal{F}_n$-measurable and $(X_{n+1} - \mu_{n+1})$ is independent of $\mathcal{F}_n$, then by (6.8) we get

$$\begin{aligned} E(\hat{S}_{n+1}^2 \mid \mathcal{F}_n) &= E(\hat{S}_n^2 + (X_{n+1} - \mu_{n+1})^2 + 2\,\hat{S}_n(X_{n+1} - \mu_{n+1}) \mid \mathcal{F}_n) \\ &= \hat{S}_n^2 + E((X_{n+1} - \mu_{n+1})^2 \mid \mathcal{F}_n) + 2\,\hat{S}_n\, E(X_{n+1} - \mu_{n+1} \mid \mathcal{F}_n) \\ &= \hat{S}_n^2 + \sigma_{n+1}^2 \end{aligned}$$

Thus, we see that $\{\hat{S}_n^2 - s_n^2\}$ is a martingale.

Exchangeable random variables. Let $X = (X_1, X_2, \cdots)$ be sequence of exchangeable random variables [see (6.14)] and let $f : \mathbf{R}^\infty \to \mathbf{R}$ be a Borel function such that $f(X) \in L(P)$. Let $\mathcal{S}_X^n$ be the permutation invariant σ-algebra based on X [see (6.17.13)] and let $U_n = S_n f(X)$ be the U-statistics associated with f [see (6.17.7)]. Set

$$(7.6.4) \qquad M_j = U_{-j} \quad \text{and} \quad \mathcal{F}_j = \mathcal{S}_X^{-j} \qquad \forall -\infty < j \le -1$$

Then $\{\mathcal{F}_j \mid -\infty < j \le -1\}$ is a filter by (6.17.14), and by (6.18.1) we have $M_j = E(f(X) \mid \mathcal{F}_j)$. Thus, M_j is of the form (7.6.1), and so $\{M_j, \mathcal{F}_j \mid j \le -1\}$ is a reversed martingale if $E|f(X)| < \infty$, a reversed submartingale if $E\,f(X) < \infty$, and a reversed supermartingale if $Ef(X) > -\infty$. In particular, if $E|X_j| < \infty$ and

$$(7.6.5) \qquad \overline{X}_n = \tfrac{1}{n} \sum_{j=1}^{n} X_j \ , \ \overline{SSD}_n = \tfrac{1}{n} SSD_n \ , \ \mathfrak{s}_n^{\,2} = \tfrac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \overline{X}_n\right)^2$$

where SSD_n is defined as above, then $\{\overline{X}_{-j}, \mathcal{F}_j \mid j \le -1\}$ is a reversed martingale, and if $EX_j^2 < \infty$ then $\{\overline{SSD}_{-j}, \mathcal{F}_j \mid j \le -1\}$ and $\{\mathfrak{s}_n^2, \mathcal{F}_j \mid j \le -1\}$ are reversed martingales [see (6.18.4)–(6.18.11)].

Wald martingales. Let $\{Z_j \mid j \ge 1\}$ be a sequence of non-negative independent random variables with finite means $\mu_j = EZ_j$. Let $L_0 = 1$ and

$$(7.6.6) \qquad L_n = \prod_{j=1}^{n} Z_j \quad , \quad M_n = L_n + \sum_{j=1}^{n} (1 - \mu_j) L_{j-1}$$

Let $\mathcal{F}_n = \sigma\{Z_1, \ldots, Z_n\}$. Then by (6.8) we have

$$E(L_{n+1} \mid \mathcal{F}_n) = L_n E(Z_{n+1} \mid \mathcal{F}_n) = \mu_{n+1} L_n$$

Hence, $\{L_n\}$ is a submartingale if $\mu_n \ge 1$ for all n , a supermartingale if $\mu_n \le 1$ for all n , and a martingale if $\mu_n = 1$ for all n . Moreover, since

$$E(M_{n+1} \mid \mathcal{F}_n) = \mu_{n+1} L_n + \sum_{j=1}^{n+1} (1 - \mu_j) L_{j-1} = M_n$$

we see that $\{M_n\}$ is a martingale. If $\mu_j > 0$ for all j , then we see that

$$(7.6.7) \qquad \hat{L}_n = \prod_{j=1}^{n} \left(\frac{Z_j}{\mu_j}\right) \qquad \text{is a martingale}$$

In particular, if $\{X_j \mid j \ge 1\}$ are independent identically distributed random variables with Laplace transform $L(t)$ [see (4.19)], then

$$(7.6.8) \qquad \Lambda_n = L(t)^{-n} \exp\left(t \sum_{j=1}^{n} X_j\right) \qquad \text{is a martingale} \qquad \forall t \in \mathcal{R}(L)$$

The examples above reveal two important features about martingales. First, they are everywhere, and then crop up at the most unexpected places. In chapter 11 of Volume II, we shall see that in any sequence whatsoever of random variables there is hidden plenty of martingales. Second, in general, it is easy to ***prove*** that a given sequence is a martingale. I have talked a lot about sequences of independent random variables, but except for the example in (6.27) I have always ***assumed*** independence, never ***proved*** it – and for a good reason. In general, it is hard work to ***prove*** independence, so the easy way out is to ***assume*** it.

7.7 Stopped Submartingales

Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ *be a submartingale where* $T = [[\alpha, \beta]]$ *is an integer interval and let* $X_{-\infty}$ *be an* $\mathcal{F}_\alpha$*-measurable random variable satisfying*

$$X_{-\infty} \le E(X_k \mid \mathcal{F}_\alpha) \qquad \forall k \in T \tag{7.7.1}$$

Let τ *be a stopping time, then* $\{X_{\tau\wedge n}, \mathcal{F}_n \mid n \in T\}$ *is a submartingale and we have*

$$X_{\tau\wedge k} \le E(X_k \mid \mathcal{F}_\tau) \qquad \forall k \in T \tag{7.7.2}$$

Moreover, if σ *is a stopping time with* $P(\sigma = -\infty) = 0$ *, then*

$$X_{\tau\wedge\sigma\wedge k} \le E(X_{\tau\wedge k} \mid \mathcal{F}_\sigma) \qquad \forall k \in T \tag{7.7.3}$$

Remarks (1): Note that we only need the random variable $X_{-\infty}$, when $\alpha = -\infty$, and this is one of the rare cases where I do not use the standard specification (7.4.8) for defining $X_{-\infty}$. There are, of course many $\mathcal{F}_\alpha$-measurable random variables satisfying (7.7.1); for instance, $X_{-\infty} \equiv -\infty$ or $X_{-\infty} = \inf_k E(X_k \mid \mathcal{F}_\alpha)$. In Section (7.11), I shall show that if $\alpha = -\infty$, then the standard specification (7.4.8) satisfies (7.7.1), but this requires some preparation.

(2): Inequalities (7.7.2+3) is our second version of *the optional sampling theorem.* In Section (7.14) we shall obtain a more general optional sampling theorem.

(3): The submartingale $\{X_{\tau\wedge n}, \mathcal{F}_n \mid n \in T\}$ is called a *stopped submartingale,* and we define *stopped supermartingales* and *martingales* similarly. Of course, (7.5.16+17) give analogous results for supermartingales and martingales, but I shall leave the formulations of these to yourself.

Proof (7.7.2): Let $k \in T$ be given and set $U_n = E(X_k \mid \mathcal{F}_n)$. If $n \ge k$, then $U_n = X_k$ because X_k is $\mathcal{F}_n$-measurable for all $n \ge k$. If $n \le k$, then by (7.5.7) [if $n \ne -\infty$] or (7.7.1) [if $n = -\infty$] we have $U_n \ge X_n$. Hence, $U_n \ge X_{k\wedge n}$ for all $n \in T^*$ and $U_\tau = E(X_k \mid \mathcal{F}_\tau) \ge X_{\tau\wedge k}$ by (7.4.6). Thus, (7.7.2) holds. Let $Y_n = X_{\tau\wedge n}$; then $EY_n \le EX_n < \infty$ by (7.7.2) and Y_n is $\mathcal{F}_n$-measurable by (7.4.1) and (7.3.2). Let $F \in \mathcal{F}_n$ be given. Then $F \cap \{\tau > n\} \in \mathcal{F}_n$ and since

$$Y_n = X_n\, 1_{\{\tau>n\}} + X_\tau\, 1_{\{\tau\le n\}} \quad , \quad Y_{n+1} = X_{n+1}\, 1_{\{\tau>n\}} + X_\tau\, 1_{\{\tau\le n\}}$$

then by (7.5.2) we find

$$\begin{aligned}\int_F Y_n dP &= \int_{F\cap\{\tau>n\}} X_n dP + \int_{F\cap\{\tau\le n\}} X_\tau dP \\ &\le \int_{F\cap\{\tau>n\}} X_{n+1}\, dP + \int_{F\cap\{\tau\le n\}} X_\tau\, dP \\ &= \int_F Y_{n+1}\, dP\end{aligned}$$

Thus, $Y_n \le E(Y_{n+1} \mid \mathcal{F}_n)$ by (6.8.3) and (6.8.6) and so $\{Y_n, \mathcal{F}_n\}$ is a submartingale.

(7.7.3): Let $Y_n = X_{\tau\wedge n}$ and set $Y_{-\infty} \equiv -\infty$. Then $Y_{-\infty}$ satisfies (7.7.1) with $\{X_k\}$ replaced by $\{Y_k\}$. Let σ be a stopping time with $P(\sigma = -\infty) = 0$. Then $Y_{\sigma\wedge k} = X_{\tau\wedge\sigma\wedge k}$ a.s. and so (7.7.3) follows by applying (7.7.2) to $\{Y_n\}$ and σ. □

7.8 The Martingale Inequalities

Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ be a submartingale where $T = [[\alpha, \beta]]$ is an integer interval, let $X_{-\infty}$ be an $\mathcal{F}_\alpha$-measurable random variable satisfying (7.7.1), and let $\mathcal{S}$ be the set of all stopping times. Then the family $\{X^+_{\sigma\wedge n} \mid \sigma \in \mathcal{S}\}$ is uniformly integrable for all $n \in T$. Let τ be a given a stopping time and let

$$W_\tau := \sup_{n\in T} X_{\tau\wedge n} \quad , \quad V_{k,\tau} := \inf_{n\in[[k,\beta]]} X_{\tau\wedge n}$$

where $k \in T$ is a given integer. Then we have the following:

(7.8.1) $$\int_{\{\tau<\infty\}} X^+_\tau dP \le \sup_{n\in T} EX^+_{\tau\wedge n} \le \sup_{n\in T} EX^+_n$$

(7.8.2) $$\int_{\{\tau<\infty\}} X^-_\tau dP \le \sup_{n\in T} E(X^+_{\tau\wedge n} - X_k) \quad \textit{if} \quad \tau \ge k \quad \textit{a.s.}$$

(7.8.3) $$\int_{\{\tau<\infty\}} |X_\tau| dP \le \sup_{n\in T} E(2X^+_{\tau\wedge n} - X_k) \quad \textit{if} \quad \tau \ge k \quad \textit{a.s.}$$

(7.8.4) $$P(W_\tau \ge t) \le \frac{1}{t} \sup_{n\in T} EX^+_{\tau\wedge n} \qquad \forall t > 0$$

(7.8.5) $$P(V_{k,\tau} \le -t) \le \frac{1}{t} \sup_{n\in T} E(X^+_{\tau\wedge n} - X_k) \quad \textit{if} \quad \tau \ge k \quad \textit{a.s. and} \quad t > 0$$

(7.8.6) *If $\inf_{n\in T} EX_n > -\infty$, then $\{X_n \mid n \in [[\alpha, k]]\}$ is uniformly integrable*

Remark By (7.5.16) and (7.5.17) we get analogous results for supermartingales and martingales, but I shall leave the proof of these to the reader [see Exc.7.8].

Proof By (7.7.2), we have that $X^+_{\tau\wedge k} \le E(X^+_k \mid \mathcal{F}_\tau)$. Hence, by (6.10.1) we have that $\{X^+_{\tau\wedge k} \mid \tau \in \mathcal{S}\}$ is uniformly integrable for all $k \in T$.

(7.8.1): Since $X^+_{\tau\wedge n} \to X^+_\tau$ on $\{\tau < \infty\}$ and $EX^+_{\tau\wedge n} \le EX^+_n$ by (7.7.2), we se see that (7.8.1) follows from Fatou's lemma (3.5).

(7.8.2+3): Since $\tau \ge k$ and $\{X_{\tau\wedge n}\}$ is a submartingale by (7.7), we have that $EX_{\tau\wedge n} \ge E\,X_{\tau\wedge k} = EX_k$ for all $n \ge k$. Recall that $t^- = t^+ - t$ for all $t \in \mathbf{R}$, Hence, we get

$$EX^-_{\tau\wedge n} = EX^+_{\tau\wedge n} - EX_{\tau\wedge n} \le E(X^+_{\tau\wedge n} - X_k) \qquad \forall n \ge k$$

Thus, (7.8.2) follows from Fatou's lemma as above. Moreover, (7.8.3) follows in the same manner using the formula: $|t| = 2t^+ - t$.

(7.8.4): Let $q \in T$ and $s > 0$ be given numbers. Then we define $W_{q,\tau} = \sup\{X_{\tau\wedge n} \mid n \in [[q,\beta]]\}$ and $\sigma = \inf\{n \in [[q,\beta]] \mid X_{\tau\wedge n} > s\}$ with the convention: $\inf\emptyset = \beta$. By (7.2.4), we have that σ is a stopping time and $W_{q,\tau}(\omega) > s$ implies $\sigma(\omega) \in T$ and $X^+_{\tau\wedge\sigma}(\omega) > s$. Hence, applying (7.8.1) on the stopping time $\tau \wedge \sigma$ and using (7.7.3), we get

$$\begin{aligned} s\,P(W_{q,\tau} > s) &\le \int_{\{\tau\wedge\sigma<\infty\}} X^+_{\tau\wedge\sigma}\,dP \\ &\le \sup_{n\in T} E(X^+_{\tau\wedge\sigma\wedge n}) \le \sup_{n\in T} E(X^+_{\tau\wedge n}) \end{aligned}$$

If $\alpha > -\infty$, then (7.8.4) follows by setting $q = \alpha$ and letting $s \uparrow t$. If $\alpha = -\infty$, then $W_{q,\tau} \uparrow W_\tau$ as $q \downarrow -\infty$; thus, (7.8.4) follows by letting $q \downarrow -\infty$ and $s \uparrow t$.

(7.8.5): Setting $\sigma = \inf\{n \in [[k,\beta]] \mid X_{\tau\wedge n} < -s\}$ and applying (7.8.2) and (7.7.3) in the same manner as above, we see that (7.8.5) follows.

(7.8.6): By assumption there exists $c \in \mathbf{R}$ such that $X_n \in L^1(P)$ and $\mu_n := EX_n \ge c$ for all $n \in T$. If $\alpha > -\infty$, then $\{X_n \mid \alpha \le n \le k\}$ is uniformly integrable by (3.22.37). So suppose that $\alpha = -\infty$. Since $\{\mu_n\}$ is increasing in n and bounded below, there exist $\mu \in \mathbf{R}$ such that $\mu_n \to \mu$ as $n \to -\infty$. Hence, if $\varepsilon > 0$ is given, there exists $q \le k$ such that $|\mu_n - \mu_q| < \varepsilon$ for all $n \le q$. Set $\rho(F) = \int_F \mid X_q \mid dP$ and let $-\infty < n \le q$ and $F \in \mathcal{F}_n$ be given. Since $F, F^c \in \mathcal{F}_n$, $\{X_n\}$ and $\{X^+_n\}$ are submartingales

and $|t| = 2t^+ - t$, we get

$$\begin{aligned}\int_F |X_n|\,dP &= 2\int_F X_n^+\,dP - \int_F X_n\,dP \\ &= 2\int_F X_n^+\,dP + \int_{F^c} X_n\,dP - \mu_n \\ &\le 2\int_F |X_q|\,dP + \mu_q - \int_F X_q\,dP - \mu_n \\ &\le 3\int_F |X_q|\,dP + |\mu_q - \mu_n| \\ &\le 3\rho(F) + \varepsilon\end{aligned}$$

Thus, by (3.24.3) and (3.13.3) we conclude that $\{X_n \mid n \le k\}$ is uniformly integrable. □

7.9 The Upcrossing Inequality

Let $\{X_n \mid n \in T\}$ be a sequence of random variables where $T = [[\alpha, \beta]]$ is an integer interval with $\alpha > -\infty$. Let $r, s \in \mathbf{R}$ be given such that $r < s$. We want to count how many times the sequence $X_\alpha, X_{\alpha+1}, \cdots$ crosses the interval $[r, s]$ in an upward direction. Set

$$\begin{aligned} F_n &= \{X_n \le r\} \quad , \quad \Gamma = \{F_n \mid n \in T\} \\ G_n &= \{X_n \ge s\} \quad , \quad \Lambda = \{G_n \mid n \in T\}\end{aligned}$$

Then $\{X_n\}$ begins its first upcrossing at the first occurrence in Γ [that is, at time Υ_Γ], it finishes the first upcrossing at the first occurrence in Λ after time Υ_Γ [that is, at time $\Upsilon_\Lambda \nabla \Upsilon_\Gamma$]. Then it begins its second upcrossing at the first occurrence in Γ after time $\Upsilon_\Lambda \nabla \Upsilon_\Gamma$ [that is, at time $\Upsilon_\Gamma \nabla (\Upsilon_\Lambda \nabla \Upsilon_\Gamma)$], etc. Thus, let us define

(7.9.1) $\qquad \rho_1 = \Upsilon_\Gamma \quad , \quad \sigma_1 := \Upsilon_\Lambda \nabla \rho_1$

(7.9.2) $\qquad \rho_{k+1} := \Upsilon_\Gamma \nabla \sigma_k \quad , \quad \sigma_{k+1} := \Upsilon_\Lambda \nabla \rho_{k+1}$

If $\{X_n\}$ admits at least k upcrossings, then the kth upcrossing begins at time ρ_k and finishes at time σ_k . Hence, $\{X_n\}$ admits at least k upcrossings if and only if $\sigma_k \in T$, $X_{\rho_k} \le r$, and $X_{\sigma_k} \ge s$. Thus, *the upcrossing number of* $\{X_n\}$ *over the interval* $[r, s]$, denoted U_{rs} , is given by

(7.9.3) $\qquad U_{rs}(\omega) := \sup\{k \in \mathbf{N} \mid \sigma_k(\omega) \in T \ , \ X_{\rho_k}(\omega) \le r \ , \ X_{\sigma_k}(\omega) \ge s\}$

with the convention that $\sup \emptyset = 0$ [there may be no upcrossing at all]. Note that if $\{\mathcal{F}_n \mid n \in T\}$ is a filter, then by (7.2.4+5) we have

(7.9.4) If X_n is $\mathcal{F}_n$-measurable for all $n \in T$, then ρ_k and σ_k are stopping times for all $k = 1, 2, \cdots$

If $\{X_n\}$ is increasing, then the upcrossing number is at most 1. If $\{X_n\}$ is a submartingale, then $\{X_n\}$ is not necessarily increasing, but we expect it to be increasing. So it is reasonable to believe that the expected number of upcrossing in a submartingale is fairly small. That is exactly, what *the upcrossing inequality* (7.9.5) tells us.

The upcrossing inequality. *Let U_{rs} denote the upcrossing number of $\{X_\alpha, X_{\alpha+1}, \ldots, X_\beta\}$ over the interval $[r, s]$ where $\{X_n, \mathcal{F}_n \mid \alpha \le n \le \beta\}$ is a submartingale, $\alpha, \beta \in \mathbf{Z}$ and $-\infty < r < s < \infty$. Then we have*

$$EU_{rs} \le \frac{1}{s-r} E\{(X_\beta - r)^+ - (X_\alpha - r)^+\} \tag{7.9.5}$$

Proof: Let $Y_n = (X_n - r)^+$, let ρ_k and σ_k be defined as above, and let $\sigma_0 = \alpha$. By (7.5.13), we know that $\{Y_n\}$ is a submartingale with $E|Y_n| < \infty$. Hence, Y_n is finite a.s. and since $\sigma_{j-1} \le \rho_j \le \beta$, then by (7.7.3) we have $Y_{\sigma_{j-1}} \le E(Y_{\rho_j} \mid \mathcal{F}_{\sigma_{j-1}})$. Thus, we have

(i) $$E(Y_{\rho_j} - Y_{\sigma_{j-1}}) \ge 0 \qquad \forall j \ge 1$$

If $m = 1 + \beta - \alpha$, then $U_{rs} < m$ and the mth upcrossing does not exist. Hence, $\sigma_m \equiv \beta$ and so we get

$$\begin{aligned} Y_\beta - Y_\alpha &= Y_{\sigma_m} - Y_{\rho_m} + Y_{\rho_m} - Y_{\sigma_{m-1}} + \cdots + Y_{\rho_1} - Y_{\sigma_0} \\ &= \sum_{j=1}^{m} (Y_{\sigma_j} - Y_{\rho_j}) + \sum_{j=1}^{m} (Y_{\rho_j} - Y_{\sigma_{j-1}}) \end{aligned}$$

If $1 \le j \le U_{rs}$, then we have $Y_{\sigma_j} \ge s - r$ and $Y_{\rho_j} = 0$. If $j > U_{rs}$, then either $\sigma_j = \rho_j = \beta$ or $Y_{\rho_j} = 0$. Hence, $Y_{\sigma_j} - Y_{\rho_j} \ge 0$ and so we find

$$Y_\beta - Y_\alpha \ge (s - r)\, U_{rs} + \sum_{j=1}^{m} (Y_{\rho_j} - Y_{\sigma_{j-1}})$$

Thus, taking expectations and using (i) we get

$$E(Y_\beta - Y_\alpha) \ge (s - r)\, EU_{rs} + \sum_{j=1}^{m} E(Y_{\rho_j} - Y_{\sigma_{j-1}}) \ge (s - r)\, EU_{rs}$$

Hence, (7.9.5) follows by dividing through with $s - r$ and inserting $Y_\beta = (X_\beta - r)^+$ and $Y_\alpha = (X_\alpha - r)^+$. □

A memento: Did you notice that something strange happened in this proof ? Take a look at (i) and set $j = 2$. By definition of ρ_2 and σ_1 it appears as if $Y_{\rho_2} = 0$ and $Y_{\sigma_1} \geq s - r$, and so apparently we have

$$Y_{\rho_2}(\omega) - Y_{\sigma_1}(\omega) \leq r - s < 0$$

which evidently contradicts (i). However, remember that this only holds for those ω's for which $\{Y_n(\omega)\}$ begins a second upcrossing. Since $\{Y_n\}$ is a submartingale, we don't really expect this to happen and we see that the probability that $\{Y_n\}$ begins a second upcrossing must be so small that it makes up for (i).

7.10 The Martingale Convergence Theorem

(1): If $\{X_n, \mathcal{F}_n \mid \alpha \leq n < \infty\}$ is a submartingale with $c = \sup_n EX_n^+ < \infty$, then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and $X_\infty^+ \in L^1(P)$. Moreover,

(7.10.1) $\quad EX_\infty^+ \leq c \quad$ *and* $\quad E|X_\infty| \leq 2c - EX_n \quad \forall n \geq \alpha$

(7.10.2) $\quad \{X_n^+ \mid n \geq \alpha\}$ *uniformly integrable* $\Rightarrow X_n \leq E(X_\infty \mid \mathcal{F}_n) \ \forall n \geq \alpha$

(7.10.3) $\quad |X_\infty(\omega)| < \infty$ *for a.a.* $\omega \in \bigcup_{n=\alpha}^{\infty} \{X_n > -\infty\}$

(2): If $\{X_n, \mathcal{F}_n \mid n \geq \alpha\}$ is a supermartingale with $d = \sup_n EX_n^- < \infty$, then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and $X_\infty^- \in L^1(P)$. Moreover,

(7.10.4) $\quad EX_\infty^- \leq d \quad$ *and* $\quad E|X_\infty| \leq 2d + EX_n \quad \forall n \geq \alpha$

(7.10.5) $\quad \{X_n^- \mid n \geq \alpha\}$ *uniformly integrable* $\Rightarrow X_n \geq E(X_\infty \mid \mathcal{F}_n) \ \forall n \geq \alpha$

(7.10.6) $\quad |X_\infty(\omega)| < \infty$ *for a.a.* $\omega \in \bigcup_{n=\alpha}^{\infty} \{X_n < \infty\}$

(3): If $\{X_n, \mathcal{F}_n \mid n \geq \alpha\}$ is a martingale such that either $c = \sup EX_n^+ < \infty$ or $d = \sup_n EX_n^- < \infty$, then $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and $X_\infty \in L^1(P)$. Moreover,

(7.10.7) $\quad E|X_\infty| \leq \lim_{n\to\infty} E|X_n| < \infty$

(7.10.8) $\quad \{X_n \mid n \geq \alpha\}$ *uniformly integrable* $\Rightarrow X_n = E(X_\infty \mid \mathcal{F}_n) \ \forall n \geq \alpha$

Proof As promised I shall only prove the submartingale case; the two other cases follow easily from (7.5.16+17). Let $X^* = \limsup X_n$ and $X_* = \liminf X_n$, let D be the set of all rational pairs $(r,s) \in \mathbf{Q}^2$ with $r < s$, and let $(r,s) \in D$ and $m \in T$ be given. Let $A_{rs} = \{X_* < r < s < X^*\}$ and let U_k denote the upcrossing number for $\{X_n \mid m \le n \le k\}$ over the interval $[r,s]$. Then $U_k \uparrow U$, where U is the upcrossing number of the sequence $\{X_n \mid n \ge m\}$ over $[r,s]$. Since $(a-b)^+ \le a^+ + b^-$ and $EX_k^+ \le c$, then by the upcrossing inequality (7.9.5) we have

$$EU_k \le \frac{1}{s-r} E\{(X_k - r)^+ - (X_m - r)^+\} \le \frac{c + r^-}{s-r} = c_0 < \infty$$

Hence, by the monotone convergence theorem (3.4), we have $EU \le c_0$ and so $U < \infty$ a.s. Let $\omega \in A_{rs}$ be given; then $X^*(\omega) > s$ and $X_*(\omega) < r$. Hence, $X_k(\omega) > s$ for infinitely many k and $X_n(\omega) < r$ for infinitely many n, but then $U(\omega) = \infty$. Thus, $A_{rs} \subseteq \{U = \infty\}$ and since $P(U = \infty) = 0$ we have $P(A_{rs}) = 0$ for all $(r,s) \in D$. Note that $X_*(\omega) \ne X^*(\omega)$ if and only if the exists $(r,s) \in D$ such that $\omega \in A_{r,s}$. Hence, by countability of D we have that $P(X_* \ne X^*) = 0$; that is $X_* = X^*$ a.s. and so $X_\infty = \lim X_n$ exists a.s.

Since $|t| = 2t^+ - t$ and EX_n is increasing in n, then by Fatou's lemma (3.5) we find

$$EX_\infty^+ \le \liminf_{n\to\infty} EX_n^+ \le c < \infty$$

$$E|X_\infty| \le \liminf_{j\to\infty} E|X_j| = \liminf_{j\to\infty}(2EX_j^+ - EX_j) \le 2c - EX_n$$

Thus, (7.10.1) is proved, and (7.10.2) follows directly from (6.10.10) and (7.5.7).

(7.10.3): Let $n \ge k \ge \alpha$ and $q \in \mathbf{N}$ be given, and set $F = \{X_k > -q\}$. Then $F \in \mathcal{F}_k$ and so by (7.5.7), we get

$$\begin{aligned}\int_F |X_n| dP &= 2\int_F X_n^+ dP - \int_F X_n dP \le 2c - \int_F X_k dP \\ &\le 2c + q\end{aligned}$$

Letting $n \to \infty$, then by Fatou's lemma (3.5) we see that $\int_F |X_\infty| dP < \infty$. Hence $|X_\infty(\omega)| < \infty$ for a.a. $\omega \in F = \{X_k > -q\}$. Since this holds for all $k \ge \alpha$ and all $q \in \mathbf{N}$, we see that (7.10.3) holds. □

7.11 The Reversed Martingale Convergence Theorem

(1): If $\{X_n, \mathcal{F}_n \mid -\infty < n \le \beta\}$ is a submartingale, then $X_{-\infty} = \lim_{n\to-\infty} X_n$ exists a.s. and $X_{-\infty}^+ \in L^1(P)$. Moreover, we have the following:

(7.11.1) $\lim_{n\to-\infty} X_n^+ = X_{-\infty}^+$ *in 1-mean*

(7.11.2) *If* $\inf_{n\le\beta} EX_n > -\infty$ *, then* $\lim_{n\to-\infty} X_n = X_{-\infty}$ *in 1-mean*

(7.11.3) $X_{-\infty} \le E(X_n \mid \mathcal{F}_{-\infty}) \qquad \forall n \le \beta$

(7.11.4) $EX_{-\infty} = \inf_{n\le\beta} EX_n = \lim_{n\to-\infty} EX_n$

(2): If $\{X_n, \mathcal{F}_n \mid n \le \beta\}$ *is a supermartingale, then* $X_{-\infty} = \lim_{n\to-\infty} X_n$ *exists a.s. and* $X_{-\infty}^- \in L^1(P)$ *. Moreover, we have the following:*

(7.11.5) $\lim_{n\to-\infty} X_n^- = X_{-\infty}^-$ *in 1-mean*

(7.11.6) *If* $\sup_{n\le\beta} EX_n < \infty$ *, then* $\lim_{n\to-\infty} X_n = X_{-\infty}$ *in 1-mean*

(7.11.7) $X_{-\infty} \ge E(X_n \mid \mathcal{F}_{-\infty}) \qquad \forall n \le \beta$

(7.11.8) $EX_{-\infty} = \sup_{n\le\beta} EX_n = \lim_{n\to-\infty} EX_n > -\infty$

(3): If $\{X_n, \mathcal{F}_n \mid -\infty < n \le \beta\}$ *is a martingale, then* $X_{-\infty} = \lim_{n\to-\infty} X_n$ *exists a.s. and* $X_{-\infty} \in L^1(P)$ *. Moreover, we have*

(7.11.9) $\lim_{n\to-\infty} X_n = X_{-\infty}$ *in 1-mean*

(7.11.10) $EX_{-\infty} = EX_n$ *and* $X_{-\infty} = E(X_n \mid \mathcal{F}_{-\infty}) \qquad \forall n \le \beta$

Proof As usual I shall only prove the submartingale case; the two other cases follow easily from (7.5.16+17). Let $m \in T$ be given integer, and let U_k be the upcrossing number of $\{X_n \mid k \le n \le m\}$ over the interval $[r, s]$. Then by the upcrossing inequality (7.9.5) we have

$$EU_k \le \frac{1}{s-r} E[(X_m - r)^+ - (X_k - r)^+] \le \frac{E(X_m - r)^+}{s-r}$$

Hence, as in the proof of (6.10) we have that $X_{-\infty} = \lim X_n$ exists a.s., and since $\{X_n^+ \mid n \le m\}$ is uniformly integrable by (7.8), then by Dunford-Pettis' theorem (3.23.4) we have that $X_\infty^+ \in L^1(P)$ and $X_n^+ \to X_{-\infty}^+$ in 1-mean. Moreover, if $\inf EX_n > -\infty$, then by (7.8) we have that $\{X_n \mid n \le m\}$ is uniformly integrable. Hence, as above, we have that $X_n \to X_{-\infty}$ in 1-mean. Thus, (7.11.1) and (7.11.2) are proved.

(7.11.3+4): Let $q \in \mathbf{R}$ be given. Then $\{X_n \vee q\}$ is a submartingale by (7.5.13) and $X_n \vee q \ge q$; so by (6.8.6) we get

$$E(X_k \vee q \mid \mathcal{F}_{-\infty}) = E(E(X_k \vee q \mid \mathcal{F}_n) \mid \mathcal{F}_{-\infty}) \ge E(X_n \vee q \mid \mathcal{F}_{-\infty})$$

for all $n \le k \le \beta$. Letting $n \to -\infty$, then by (6.10.8) we find

$$E(X_k \vee q \mid \mathcal{F}_{-\infty}) \ge \liminf_{n \to -\infty} E(X_n \vee q \mid \mathcal{F}_{-\infty}) \ge X_{-\infty} \vee q$$

Letting $q \downarrow -\infty$ and using (6.10.4), we see that (7.11.3) holds. In particular, we see that $EX_{-\infty} \le c$ where $c = \inf EX_n$. If $c = -\infty$, then clearly we have that $c = EX_{-\infty}$. If $c > -\infty$, then $X_n \to X_{-\infty}$ in 1-mean by (7.11.2). Hence, (7.11.4) follows from (7.5.8) and the theorem is proved. □

7.12 Lévy's Theorem

Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter where $T = [[\alpha, \beta]]$ is an integer interval and let $\{M_n \mid n \in T\}$ be random variables. Then we have that

(7.12.1) $\{M_n, \mathcal{F}_n \mid n \in T\}$ *is a uniformly integrable martingale if and only if there exists* $M \in L^1(P)$ *such that* $M_n = E(M \mid \mathcal{F}_n)$ *for all* $n \in T$

If $\alpha = -\infty$, we have

(7.12.2) $X \in L(P) \Rightarrow \lim_{n \to -\infty} E(X|\mathcal{F}_n) = E(X|\mathcal{F}_{-\infty})$ *a.s.*

(7.12.3) $X \in L^1(P) \Rightarrow \lim_{n \to -\infty} E(X|\mathcal{F}_n) = E(X|\mathcal{F}_{-\infty})$ *a.s. and in 1-mean*

Similarly, if $\beta = \infty$, we have

(7.12.4) $X \in L(P) \Rightarrow \lim_{n \to \infty} E(X|\mathcal{F}_n) = E(X|\mathcal{F}_{\infty})$ *a.s.*

(7.12.5) $X \in L^1(P) \Rightarrow \lim_{n \to \infty} E(X|\mathcal{F}_n) = E(X|\mathcal{F}_{\infty})$ *a.s. and in 1-mean*

Proof (7.12.1): If $M_n = E(M \mid \mathcal{F}_n)$ for some $M \in L^1(P)$, then by (6.10.1) and (7.6.1) we have that $\{M_n, \mathcal{F}_n\}$ is a uniformly integrable martingale. Conversely, suppose that $\{M_n, \mathcal{F}_n\}$ is a uniformly integrable martingale and set $M = M_\beta$. Then $M \in L^1(P)$ and $M_n = E(M_\beta \mid \mathcal{F}_n)$ by (7.5.11) [if $\beta < \infty$] or (7.10.8) [if $\beta = \infty$].

(7.12.2+3): Suppose that $\alpha = -\infty$ and that $EX^+ < \infty$. If we set $X_n = E(X \mid \mathcal{F}_n)$, then by (7.6.1) we have that $\{X_n, \mathcal{F}_n \mid -\infty < n \le \beta\}$ is a reversed submartingale. Thus, by the reversed martingale convergence theorem

(7.11) we know that $X_n \to X_{-\infty}$ a.s. and by (1.40.3), we have that $X_{-\infty}$ is $\mathcal{F}_{-\infty}$-measurable. Hence, by (7.11.3), (6.8.8), (6.10.1) and (6.10.10) we find

$$\begin{aligned} X_{-\infty} &\le E(X_n \mid \mathcal{F}_{-\infty}) = E(E(X \mid \mathcal{F}_n) \mid \mathcal{F}_{-\infty}) = E(X \mid \mathcal{F}_{-\infty}) \\ &= \liminf_{k \to -\infty} E(X_k \mid \mathcal{F}_{-\infty}) \le E(X_{-\infty} \mid \mathcal{F}_{-\infty}) = X_{-\infty} \end{aligned}$$

Thus, $X_{-\infty} = E(X \mid \mathcal{F}_{-\infty})$ and so (7.12.2) follows in case that $EX^+ < \infty$. The case $EX^- < \infty$ follows in a similar manner. If $X \in L^1(P)$, then $\{X_n\}$ is a martingale and so (7.12.3) follows from (7.12.2) and (7.11.9).

(7.12.4+5) are proved in exactly the same manner as (7.12.2+3). □

7.13 Optional Times

Consider the martingale strategy explained in (7.5). The major drawback of this strategy is that it requires an infinite fortune to carry out. Doubling your stake every time you lose, makes the stakes to grow incredibly fast. For instance, if you have lost 20 times, then your stake at the 21st play should be more than \$1.000.000 . So you may easily be ruined before you get your payback. This means that not all stopping strategies are optional for you. I shall now introduce a class of T-times with the property that a given sequence of random variables $\{Y_n\}$ [for instance, your net loss at time n] does not become to large before your strategy τ tells you to stop.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\{Y_n \mid n \in T\}$ be a sequence of random variables where $T = [[\alpha, \beta]]$ is a given integer interval. Then a T-time τ is said to be *optional* for $\{Y_n\}$ if there exists $q \in T$ such that

(7.13.1) $\{Y_{\tau \wedge n} \mid n \in T\ ,\ n \ge q\}$ is uniformly integrable

Remember that if $\alpha = -\infty$, then we define $Y_{-\infty}$ by the standard specification (7.4.7), and if $\beta = \infty$, then we define Y_∞ by the standard specification (7.4.8). Since $\tau \wedge n$ runs through all times $\le \tau$ when n runs through T, we see that (7.13.1) ensures that $|Y_n|$ does not become too large before time τ. Observe that $Y_{\tau \wedge n}(\omega) = Y_\tau(\omega)$ for all $n \ge \tau(\omega)$ and $Y_{\tau \wedge n}(\omega) = Y_n(\omega)$ for all $n \le \tau(\omega)$. Hence, we find

$$Y_\tau(\omega) = \begin{cases} \lim\limits_{n \to \infty} Y_{\tau \wedge n}(\omega) & \text{if } \tau(\omega) < \infty \text{ or } \lim\limits_{n \to \infty} Y_n(\omega) \text{ exist} \\ 0 & \text{otherwise} \end{cases}$$

Hence, $|Y_\tau| \le \liminf |Y_{\tau\wedge n}|$ and so by Fatou's lemma (3.5), and Dunford–Pettis' theorem (3.23.4), we have

(7.13.2) If τ is an optional T-time for the sequence $\{Y_n \mid n \in T\}$, then

$$E|Y_\tau| < \infty \quad \text{and} \quad \lim_{n\to\infty} \int_{\{\tau<\infty\}} |Y_{\tau\wedge n} - Y_\tau|\, dP = 0$$

Let τ be a T-time such that $P(q \le \tau \le k) = 1$ for some $q, k \in T$. If $n \ge q$, then $Y_{\tau\wedge n}(\omega)$ equals one of the numbers $Y_q(\omega), Y_{q+1}(\omega), \ldots, Y_k(\omega)$. Hence, $|Y_{\tau\wedge n}| \le |Y_q| + \cdots + |Y_k|$ for all $n \in T$ with $n \ge q$. Thus, we find

(7.13.3) If $P(q \le \tau \le k) = 1$ for some $q, k \in T$, and if $E|Y_j| < \infty$ for all $q \le j \le k$, then τ is optional for $\{Y_n\}$

Suppose that $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a submartingale. If $\alpha = -\infty$, then by the reversed martingale convergence theorem (7.11) we have that the standard specification $X_{-\infty}$ given by (7.4.8) satisfies (7.7.1). Hence, by (7.8.6) [if $\alpha = -\infty$] or (7.13.3) [if $\alpha > -\infty$] we obtain the following result:

(7.13.4) If $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a submartingale and τ is a stopping time such that $P(\tau \le k) = 1$ for some $k \in T$, then τ is optional for $\{X_n^+\}$

Of course, we have analogue results for supermartingales and martingales [see (7.5.16+17)].

We say that $\{X_n \mid n \in T\}$ is a *pointwise a.s. monotone* sequence of random variables if there exists a null set $N \in \mathcal{F}$ such that for each $\omega \in \Omega \setminus N$, we have that $\{X_n(\omega)\}$ is either increasing or decreasing in n, but the sequence may be increasing for some ω and decreasing for others. Suppose that $\{X_n \mid n \in T\}$ is a pointwise a.s. monotone sequence of random variables and that τ is a T-time. Then we have that $X_{\tau\wedge n} \to X_\tau$ a.s. and $|X_{\tau\wedge n}| \le |X_\tau| + |X_q|$ a.s. for all $q \in T$ and all $n \ge q$. Hence, by (7.13.2) we get

(7.13.5) If $\{X_n \mid n \in T\}$ is pointwise a.s. monotone and τ is a T-time, then $X_{\tau\wedge n} \to X_\tau$ a.s. If, moreover, X_q is integrable for just one integer $q \in T$, then τ is optional for $\{X_n\}$ if and only if $E|X_\tau| < \infty$

7.14 The Optional Sampling Theorem

(1): Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ *be a submartingale where* $T = [[\alpha, \beta]]$ *is an integer interval, and let* τ *and* σ *be stopping times such that* τ *is optional for* $\{X_n^+\}$ *. Then* $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ *a.s. and* $EX_\tau^+ < \infty$ *. Moreover, we have*

(7.14.1) $\quad X_{\sigma\wedge\tau} \le E(X_\tau \mid \mathcal{F}_\sigma) \quad$ *and* $\quad EX_{\tau\wedge\sigma} \le EX_\tau < \infty$

(2): If $\{X_n, \mathcal{F}_n \mid n \in T\}$ *is a supermartingale, and* τ *and* σ *are stopping times such that* τ *is optional for* $\{X_n^-\}$ *, then* $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ *a.s. and* $EX_\tau^- < \infty$ *. Moreover, we have*

(7.14.2) $\quad X_{\sigma\wedge\tau} \ge E(X_\tau \mid \mathcal{F}_\sigma) \quad$ *and* $\quad EX_{\tau\wedge\sigma} \ge EX_\tau > -\infty$

(3): If $\{X_n, \mathcal{F}_n \mid n \in T\}$ *is a martingale, and* τ *and* σ *are stopping times such that* τ *is optional for* $\{X_n\}$ *, then* $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ *a.s. and in 1–mean, and* $E|X_\tau| < \infty$ *. Moreover, we have*

(7.14.3) $\quad X_{\sigma\wedge\tau} = E(X_\tau \mid \mathcal{F}_\sigma) \quad$ *and* $\quad EX_{\tau\wedge\sigma} = EX_\tau \ne \pm\infty$

In particular, if $\sigma \le \tau$ *, then* $X_\sigma = E(X_\tau \mid \mathcal{F}_\sigma)$ *and* $EX_\sigma = EX_\tau \ne \pm\infty$ *.*

Remark: This is our final version of the optional sampling theorem. In applications, you have to check the appropriate optionality condition. Note that (7.13.3)–(7.13.5) provide three criteria for optionality, and in Section (7.16) you will find seven more criteria for optionality. Remember that we use the standard specification (7.4.7) for defining $X_{-\infty}$ when $\alpha = -\infty$, and (7.4.8) for defining X_∞ when $\beta = \infty$.

Proof As usual, I shall only prove the submartingale case; the two other cases follow easily from (7.5.16+17). Let $X_{-\infty} = X_\alpha$ [see (7.4.8)]. Then by (7.5.7) [if $\alpha > -\infty$] or the reversed martingale theorem (7.11.3) [if $\alpha = -\infty$], we see that (7.7.1) holds. Hence, if $Z_n = X_{\tau\wedge n}$ for $n \ge \alpha$, then $Z_\beta = X_\tau$ and by (7.7) and (7.5.15), we have that $\{Z_n, \mathcal{F}_n \mid n \in T\}$ and $\{Z_n^+, \mathcal{F}_n \mid n \in T\}$ are submartingales. Moreover, by optionality of τ there exists $q \in T$ such that $\{Z_n^+ \mid n \ge q\}$ is uniformly integrable and by (7.8.6) we see that $\{Z_n^+ \mid n \le q\}$ is uniformly integrable. Thus, the whole sequence $\{Z_n^+ \mid n \in T\}$ is uniformly integrable. Hence, by (7.5.7) [if $\beta < \infty$] or the martingale convergence theorem (7.10) [if $\beta = \infty$], we conclude that $X_{\tau\wedge n} = Z_n \to Z_\beta = X_\tau$ a.s., $EX_\tau^+ < \infty$ and

(i) $\quad Z_n \le U_n \quad \forall n \in T \cup \{\beta\}$ where $U_n := E(X_\tau \mid \mathcal{F}_n) \quad \forall n \in T^*$

Suppose that $\alpha = -\infty$. Then by (7.12.2) and the reversed martingale convergence theorem (7.11) we have $U_j \to U_{-\infty}$ and $Z_j \to Z_{-\infty}$ a.s. when $j \to -\infty$. Hence, by (i) we see that $Z_n \le U_n$ for all $n \in T^*$ and so $X_{\sigma\wedge\tau} = Z_\sigma \le U_\sigma$. Moreover, by (7.4.6) with $Y_n := X_\tau$ for all $n \in T^*$, we conclude that $U_\sigma = E(X_\tau \mid \mathcal{F}_\sigma)$. Thus, $X_{\sigma\wedge\tau} \le E(X_\tau \mid \mathcal{F}_\sigma)$ and the theorem is proved. □.

7.15 Review of the Optional Sampling Theorem

The optional sampling theorem is of paramount importance in the theory as well as applications of martingales. Don't let the short proof deceive you; we have used every bit of the results in this chapter to prove it. Let us consider the real-life implications of the optional sampling theorem. Recall that we left our good friend Ann in a gambling house. Now any reasonable gambling house will devise the games such that the odds are against the players and in favor of the gambling house. This means that Ann's net-profit S_n after the nth game forms a supermartingale. Hence, (7.14.2) says that if Ann uses a stopping strategy which is optional for $\{S_n^-\}$, then the longer she plays the more she loses. Thus, the only way to beat the odds is to use non-optional stopping strategies. But then her loss S_n^- will become unboundedly large before the stopping strategy tells her to stop. Thus, if Ann's fortune is finite, then there is a positive (and even fairly high) probability that she will be ruined before time τ. On the other hand if Ann's fortune is infinite, then she cannot increase her fortune by gambling, and so there is no point in gambling. In plain words the optional sampling theorem may be put as follows: **You can't beat the odds!!!** You may at this point argue that stopping strategies just may be too simple and that you may come out with a positive mean profit if you choose more sophisticated strategies, such as letting your stake depend on the future history and/or change between various games. However, as we shall see in Section (7.17) this doesn't change the fact that unless your fortune is infinite there are no non-clairvoyant strategies which can change the given odds.

7.16 Criteria for Optionality

Let $T = [[\alpha, \infty]]$ be an integer interval and let $\{Y_n \mid n \in T\} \subseteq L^1(P)$ be random variables. Let τ be a T-time and set

$$R_n = 1_{\{\tau=\infty\}} Y_n \quad , \quad L_n = 1_{\{\tau>n\}} Y_n \quad , \quad m = \int_{\{\tau<\infty\}} |Y_\tau| \, dP$$

If $k \in T$ is any given number, then the following six statements are equivalent:

(7.16.1) τ *is optional for* $\{Y_n\}$

(7.16.2) $\{Y_{\tau\wedge n} \mid n \geq k\}$ *is uniformly integrable*

(7.16.3) $m < \infty$ *and* $\{L_n \mid n \geq k\}$ *is uniformly integrable*

(7.16.4) $m < \infty$, $\{R_n \mid n \geq k\}$ *is uniformly integrable, and*

$$\lim_{n\to\infty} \int_{\{n<\tau<\infty\}} |Y_n|\, dP = 0$$

(7.16.5) $m < \infty$, $\{R_n \mid n \geq k\}$ *is uniformly integrable, and*

$$\lim_{n\to\infty} \int_{\{n<\tau<\infty\}} |Y_\tau - Y_{\tau\wedge n}|\, dP = 0$$

(7.16.6) $m < \infty$ *and* σ *is optional for* $\{Y_n\}$ *for any* T*-time* $\sigma \leq \tau$ *with*

$$\int_{\{\sigma<\infty\}} |Y_\sigma|\, dP < \infty$$

In particular, if τ *and* σ *are optional for* $\{Y_n\}$, *so are* $\sigma \vee \tau$ *and* $\sigma \wedge \tau$.

Remark: Suppose that τ is a proper T-time. Then $m = E|Y_\tau|$ and $R_n = 0$ a.s. and so $\{R_n \mid n \in T\}$ is evidently uniformly integrable.

Proof: Throughout the proof let the random variables U , U_n and V_n be defined by

$$U = 1_{\{\tau<\infty\}} |Y_\tau| \quad , \quad U_n = 1_{\{\tau<\infty\}} |Y_\tau - Y_{\tau\wedge n}| \quad , \quad V_n = 1_{\{n<\tau<\infty\}} Y_n$$

(7.16.1) $\Rightarrow$ (7.16.2): By (7.13.1) there exists $q \in T$ such that $\{Y_{\tau\wedge n} \mid n \geq q\}$ is uniformly integrable. Since $Y_{\tau\wedge q}$ and Y_j are integrable, we have

$$M := |Y_{\tau\wedge q}| + \sum_{j=k\wedge q}^{q} |Y_j| \in L^1(P)$$

and since $|Y_{\tau\wedge n}| \leq M$ for all $k \leq n \leq q$, we see that (7.16.2) holds.

(7.16.2) $\Rightarrow$ (7.16.3): This follows from (7.13.2) because $|L_n| \leq |Y_{\tau\wedge n}|$.

(7.16.3) $\Rightarrow$ (7.16.4): Since $|R_n| \leq |L_n|$ and $|V_n| \leq |L_n|$, we have that $\{R_n \mid n \geq k\}$ and $\{V_n \mid n \geq k\}$ are uniformly integrable. Hence, (7.16.4) follows from Dunford-Pettis' theorem (3.23.4) because $V_n \to 0$.

(7.16.4) $\Rightarrow$ (7.16.5). Since $|Y_n(\omega) - Y_{\tau\wedge n}(\omega)| = 0$ for all $\omega \in \{\tau \leq n\}$, we have

$$\begin{aligned}\int_{\{\tau<\infty\}} |Y_\tau - Y_{\tau\wedge n\}}|\, dP &= \int_{\{n<\tau<\infty\}} |Y_\tau - Y_n|\, dP \\ &\leq \int_{\{n<\tau<\infty\}} U\, dP + \int_{\{n<\tau<\infty\}} |Y_n|\, dP\end{aligned}$$

The first term tends to 0 because $EU = m < \infty$ and the second term tends to 0 by assumption.

(7.16.5) $\Rightarrow$ (7.16.6): Let $\varepsilon > 0$ be given and let σ be a T-time such that $\sigma \leq \tau$ and $EW < \infty$, where $W = 1_{\{\sigma<\infty\}} |Y_\sigma|$. Since $EU_n \to 0$ by assumption, there exists $p \geq k$ so that $EU_n \leq \varepsilon$ for all $n \geq p$. Moreover, since $|Y_{\sigma\wedge n}| \leq |R_n| + U_n + U + W$, we have

$$\int_F |Y_{\sigma\wedge n}|\, dP \leq \varepsilon + \int_F (U+W)\, dP + \sup_{j\geq k} \int_F |R_j|\, dP \qquad \forall F \in \mathcal{F}$$

Recalling that $E(U+W) < \infty$ and $\{R_j \mid j \geq k\}$ is uniformly integrable by assumption, then by (3.23.4) we conclude that σ is optional for $\{Y_n\}$.

(7.16.6) $\Rightarrow$ (7.16.1): Evident!

Thus, the equivalence of (7.16.1)–(7.16.6) is established. Using the equivalence of (7.16.1) and (7.16.3), we see that $\sigma \wedge \tau$ and $\sigma \vee \tau$ are optional for $\{Y_n\}$ whenever σ and τ are optional for $\{Y_n\}$. □

7.17 Martingale Differences

Suppose that you are visiting a gambling house. Let D_0 be your fortune before you join the game and let D_j denote your gain or loss in the jth game if you put \$1 at stake. Moreover, let $\mathcal{F}_0$ represent your initial information about the game, and let $\mathcal{F}_j$ be your information right after the jth game for $j \geq 1$. You may then let your stake, say V_j, in the jth game depend on your information before the jth game, that is, V_j can be taken to be an $\mathcal{F}_{j-1}$-measurable random variable. Then your fortune after the first n games equals $D_0 + V_1 D_1 + \cdots + V_n D_n$. We shall now see that no matter how you choose your stakes V_j you can't change the odds. Formally, we consider the following setting:

Let $\{\mathcal{F}_n \mid n \geq 0\}$ be a filter and let $\{D_j \mid j \geq 0\} \subseteq L(P)$ be **finite** random variables such that D_j is $\mathcal{F}_j$-measurable for all $j \geq 0$. Then we say that the sequence $\{D_n, \mathcal{F}_n \mid n \geq 0\}$ is

- a *submartingale difference* if $E(D_{n+1} \mid \mathcal{F}_n) \geq 0$ for all $n \geq 0$
- a *supermartingale difference* if $E(D_{n+1} \mid \mathcal{F}_n) \leq 0$ for all $n \geq 0$
- a *martingale difference* if $E(D_{n+1} \mid \mathcal{F}_n) = 0$ for all $n \geq 0$

Thus, if your stake at the jth game equals \$1 for all $j \geq 1$ and S denotes your fortune after the first n games equals, then

(7.17.1) $$S_n = \sum_{j=0}^{n} D_j \quad \text{and} \quad D_{n+1} = S_{n+1} - S_n \qquad \forall n = 0, 1, 2, \ldots$$

and S_n is $\mathcal{F}_n$-measurable and finite for all $n \geq 0$. Hence by (6.8.7) we see that $E(S_{n+1} \mid \mathcal{F}_n) = S_n + E(D_{n+1} \mid \mathcal{F}_n)$, and so we get [see also Exc.7.3]

(7.17.2) $\{S_n, \mathcal{F}_n \mid n \geq 0\}$ is a submartingale, supermartingale or martingale if and only if $\{D_n, \mathcal{F}_n \mid n \geq 0\}$ is a submartingale difference with $ES_n^+ < \infty$, a supermartingale difference with $ES_n^- < \infty$ or a martingale difference with $E|S_n| < \infty$ for all $n \geq 0$

Suppose that you decide to let your stake, say V_j , in the jth game depend on the circumstances. Then your outcome in the jth game equals $V_j D_j$, and your fortune after the nth game equals

(7.17.3) $$T_n = D_0 + V_1 D_1 +, \ldots, + V_n D_n$$

Being non-clairvoyant, we must assume that $\{V_j \mid j \geq 1\}$ is *predictable* in the sense that V_j is $\mathcal{F}_{j-1}$-measurable for all $j \geq 1$. If so, we have

$$E(T_{n+1} \mid \mathcal{F}_n) = T_n + V_{n+1} E(D_{n+1} \mid \mathcal{F}_n)$$

provided that $T_j \in L(P)$ and $V_j D_j \in L(P)$ for all $j \geq 1$. Hence, if $\{V_j \mid j \geq 1\}$ is predictable, then we have the following:

(7.17.4) If $\{D_n, \mathcal{F}_n \mid n \geq 0\}\}$ is a martingale difference such that $E|V_j D_j| < \infty$ for all $j \geq 0$, then $\{T_n, \mathcal{F}_n \mid n \geq 0\}$ is a martingale

(7.17.5) Suppose that $V_j \geq 0$ for all $j \geq 1$. If $\{D_n, \mathcal{F}_n \mid n \geq 0\}$ is a submartingale difference with $EV_j D_j^+ < \infty$ for all $j \geq 0$ [respectively a supermartingale difference with $EV_j D_j^- < \infty$ for all $j \geq 0$], then $\{T_n, \mathcal{F}_n \mid n \geq 0\}$ is a submartingale [respectively a supermartingale]

Thus, the message is that no matter how sophisticated the strategies you choose are, you cannot change the odds. This may be the reason that so few probabilists are serious gamblers. The morale of this story may be summed up as follows: If you want to make a living from gambling, start your own gambling house, but don't gamble in others.

Let $V_j = 1_{\{\tau \geq j\}}$ where τ is a stopping time. Then V_j is $\mathcal{F}_{j-1}$-measurable, and if T_n is defined by (7.17.3), then we have

$$(7.17.6) \qquad T_n = \sum_{j=0}^{n} 1_{\{\tau \geq j\}} D_j = \sum_{j=0}^{n \wedge \tau} D_j = S_{n \wedge \tau}$$

which corresponds to the strategy of stopping at time τ .

Recall the original martingale strategy explained in (7.5). This strategy is a typical example of a strategy with random stakes. In this case, $D_1, D_2, \ldots$ are independent with $P(D_j = 1) = p$ and $P(D_j = -1) = 1 - p$, and $V_j = 2^{\alpha_j}$ where α_j is the number of plays between the last red and $(j-1)$. Thus, V_j is $\mathcal{F}_{j-1}$-measurable. From (7.6), we know that $\{S_n\}$ is a submartingale, a supermartingale or a martingale if $p \geq \frac{1}{2}$, $p \leq \frac{1}{2}$ or $p = \frac{1}{2}$, respectively. By (7.17.4+5), we see that $\{T_n\}$ behaves in the same way. Thus, the classical martingale strategy can **not** change a supermartingale into a submartingale.

Let us take a closer look at the optional sampling theorem. It is equation (7.14.1) which is important in applications, and one may ask if the optionality condition on τ is really necessary. The answer is: "essentially yes" in the following precise sense:

7.18 The Converse Optional Sampling Theorem

Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter where $T = [[\alpha, \beta]]$ is an integer interval, and let $\{X_n \mid n \in T^*\} \subseteq L(P)$ be random variables such that X_n is $\mathcal{F}_n$-measurable for all $n \in T^*$. Suppose that τ is a stopping time satisfying

(7.18.1) $\quad X_\tau \in L(P)$ *and* $X_{\tau \wedge n} \leq E(X_\tau \mid \mathcal{F}_n) \quad \forall n \in T$

(7.18.2) $\quad$ *If* $\alpha = -\infty$, *then* $X_{-\infty} \leq \limsup_{n \to -\infty} X_{\tau \wedge n}$ *a.s.*

If $\mathcal{S}$ is the set of all stopping times, then we have the following:

(7.18.3) $\quad X_{\tau \wedge \sigma} \leq E(X_\tau \mid \mathcal{F}_\sigma) \quad \forall \sigma \in \mathcal{S}$

(7.18.4) $\quad$ *If* $EX_\tau^+ < \infty$, *then* $\{X_{\tau \wedge \sigma}^+ \mid \sigma \in \mathcal{S}\}$ *is uniformly integrable and* σ *is optional for* $\{X_n^+\}$ *for every* $\sigma \in \mathcal{S}$ *with* $\sigma \leq \tau$

In particular, if $\{X_n, \mathcal{F}_n \mid n \in T\}$ *is a submartingale and* τ *is a stopping time, then (7.18.2) holds if we use the standard specifications (7.4.7+8) to define* $X_{\pm\infty}$ *when* $\alpha = -\infty$ *and/or* $\beta = \infty$ *. Moreover, we have the following:*

(7.18.5) τ *is optional for* $\{X_n^+\}$ *if and only if* τ *satisfies (7.18.1) and* $EX_\tau^+ < \infty$ *; and if so, then* σ *is optional for* $\{X_n^+\}$ *for every* $\sigma \in \mathcal{S}$ *with* $\sigma \leq \tau$

(7.18.6) *If* $\beta < \infty$ *, then every stopping* $\sigma \in \mathcal{S}$ *is optional for* $\{X_n^+\}$

Thus, except for finiteness of EX_τ^+ *, we see that (7.14.1) is equivalent to the optionality of* τ *. As usual, we have analogue results for supermartingales and martingales.*

Remarks: (1): If $\alpha = -\infty$ and $\lim_{n\to-\infty} X_n$ exists a.s. [for instance, if $\{X_n, \mathcal{F}_n\}$ is a submartingale or a supermartingale], then the standard specification (7.4.7) satisfies (7.18.2) for every stopping time τ .

(2): Suppose that $T = \{-1, -2, -3, \cdots\}$. Then by (7.18.6) we have that any stopping time τ is optional for $\{X_n^+\}$, and thus, (7.14.1) holds for all stopping times σ and τ . Moreover, by the reversed martingale convergence theorem (7.11) that any submartingale with time set T converges as $n \to -\infty$ to a limit satisfying (7.7.1). Hence, reversed submartingales, supermartingales and martingales behave much more regularly than those with the positive integers as time set. There is a meta-mathematical and rather airy explanation of this phenomenon, which is also found in other areas of probability. Suppose that $\{X_n \mid -\infty < n < 0\}$ is a stochastic process. Then at every time $n < 0$ the process has "lived" for an infinite time span, and so one may expect that the process has stabilized and settled down, much more than if the process is still young and only have lived in a finite amount of time. This kind of reasoning is, of course, mathematically unjustified and must be verified case by case, but it is a sound intuition.

Proof: (7.18.3): Since $X_\tau = X_{\tau\wedge\beta}$ is $\mathcal{F}_\beta$ -measurable, then by (7.17.7+8) and Levy's theorem (7.12.2) we see that $X_{\tau\wedge n} \leq E(X_\tau \mid \mathcal{F}_n)$ for all $n \in T^*$. Hence, (7.18.3) follows from (7.4.6).

(7.18.4): Follows from (7.18.3) and (6.10.1).

Suppose that $\{X_n, \mathcal{F}_n\}$ is a submartingale. If we use the standard specifications (7.4.7+8) to define $X_{\pm\infty}$ when $\alpha = -\infty$ and/or $\beta = \infty$, then, as noted above, we have that (7.18.2) holds for every stopping time τ . Moreover, (7.18.5) follows directly from (7.18.4) and the optional sampling theorem (7.14.1). Finally, note that (7.18.6) follows from (7.13.4). □

7.19 Non-Negative Supermartingales

Let $\{X_n, \mathcal{F}_n \mid n \geq 1\}$ *be a non-negative supermartingale. Then* $X_\infty = \lim_{n\to\infty} X_n$ *exists a.s. and*

(7.19.1) $X_{\sigma\wedge\tau} \geq E(X_\tau \mid \mathcal{F}_\sigma) \qquad \forall \tau, \sigma \in \mathcal{S}$

(7.19.2) $X_n \geq E(X_\infty \mid \mathcal{F}_n) \qquad \forall n \geq 1$

(7.19.3) $EX_\infty \leq \lim_{n\to\infty} EX_n$

where $\mathcal{S}$ *is the set of all stopping times. Moreover, if* $EX_1 < \infty$ *and* $\psi : [0,\infty[\to \mathbf{R}$ *is a continuous function such that* $\frac{\psi(t)}{t} \to 0$ *as* $t \to \infty$ *, then we have the following:*

(7.19.4) $E\psi(X_\infty) = \lim_{n\to\infty} E\psi(X_n)$

(7.19.5) *If* $\psi(t) > \psi(0) = 0$ *for all* $t > 0$ *, then* $X_\infty = 0$ *a.s. if and only if* $\lim_{n\to\infty} E\psi(X_n) = 0$

(7.19.6) *If* $0 < q < 1$ *is any given number, then* $X_\infty = 0$ *a.s. if and only if* $\lim_{n\to\infty} EX_n^q = 0$

Proof: Since $X_n^- \equiv 0$ for all $n \geq 1$, we see that any stopping time is optional for $\{X_n^-\}$ and $EX_n^- = 0$. Hence, (7.19.1)–(7.19.3) follows from the optional sampling theorem (7.14) and the martingale convergence theorem (7.10). So suppose that $EX_1 < \infty$, and let $\psi : [0,\infty[\to \mathbf{R}$ be as in the theorem. Since $EX_n \leq EX_1$ and $X_n \xrightarrow{\sim} X_\infty$, then (7.19.4) follows from (5.12.4). Moreover, if $\psi(t) > \psi(0) = 0$ for all $t > 0$, then $X_\infty = 0$ a.s. if and only if $E\psi(X_\infty) = 0$, and so (7.19.5+6) follow from (7.19.4). □

7.20 Convex Functions of Martingales

Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ *be a martingale and let* $\varphi :]a,b[\to \mathbf{R}$ *be a convex function, where* $-\infty \leq a < b \leq \infty$ *and* $P(a \leq X_n \leq b) = 1$ *for all* $n \in T$ *. Let* $Y_n = \varphi(X_n)$ *and suppose* τ *is a stopping time such that* $\varphi(X_\tau) \in L(P)$ *and* τ *is optional for* $\{X_n\}$ *. Then we have the following:*

(7.20.1) $Y_{\tau\wedge\sigma} \leq E(Y_\tau \mid \mathcal{F}_\sigma) \qquad \forall \sigma \in \mathcal{S}$

(7.20.2) *If* $EY_\tau^+ < \infty$ *, then* τ *is optional for* $\{Y_n^+\}$

where $\mathcal{S}$, as usual, denotes the set of all stopping times.

Similarly, suppose that $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a submartingale such that $P(a \le X_n \le b) = 1$ for all $n \in T$, and let τ be a stopping time such that τ is optional for $\{X_n^+\}$. If $Y_n = \varphi(X_n)$ for some increasing convex function $\varphi :]a, b[\to \mathbf{R}$ such that $\varphi(X_\tau) \in L(P)$, then we have the following:

(7.20.3) $\quad Y_{\tau\wedge\sigma} \le E(Y_\tau \mid \mathcal{F}_\sigma) \qquad \forall \sigma \in \mathcal{S}$

(7.20.4) $\quad$ *If $EY_\tau^+ < \infty$, then τ is optional for $\{Y_n^+\}$*

I shall leave the formulation of the supermartingale case to the reader.

Remark: Recall that under the hypothesis of the theorem, we have that $\{Y_n\}$ is a submartingale whenever $EY_n < \infty$ for all $n \in T$ [see (7.5.19)]. Hence, (7.20.1) is a cheap but useful extension of the optional sampling theorem (7.14) because τ may be optional for $\{X_n\}$ without being optional for $\{Y_n\}$.

Proof: (7.20.1+2): By the optional sampling theorem (7.14.3), we have that $X_{\tau\wedge\sigma} = E(X_\tau \mid \mathcal{F}_\sigma)$, but then (7.20.1) follows directly from Jensen's inequality (6.9.3), and (7.20.2) follows from (7.20.1) and (6.10.1).

(7.20.3+4) follows in the same manner. □

7.21 The Ballot Theorem

Let $X_1, \ldots, X_q$ be non-negative integer-valued random variables such that $(X_1, \ldots, X_q)$ are exchangeable [see (6.12)]. If $S_n = X_1 + \cdots + X_n$ for $1 \le n \le q$, then we have

$$(7.21.1) \qquad P\left(\max_{1\le n\le q} \frac{S_n}{n} < 1 \,\middle|\, S_q\right) = P\left(\max_{1\le n\le q} \frac{S_n}{n} < 1 \,\middle|\, \mathcal{S}_X^q\right) = \left(1 - \frac{S_q}{q}\right)^+$$

where $\mathcal{S}_X^1, \ldots, \mathcal{S}_X^q$ are the permutation invariant σ-algebras [see (6.17.8)].

Proof: Let $Y_j = X_j \wedge q$ and $U_n = \frac{1}{n} T_n$ where $T_n = Y_1 + \cdots + Y_n$. Then $Y_1, \ldots, Y_q$ are exchangeable and integrable. Hence, by (6.19.6) we have that $U_n = E(Y_1 \mid \mathcal{S}_X^n)$ for all $1 \le n \le q$. Thus, by (7.6) we have that $\{W_j, \mathcal{G}_j \mid -q \le j \le -1\}$ is a martingale where $W_j = U_{-j}$ and $\mathcal{G}_j = \mathcal{S}_X^{-j}$ for $j = -1, -2, \ldots, -q$. By (7.2.4), we have that

$$\tau := \inf\{-q \le j \le -1 \mid U_{-j} \ge 1\} \qquad (\inf \emptyset = -1)$$

is a stopping time for $\{\mathcal{G}_j \mid -q \le j \le -1\}$. Moreover, if $\xi = -\tau$ and $A = \{\max\{U_1, \ldots, U_q\} < 1\}$, then I claim that for any $1 \le n \le q$ and any $\omega \in \Omega$ we have

(i) $$U_\xi(\omega) = U_q(\omega) - 1_A(\omega) + (1 - U_q(\omega))^+$$

(ii) $$U_n(\omega) < 1 \iff \tfrac{1}{n} S_n(\omega) < 1$$

Proof of (i): Suppose that $\omega \in A$; then $\xi(\omega) = 1$, $U_\xi(\omega) = Y_1(\omega) < 1$ and $U_q(\omega) < 1$. Hence, $U_q(\omega) + (1 - U_q(\omega))^+ = 1 = 1_A(\omega)$ and $U_\xi(\omega) = 0$ because $Y_1(\omega)$ is a non-negative integer. Thus, (i) holds whenever $\omega \in A$. Suppose that $\xi(\omega) = q$, then $U_q(\omega) \ge 1$ and $1_A(\omega) = 0$, so (i) holds whenever $\xi(\omega) = 1$. Finally, suppose that $\omega \notin A$ and $1 \le \xi(\omega) = n < q$; then $U_n(\omega) \ge 1$ and $U_{n+1}(\omega) < 1$. Hence, $n \le T_n(\omega) \le T_{n+1}(\omega) < n + 1$ and so $T_n(\omega) = n$ because $T_n(\omega)$ is an integer. Thus, $U_\xi(\omega) = U_n(\omega) = 1$, and since $U_q(\omega) < 1$ and $1_A(\omega) = 0$, we see that (i) holds for any $\omega \in \Omega$.

Proof of (ii): Suppose that $U_n(\omega) < 1$. Then $Y_j(\omega) < n \le q$ for all $j = 1, \ldots, n$, and so $Y_j(\omega) = X_j(\omega)$ by definition of Y_j . Hence, $U_n(\omega) = \frac{1}{n} S_n(\omega) < 1$. Conversely, by definition we have $U_n(\omega) \le \frac{1}{n} S_n(\omega)$. Thus, the converse implication holds.

Note that τ is optional for $\{W_j\}$ by (7.13.3), and U_q is $\mathcal{S}_X^q$-measurable by (6.17.11). Hence, by (i) and the optional sampling theorem (7.14.3) we get

$$\begin{aligned} P(A \mid \mathcal{S}_X^q) &= E(U_q - U_\xi + (1 - U_q)^+ \mid \mathcal{S}_X^q) \\ &= U_q + (1 - U_q)^+ - E(W_\tau \mid \mathcal{S}_X^q) \\ &= U_q + (1 - U_q)^+ - W_{-q} = (1 - U_q)^+ \end{aligned}$$

and by (ii) we get

$$A = \left\{ \max_{1 \le n \le q} \frac{S_n}{n} < 1 \right\} \quad , \quad (1 - U_q)^+ = \left(1 - \frac{S_q}{q}\right)^+$$

Thus, the last equality in (7.2.1) holds, and since $\sigma(S_q) \subseteq \mathcal{S}_X^q$ and $(1 - U_q)^+$ is $\sigma(S_q)$-measurable, we see that the first equality follows from (6.8.9). □

Remark: Equation (7.21.1) is called *the ballot theorem* for the following reason: Suppose that our two good friends, Mike and Jane have joined an electoral campaign. The poll is over, and they are sitting and watching the count of the votes, when Jane poses the following question to Mike: "Suppose that I finally, get k more votes than you. What is the probability that I will lead ahead of you all the time during the count?" The answer is given by the ballot theorem, and the correct answer is $\frac{k}{q}$, where q is the total number of voters. Let us look at a model:

Let J_n and M_n denote the number of votes for Jane and Mike, respectively, when the n first votes are counted. Then Jane is asking for the following probability:

$$p = P(M_n < J_n \quad \forall 1 \le n \le q \mid J_q = k + M_q)$$

Let us define X_j as follows:

$$X_j = \begin{cases} 0 & \text{if vote number } j \text{ is for Jane} \\ 2 & \text{if vote number } j \text{ is for Mike} \\ 1 & \text{otherwise} \end{cases}$$

It is then easily checked that

$$p = P\left(\max_{1 \le n \le q} \frac{S_n}{n} < 1 \,\middle|\, S_q = q - k \right)$$

where $S_n = X_1 + \cdots X_n = n + M_n - J_n$. Thus, if we suppose that the votes are counted in a random order such that $(X_1, \ldots, X_q)$ are exchangeable random variables, then by (7.21.1) we have $p = \frac{k}{q}$.

For instance, if there are 2 million voters, then Jane should get at least 1.9 million more votes than Mike in order to have 95% chance of leading ahead of Mike all the time during the count. This astonishing fact tells a little about how difficult it is to make a reliable Gallup poll. It is also surprising that the probability p only depends on k and the total number of voters including those voters who have not voted for any of the two.

Actually, Jane's question was first posed by a Danish mathematician and politician Carl Cristopher Georg Andræ (1812–1893). Andræ invented several methods of election. As a matter of fact, there is a good deal of non-trivial mathematics behind most of the election methods in use.

7.22 Rescaling of Scaled Martingales

Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter and let $\{X_n \mid n \in T\}$ be a sequence of random variables. Then we say that $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a *scaled submartingale* if there exist numbers $\alpha_n \in \mathbf{R}$ and $\beta_n > 0$ such that [cf. (7.5.1) and (7.5.2)]

(7.22.1) $\quad X_n$ is $\mathcal{F}_n$ -measurable and $EX_n^+ < \infty \quad \forall n \in T$

(7.22.2) $\quad \alpha_n + \beta_n X_n \le E(X_{n+1} \mid \mathcal{F}_n) \qquad$ if $n \in T$ and $n + 1 \in T$

We define *scaled supermartingales* and *scaled martingales* similarly. The parameters $\{\alpha_n\}$ are called *the position parameters* and may be interpreted as a deterministic trend; the parameters $\{\beta_n\}$ are called *the scale parameters* and may be interpreted as a deterministic growth rate. In (7.6), we saw that sums and products of independent random variables are scaled martingales. Any scaled submartingale, supermartingale or martingale can be rescaled to a proper submartingale, supermartingale or martingale, respectively, as follows. Let $A_n, B_n \in \mathbf{R}$ be real numbers satisfying

(7.22.3) $\quad B_n > 0 \quad \forall n \in T \quad$ and $\quad B_n = \beta_n B_{n+1} \quad \forall n \in T \quad$ with $\quad n+1 \in T$

(7.22.4) $\quad A_{n+1} = \beta_n A_n + \alpha_n \quad \forall n \in T \quad$ with $\quad n+1 \in T$

Note that if $k \in T$ is fixed and $A_k \in \mathbf{R}$ and $B_k > 0$ are given real numbers, then (7.22.3+4) recursively defines uniquely A_n and B_n for all $n \in T$. If A_n and B_n satisfies (7.22.3+4) and

(7.22.5) $\quad Y_n = B_n(X_n - A_n) \qquad \forall n \in T$

then $Y_n = B_{n+1}(\beta_n X_n + \alpha_n - A_{n+1})$ and

$$E(Y_{n+1} \mid \mathcal{F}_n) = B_{n+1}(E(X_{n+1} \mid \mathcal{F}_n) - A_{n+1})$$

Thus, we get:

(7.22.6) Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a scaled submartingale, or a supermartingale, or a martingale with position parameters $\alpha_n \in \mathbf{R}$ and scale parameters $\beta_n > 0$. If A_n and B_n satisfies (7.22.3+4), then $\{B_n(X_n - A_n), \mathcal{F}_n \mid n \in T\}$ is a submartingale, or a supermartingale, or a martingale

Moreover, we have the following:

(7.22.7) Suppose that $\{X_n, \mathcal{F}_n \mid n \in T\}$ is a non-negative scaled supermartingale with position parameters $\alpha_n \geq 0$ and scale parameters $\beta_n > 0$. If $\varphi : [0,\infty[\to [0,\infty[$ is an increasing concave function, then $\{\varphi(X_n), \mathcal{F}_n \mid n \in T\}$ is a scaled supermartingale with position parameters $\hat{\alpha}_n = \varphi(\alpha_n)$ and scale parameters $\hat{\beta}_n = \beta_n \vee 1$ for $n \in T$

To see this, let $\alpha, \gamma, y \in \mathbf{R}_+$ be given such that $\gamma \geq 1$. Since $\varphi(0) \geq 0$ and φ is concave, we have

$$\varphi(\alpha) \geq \frac{\alpha}{\alpha+\gamma y}\varphi(\alpha+\gamma y) + \frac{\gamma y}{\alpha+\gamma y}\varphi(0) \geq \frac{\alpha}{\alpha+\gamma y}\varphi(\alpha+\gamma y)$$
$$\gamma\varphi(y) \geq \frac{\gamma y}{\alpha+\gamma y}\varphi(\alpha+\gamma y) + \frac{\alpha+(\gamma-1)y}{\alpha+\gamma y}\gamma\varphi(0) \geq \frac{\gamma y}{\alpha+\gamma y}\varphi(\alpha+\gamma y)$$

Adding these inequalities, we find $\varphi(\alpha) + \gamma\,\varphi(y) \geq \varphi(\alpha + \gamma\, y)$ for all $\alpha \geq 0$ all $y \geq 0$ and all $\gamma \geq 1$. Moreover, since $\alpha_n + \beta_n X_n \leq \alpha_n + \hat{\beta}_n X_n$ and φ is increasing and concave, then by Jensen's equality (6.9.3) we find

$$\begin{aligned} E(\varphi(X_{n+1}) \mid \mathcal{F}_n) &\leq \varphi(E(X_{n+1} \mid \mathcal{F}_n)) \leq \varphi(\alpha_n + \beta_n X_n) \\ &\leq \varphi(\alpha_n + \hat{\beta}_n X_n) \leq \varphi(\alpha_n) + \hat{\beta}_n\,\varphi(X_n) \end{aligned}$$

Hence, (7.22.7) follows.

7.23 The Most Favorable Opting Time

Let $\{\mathcal{F}_n \mid n \geq 0\}$ be a filter and let $\{S_n \mid n \geq 0\}$ be non-negative random variables such that S_n is $\mathcal{F}_n$-measurable for all $n \geq 0$. Let $p, a \geq 0$ be given non-negative numbers and let $\varphi : [0, \infty[$ be an increasing convex function satisfying

(7.23.1) $\quad E(\varphi(S_{n+1}) \mid \mathcal{F}_n) \leq (1+p)\varphi(S_n) \qquad \forall n \geq 0$

(7.23.2) $\quad \varphi(a) > 0$ *and* $\varphi'(t)$ *is continuous and strictly positive on* $]a, \infty[$

(7.23.3) $$\int_0^\infty \left(1 - \frac{\varphi'(t)}{b}\right) dt > \frac{\varphi(0)}{b} + a$$

where $b = \lim_{n\to\infty} \varphi'(t) = \sup_{t\geq 0} \varphi'(t)$. *Then the equation:*

(7.23.4) $$x - \frac{\varphi(x)}{\varphi'(x)} = a$$

admits a minimal solution c satisfying $c > a$. Moreover, if we define

(7.23.5) $\quad V_n = (1+p)^{-n}(S_n - a)^+ \quad , \quad V_\infty = \limsup_{n\to\infty} V_n$

(7.23.6) $\quad \sigma = \inf\{n \geq 0 \mid S_n \geq c\} \qquad (\inf \emptyset = \infty)$

$$(7.23.7) \qquad \psi(t) = \begin{cases} \dfrac{\varphi(t)}{\varphi'(c)} & \text{if } 0 \le t \le c \\ t - a & \text{if } t \ge c \end{cases}$$

then σ *is a stopping time and we have*

$$(7.23.8) \qquad EV_\tau \le E\psi(S_0) \quad \text{and} \quad E(V_\tau \mid \mathcal{F}_0) \le \psi(S_0) \qquad \forall \tau \in \mathcal{S}$$

$$(7.23.9) \qquad E(V_\tau \mid \mathcal{F}_\sigma) \le V_{\tau \wedge \sigma} \qquad \forall \tau \in \mathcal{S}$$

where $\mathcal{S}$ *denotes the set of all stopping times for the filter* $\{\mathcal{F}_n \mid n \ge 0\}$.

Remarks: (1): In the financial world the word option has a special meaning. A company may offer you *an option contract* on a certain amount of shares of stocks, say 100 shares, for a certain period, say five years. This means that at any time within the next five years you may purchase the asset at a fixed stated price regardless of the current value of the stocks on the stock market (in the language of finance: you *opt your asset*) or you may choose not to purchase at all. Since only a fool will opt the asset at a time when the stated price is below the current market value, you can never lose on an option contract. A company which needs capital, can stimulate the sales of their stock by offering an option on a certain number of shares per share that you buy now. Suppose that you hold an option contract, then you have a problem of finding the most favorable time to opt the asset. To solve this problem, let us make a model:

(2): Let S_n denote the value of the asset on the market and let $\mathcal{F}_n$ be your available information at time $n \ge 0$. Let us suppose that you know the market price of the asset at every time n , that is, S_n is $\mathcal{F}_n$-measurable for all $n \ge 0$. Let $a \ge 0$ be the stated price at which you have the option of buying the asset, and let $p \ge 0$ be the market interest. If $S_n \ge a$, then you can make a profit of $S_n - a$ by opting the asset at time n . If $S_n < a$, then it is foolish to opt the asset. Thus, $(S_n - a)^+$ represents the value of the option at time n . Discounting this to time 0 with the market rate p , we see that if you opt the asset at time n , then V_n as given by (7.23.5) is the value of the option (in time 0 money). Being non-clairvoyant the most favorable opting time must be found among the stopping times for $\{\mathcal{F}_n\}$. Hence, the optimal value of the option equals $v = \sup\{EV_\tau \mid \tau \in \mathcal{S}\}$. We are searching that (or those) stopping time(s) η satisfying $EV_\eta = v$ if such stopping time exists at all. Suppose that (7.23.1)–(7.23.3) hold for some given increasing convex function φ . Note that (7.23.1) states that you expect that the value of the asset grows slower than the market rate. The stopping time σ given by (7.23.6), may not be the optimal opting time, but (7.23.9) shows that the optimal opting time (if it exists) must be $\le \sigma$. In particular, if $S_0(\omega) \ge c$, then $\sigma(\omega) = 0$ and the best you can do is to opt the asset immediately.

(3): If $\varphi(t) = t^q$ for some $q > 1$, then φ is an increasing convex function satisfying (7.23.2+3) for any $a > 0$, and $c = a + \frac{a}{q-1}$ is the only solution to (7.23.4). Similarly, if $\varphi(t) = e^{\lambda t}$ for some $\lambda > 0$, then φ is an increasing convex function satisfying (7.23.2+3) for any $a \geq 0$, and $c = a + \frac{1}{\lambda}$ is the only solution to (7.23.4). Actually if φ is strictly convex on $]a, \infty[$, then it follows easily that (7.23.4) admits one and only one solution. Moreover, note that if $b = \infty$, then (7.23.3) holds trivially because the left-hand side equals ∞ and the right-hand side equals a .

Proof: Since φ is convex and increasing, we have that φ' is increasing and non-negative. Hence, if we define

$$\xi(t) := t - \frac{\varphi(t)}{\varphi'(t)} = \int_0^t \left(1 - \frac{\varphi'(u)}{\varphi'(t)}\right) du - \frac{\varphi(0)}{\varphi'(t)} \qquad \forall t \geq 0$$

then ξ is increasing on $[0, \infty[$ and continuous on $]a, \infty[$. Since $\varphi(a) > 0$, we have $\xi(a) < a$ and by (7.23.3), we have $\lim_{t\to\infty} \xi(t) > a$. Thus, the set of solutions to equation (7.23.4) is a closed non-empty subset of $]a, \infty[$. Hence, the minimal solution c exists and satisfies $c > a$. Note that by (7.23.2) we have that φ is strictly increasing on $[a, \infty[$, so let $\tilde{\varphi} : [\varphi(a), \infty[\to [a, \infty[$ denote the inverse function of φ restricted to $[a, \infty[$. Then it follows easily that

$$\kappa(s) = \begin{cases} \tilde{\varphi}(s) - a & \text{if } s \geq \varphi'(c) \\ \dfrac{s}{\varphi'(s)} & \text{if } 0 \leq s \leq \varphi'(c) \end{cases}$$

is a concave increasing function satisfying: $\kappa(\varphi(t)) = \psi(t)$ for all $t \geq 0$ where $\psi(t)$ is given by (7.23.7). Moreover, by convexity of φ we have

$$\frac{\varphi(t)}{\varphi'(c)} \geq \frac{\varphi'(c)\,(t-c) + \varphi(c)}{\varphi'(c)} = t - \xi(c) = t - a$$

Hence, we conclude that $\psi(t) \geq (t-a)^+$ for all $t \geq 0$.

By (7.23.1), we have that $\{\varphi(X_n)\}$ is a scaled supermartingale with position parameters $\alpha_n = 0$ and scale parameters $\beta_n = 1 + p$. By the argument above, we have that $\psi(t) = \kappa(\varphi(t))$ and that κ is an increasing concave function. Hence, by (7.22.7) we have that $\{\psi(S_n)\}$ is a scaled supermartingale with position parameters $\hat{\alpha}_n = \psi(0) = 0$ and scale parameters $\hat{\beta}_n = 1 + p$. Thus, by rescaling [see (7.22.8)], we have that $T_n := (1+p)^{-n}\,\psi(S_n)$ is a non-negative supermartingale, and by (7.19) and the inequality: $\psi(t) \geq (t-a)^+$ we see that $T_\infty := \lim T_n$ exists a.s. and $T_n \geq V_n$ for all $0 \leq n \leq \infty$. Let

$\tau \in \mathcal{S}$ be given and set $A = \{\sigma < \tau\}$ and $B = \{\sigma \geq \tau\} = A^c$. If $\omega \in A$, then $\sigma(\omega) = \tau(\omega) \wedge \sigma(\omega) < \infty$ and $S_\sigma(\omega) \geq c > a$. Hence, by (7.23.5+6) we get

$$1_A T_{\tau\wedge\sigma} = 1_A (1+p)^{-n} \psi(S_{\tau\wedge\sigma}) = 1_A (1+p)^{-n} (S_{\tau\wedge\sigma} - a) = 1_A V_{\tau\wedge\sigma}$$

By (7.3.2+5) and (7.4.1), we have that $1_B V_\tau = 1_B V_{\tau\wedge\sigma}$ is $\mathcal{F}_\sigma$-measurable, and since $V_\tau \leq T_\tau$ and $A \in \mathcal{F}_\sigma$, then by (7.19.1) and (6.8) we get

$$E(V_\tau \mid \mathcal{F}_\rho) \leq E(T_\tau \mid \mathcal{F}_\rho) \leq T_{\tau\wedge\rho} \quad \forall \rho \in \mathcal{S}$$

$$\begin{aligned} E(V_\tau \mid \mathcal{F}_\sigma) &= E(1_A V_\tau \mid \mathcal{F}_\sigma) + E(1_B V_\tau \mid \mathcal{F}_\sigma) \leq 1_A E(T_\tau \mid \mathcal{F}_\sigma) + 1_B V_{\tau\wedge\sigma} \\ &\leq 1_B T_{\tau\wedge\sigma} + 1_B V_{\tau\wedge\sigma} = V_{\tau\wedge\sigma} \end{aligned}$$

Thus, (7.23.9) follows from the second inequality, and since $T_0 = \psi(S_0)$, we see that (7.23.8) follows from the first inequality with $\rho = 0$. □

7.24 An Epidemic Model

Let $\{X_0, \xi_{jn} \mid j \geq 1, n \geq 0\} \subseteq L^2(P)$ be independent, non-negative, integer-valued random variables and suppose that there exist $\mu > 0$ and $\sigma^2 > 0$ such that

(7.24.1) $\quad E\xi_{jn} = \mu$ *and* $\operatorname{var}(\xi_{jn}) = \sigma^2 \quad \forall j \geq 1\ \forall n \geq 0$

Let $\mathcal{F}_0 = \sigma(X_0)$ and $\mathcal{F}_n = \sigma\{X_0, \xi_{jk} \mid j \geq 1,\ 0 \leq k \leq n-1\}$ for $n \geq 1$. Moreover, for $n = 0, 1, 2, \ldots$ we define

(7.24.2) $$m := EX_0 \quad , \quad X_{n+1} := \sum_{1 \leq j \leq X_n} \xi_{jn} = \sum_{k=1}^{\infty} \sum_{j=1}^{k} 1_{\{X_n = k\}} \xi_{jn}$$

(7.24.3) $$S_n := \sum_{i=0}^{n} X_i \quad , \quad S := \sum_{i=0}^{\infty} X_i \quad , \quad M_n := \mu^{-n} X_n$$

(7.24.4) $$\tau := \inf\{n \geq 0 \mid X_n = 0\} \quad (\inf \emptyset := \infty)$$

(7.24.5) $$\alpha_n := \inf_{j \geq 1} P(\xi_{jn} = 0) \quad , \quad \beta_n := \sup_{j \geq 1} P(\xi_{jn} = 0) \quad , \quad \alpha := \inf_{n \geq 0} \alpha_n$$

and we let $q_n(s)$ denote the generating function for X_n. Then τ is a stopping time for $\{\mathcal{F}_n \mid n \geq 0\}$ such that $P(\tau = 0) = P(X_0 = 0)$, and $\{M_n, \mathcal{F}_n \mid n \geq 0\}$ is a non-negative martingale such that the limit $M_\infty = \lim_{n\to\infty} M_n$ exists a.s. and $EM_\infty \leq EM_n = m$ for all $n \geq 0$. Moreover, for every $n \geq 0$ we have

(7.24.6) $$E(X_{n+1} \mid \mathcal{F}_n) = \mu X_n \quad , \quad E(X_{n+1}^2 \mid \mathcal{F}_n) = \sigma^2 X_n + \mu^2 X_n^2$$

$$(7.24.7) \qquad EX_n = m\,\mu^n \quad , \quad ES = \begin{cases} \frac{m}{\mu-1} & \text{if } \mu < 1 \\ \infty & \text{if } \mu \geq 1 \text{ and } m > 0 \end{cases}$$

$$(7.24.8) \qquad EX_n^2 = \begin{cases} \mu^{2n}\,EX_0^2 + \dfrac{m\,\sigma^2\left(\mu^{2n-1} - \mu^{n-1}\right)}{\mu - 1} & \text{if } \mu \neq 1 \\ n\,m\,\sigma^2 + EX_0^2 & \text{if } \mu = 1 \end{cases}$$

$$(7.24.9) \qquad P(\tau < \infty) = P(S < \infty) = P\Big(\lim_{n\to\infty} X_n = 0\Big) = P(\exists\, n \,:\, X_n = 0)$$

$$(7.24.10) \qquad q_n(\alpha_n) \leq P(\tau \leq n+1) = P(X_{n+1} = 0) \leq q_n(\beta_n)$$

In addition, we have the following:

The subcritical case: *Suppose that* $\mu < 1$. *Then* $\tau < \infty$ *a.s.,* $X_n \to 0$ *a.s. and* $ES = \frac{m}{1-\mu} < \infty$.

The critical case: *Suppose that* $\mu = 1$. *Then* $ES = \infty$ *and the limit* $X_\infty = \lim_{n\to\infty} X_n$ *exist a.s. and satisfies* $EX_\infty \leq m$. *If, moreover,* $\alpha > 0$, *then* $\tau < \infty$ *a.s.,* $X_n \to 0$ *a.s. and* $S < \infty$ *a.s.*

The supercritical case: *Suppose that* $\mu > 1$. *Then* $X_n(\omega) \to \infty$ *for a.a.* $\omega \in \{M_\infty > 0\}$ *and we have*

$$(7.24.11) \qquad P(M_\infty > 0) \leq P(\tau = \infty) = P(S = \infty) \leq 1 - \sup_{n\geq 0} q_n(\alpha_n)$$

$$(7.24.12) \qquad EM_\infty = m \quad , \quad \mathrm{var}(M_\infty) = \mathrm{var}(X_0) + \frac{m\,\sigma^2}{\mu(\mu-1)}$$

Remarks: (1): Suppose that an epidemic has broken out. Let us say that at time 0 there are X_0 infected persons. Each of these will infect a random number of new persons, and each of these infects a random number of new persons, etc. What happens? Will the epidemic explode and finally, wipe out the whole population, or will the epidemic finally, die out? Let us look at a model: At the start of the epidemic, there are X_0 infected persons [the 0th generation], which we may enumerate: $1, 2, \ldots, X_0$. Let ξ_{j0} be the random number that Mr. j of the 0th generation will infect. Then the total number of persons in the first generation, that is, the persons infected by the 0th generation, equals X_1 as given by (7.24.2). Let us enumerate the first generation: $1, 2, \cdots, X_1$, and let ξ_{j1} be the number of persons, that Mrs. j of the first generation will infect. Then the total number of persons in the second generation equals X_2 as given by (7.24.2), etc. Thus, if X_n is given by (7.24.2), then X_n the total number of persons in the nth generation of the epidemic and S is the total number of infected during the whole epidemic. Note that if $X_n(\omega) = 0$, then $X_k(\omega) = 0$ for all $k \geq n$. Hence, if the epidemic has died out, then it will never revive again. Moreover, the stopping time τ given by (7.24.4) is the first time that the epidemic dies out if

this ever happens, and $\{\tau = \infty\}$ is the event that the epidemic goes on forever. Note that μ is the average number that an infected person will infect, and for this reason the number μ is called *the infection rate.*

(2): In the subcritical case we have that an infected person on average will infect strictly less than 1 person, and so it is no wonder that the epidemic eventually dies out. As a matter of fact, since $M_\infty < \infty$ a.s. and $X_n = \mu^n M_n$, we see that X_n decreases exponentially fast to zero. In the critical case we have that an infected person on average will infect exactly 1 person, and so one should expect that the epidemic could survive forever; but if $\alpha > 0$ [for instance, if the ξ_{jn} th all have the same distribution and $P(\xi_{jn} = 0) > 0$], then the theorem above shows that this is **not** the case and the epidemic dies out with probability 1 . However, even though the total number of infected persons is finite a.s., the expected number of infected persons is infinite. In the supercritical case we have that an infected person in average will infect strictly more than 1 person, and so it is no wonder that the epidemic may explode and that $X_n = \mu^n M_\infty$ grows exponentially fast to infinity for all $\omega \in \{M_\infty > 0\}$. However, it is surprising that the epidemic may die out with positive probability, which by (7.24.11) will be the case if just one of the α_n th are strictly positive.

(3): In 1964 we had the last outbreak of smallpox in Denmark. At that time the infection rate μ was estimated to $\mu = 0.98 \pm 0.04$. Thus, in spite of a very high vaccination rate, smallpox is just around the critical case. Let me point out that the epidemic model described above is the simplest possible random model, and we have discounted any particular knowledge about the disease, such as incubation time, immunity, intermediate hosts, geographical diversity, etc. For instance, the incubation time seems to present the most serious problem in the attempt to make a reliable model for the AIDS-epidemic. The incubation time for AIDS is very long and with a large variance. Many distributions have been suggested, but at the present time (1991), we are far from a consensus about the distribution of the incubation time.

(4): The model described in the theorem above contains many implicit assumptions which may not be fulfilled in real life. For instance, it is assumed that there at any time is an unlimited number of persons who can be infected. This assumption is reasonable if the epidemic only attacks a small proportion of the total population. However, for epidemics like the plague, which attacks a substantial part of the population, this will not be so. In Table 7.1 you will find data from the last plague epidemic in Denmark in 1711. From fall in 1710 to the end of 1711 we had the last outbreak of plague in Copenhagen. In that year there were 22.535 deaths recorded out of a total population of approximately 60.000 (before the plague). Table 7.1 contains the weekly number of newborn babies and deaths as recorded by a Danish newspaper from January 1712, and in a report from the Chief of Police to the Danish king Christian VI we know the number of deaths

Table 7.1

Weekly mortality in Copenhagen 1711 as repoted by a Danish newspaper January 1712								
Week	Newborn	Deaths	Week	Newborn	Deaths	Week	Newborn	Deaths
1	43	84	19	41	55	36	24	2326
2	35	86	20	41	55	37	20	1751
3	24	61	21	34	54	38	21	1312
4	34	80	22	27	50	39	10	974
5	39	61	23	40	54	40	13	613
6	48	62	24	27	65	41	15	369
7	44	63	25	36	95	42	12	241
8	29	72	26	33	126	43	11	176
9	33	68	27	34	228	44	14	92
10	31	68	28	36	342	45	15	84
11	17	55	29	29	748	46	11	60
12	29	46	30	32	1032	47	20	64
13	32	49	31	41	1951	48	22	51
14	35	45	32	38	1831	49	19	47
15	39	52	33	35	1949	50	12	41
16	36	65	34	37	2393	51	20	28
17	30	50	35	27	2119	52	11	22
18	35	57				*Total*	*1464*	*22535*

and newborn in all the dioceses of Denmark in the year 1711. For instance, in the diocese of Aarhus there were 3.111 newborn and 2.237 deaths recorded. Since the number of newborn in Copenhagen in 1711 was 1.464, the normal mortality should be approximative 1.000; thus, by far most of the 22.535 deaths were caused by the plague. Note that the number deaths in the winter and spring time was approximately twice the number of newborns. But in week 26 (June 28 to July 4) the epidemic started to explode and the weekly mortality doubled until the epi-

demic reached its maximum in week 34 (August 23 to August 30) with a weekly death toll of more than twice the normal yearly mortality in Copenhagen at that time. There are two kinds of plague: pneumonic and bubonic, both are highly lethal, but the bubonic plague is much more infectious than the pneumonic plague. As normal for winter time, the plague epidemic in Copenhagen started with pneumonic plague and changed to bubonic plague at summer time, which may explain the sudden increase at the end of June. Note that from week 26 to week 34 the epidemic shows an exponential increase, which is in accordance with the theory, but then it starts to decrease exponentially, which seems to contradict the theory. However, in the model described above we have assumed that there is an unlimited number of persons, who can be infected – an assumption which certainly was not fulfilled in Copenhagen in 1711 – and the point is that due to the few uninfected but susceptible persons left the infection rate μ changes from being > 1 to be < 1 , which explains the exponential decrease in the fall of 1711. This feature is typical for highly infectious epidemics, such as plague. At the beginning of such an epidemic, the assumptions of the above theorem are reasonably satisfied, but at a certain time there will be so few persons left to be infected that the infection rate μ will decrease. Consequently, a highly infectious epidemic will show the following picture: At the beginning it grows exponentially, then after some time it flattens out, and finally, it starts to decrease exponentially.

Proof: Evidently, we have that $\{\mathcal{F}_n \mid n \geq 0\}$ is a filter, and an easy induction argument shows that X_n is $\mathcal{F}_n$-measurable, non-negative and integer-valued. Hence, by (7.2.4) and (7.2.6) we have that τ is a stopping time such that $P(\tau = 0) = P(X_0 = 0)$. Let $\Xi_{kn} = \xi_{1n} + \cdots + \xi_{kn}$ if $\varphi : \mathbf{R}_+ \to \mathbf{R}_+$ is an arbitrary function, then we define $\varphi^{<n>} : \mathbf{N}_0 \to \mathbf{R}_+$ as follows:

(i) $\varphi^{<n>}(0) = \varphi(0) \quad , \quad \varphi^{<n>}(k) = E\varphi(\Xi_{kn}) \quad \forall k \geq 1 \; \forall n \geq 0$

Since Ξ_{kn} is independent of $\mathcal{F}_n$ and X_n is $\mathcal{F}_n$-measurable, then by (6.8) and (6.10.6) we find

$$E(\varphi(X_{n+1}) \mid \mathcal{F}_n) = E\left(1_{\{X_n=0\}}\,\varphi(0) + \sum_{k=1}^{\infty} 1_{\{X_n=k\}}\,\varphi(\Xi_{kn}) \;\middle|\; \mathcal{F}_n\right)$$
$$= 1_{\{X_n=0\}}\,\varphi(0) + \sum_{k=1}^{\infty} 1_{\{X_n=k\}}\,\varphi^{<n>}(k) = \varphi^{<n>}(X_n)$$

Thus, by (6.8.4) we get

(ii) $E(\varphi(X_{n+1}) \mid \mathcal{F}_n) = \varphi^{<n>}(X_n) \qquad \forall n \geq 0$

(iii) $E\varphi(X_{n+1}) = E\varphi^{<n>}(X_n) \qquad \forall n \geq 0$

(7.24.6): Setting $\varphi(t) = t$, we find that $\varphi^{<n>}(k) = k\mu$, and setting $\varphi(t) = t^2$, we find that $\varphi^{<n>}(0) = 0$ and [see (4.4.3) and (4.4.18)]

$$\varphi^{<n>}(k) = E(\Xi_{kn}^2) = \mathrm{var}(\Xi_{kn}) + (E\Xi_{kn})^2 = k\sigma^2 + \mu^2 k^2$$

Hence, (7.24.6) follows from (ii)

Note that (7.24.6) shows that $\{X_n\}$ is a scaled martingale. Thus, by rescaling [see (7.10.8)] we see that $\{M_n, \mathcal{F}_n \mid n \geq 0\}$ is a non-negative martingale with $EM_n = EM_0 = EX_0 = m$ [see (7.5.12)]. Hence, by (7.19.3) we have that $M_\infty = \lim M_n$ exists a.s. and $EM_\infty \leq m$.

(7.24.7+8): By (7.24.6), we have that $EX_{n+1} = \mu EX_n$ and $EX_{n+1}^2 = \sigma^2 EX_n + \mu^2 EX_n^2$. Hence, (7.24.7+8) follows by induction in n and Beppo Levi's theorem (3.7).

(7.24.9+10): Note that if $X_n(\omega) = 0$, then $X_k(\omega) = 0$ for all $k \geq n$, and since X_n is integer-valued we find

$$\{\tau \leq n\} = \{X_n = 0\} \quad , \quad \{\tau < \infty\} = \bigcup_{n=0}^{\infty} \{X_n = 0\} = \{S < \infty\}$$

Hence, (7.24.9) holds and $P(\tau \leq n) = P(X_n = 0)$ for all $n \geq 0$. Setting $\varphi = 1_{\{0\}}$ gives $\varphi^{<n>}(0) = 0$ and

$$\varphi^{<n>}(k) = P(\Xi_{kn} = 0) = \prod_{j=1}^{k} P(\xi_{jn} = 0)$$

because $\xi_{1n}, \xi_{2n}, \ldots$ are independent and non-negative. Hence, by (iii) we get

$$\text{(iv)} \qquad P(X_{n+1} = 0) = P(X_n = 0) + \sum_{k=1}^{\infty} P(X_n = k) \prod_{j=1}^{k} P(\xi_{jn} = 0)$$

Thus, the first and last inequality in (7.24.11) follows.

The subcritical case: Since $M_\infty < \infty$ a.s. and $\mu < 1$, we have that $X_n = \mu^n M_n \to 0$ a.s. Hence, by (7.24.7) and (7.24.9) we have $ES = \frac{m}{1-\mu} < \infty$ and $\tau < \infty$ a.s.

The critical case: In this case, we have $X_n = M_n$ and $X_n \to M_\infty = X_\infty$ a.s. and $EX_\infty \leq m$. Let $q_\infty(s)$ be the generating function for X_∞ . Then q_∞ is increasing on $[0,1]$, and q_∞ is strictly increasing on $[0,1]$ if and only if $P(X_\infty \neq 0) > 0$. Moreover, by (5.11.3), (5.15.2) and (7.24.10) we get

$$P(X_\infty = 0) = q_\infty(0) = \lim_{n\to\infty} P(X_{n+1} = 0) \geq \lim_{n\to\infty} q_n(\alpha) = q_\infty(\alpha)$$

Hence, if $\alpha > 0$, then $X_\infty = 0$ a.s., and so $\tau < \infty$ and $S < \infty$ a.s. by (7.24.9).

The supercritical case: Since $\mu > 1$ and $X_n = \mu^n M_n$, we see that $X_n(\omega) \to \infty$ for a.a $\omega \in \{M_\infty > 0\}$, and so (7.24.11) follows from (7.24.9+10). By (7.24.8) we have

$$EM_n^2 = \mu^{-2n} EX_n^2 \to \operatorname{var}(X_0) + m^2 + \frac{m\sigma^2}{\mu(\mu-1)}$$

Hence, by (3.24.5) we have that $\{M_n\}$ is uniformly integrable, and by Fatou's lemma we have that $EM_\infty^2 < \infty$. Thus, applying (7.20.2) on $\{M_n\}$ with $\varphi(t) = t^2$ and $\tau \equiv \infty$, we see that $\{M_n^2\}$ is uniformly integrable. So by Dunford-Pettis' theorem (3.23.4) we have

$$EM_\infty = \lim_{n\to\infty} EM_n \quad , \quad EM_\infty^2 = \lim_{n\to\infty} EM_n^2$$

Thus, (7.24.12) follows. □

7.25 Gaussian Martingales

Let $(X_0, X_1, X_2, \ldots)$ be a Gaussian sequence; that is, $(X_0, \ldots, X_n)$ has a normal distribution for all $n \geq 0$, and let $\mu_n = EX_n$, $\sigma_{ij} = \operatorname{cov}(X_i, X_j)$ and $\sigma_i^2 = \sigma_{ii} = \operatorname{var}(X_i)$. If $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ for $n \geq 0$, then the following four statements are equivalent:

(7.25.1) $\{X_n, \mathcal{F}_n \mid n \geq 0\}$ *is a martingale*

(7.25.2) $\mu_n = \mu_0$ *and* $\sigma_{k\wedge n}^2 = \sigma_{kn} \quad \forall k, n \geq 1$

(7.25.3) *There exist independent normally distributed random variables $W_0, W_1, W_2, \ldots$ with $X_n = \mu_0 + W_0 + \cdots + W_n$ and $EW_n = 0$ for all $n \geq 0$*

(7.25.4) *There exist uncorrelated random variables $V_0, V_1, V_2, \ldots$ such that $X_n = \mu_0 + V_0 + \cdots + V_n$ and $EV_n = 0$ for all $n \geq 0$*

Proof: (7.25.1) $\Rightarrow$ (7.25.2): Follows from (7.5.12) and Exc.7.2.

(7.25.2) $\Rightarrow$ (7.25.3): Let $W_0 = X_0 - \mu_0$ and $W_n = X_n - X_{n-1}$ for all $n \geq 1$, then $X_n = \mu_0 + W_0 + \cdots + W_n$, and by (4.22.16) we have that

$(W_0, \dots, W_n) \sim N_n(0, \Sigma_n)$, where

$$\Sigma_n = \begin{Bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{Bmatrix} \{\sigma^2_{i \wedge j}\} \begin{Bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{Bmatrix}$$

$$= \begin{Bmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 - \sigma_0^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 - \sigma_{n-1}^2 \end{Bmatrix}$$

Hence, by (4.22.14) we conclude that $W_0, W_1, \dots$ are independent normally distributed random variables with mean 0 .

(7.25.3) $\Rightarrow$ (7.26.4): Evident!

(7.25.4) $\Rightarrow$ (7.25.1): Since $V_0 = X_0 - \mu_0$ and $V_n = X_n - X_{n-1}$ we see that $V_0, V_1, V_2, \dots$ is a Gaussian sequence. Hence, by (4.22.14) we have that $V_0, V_1, V_2, \dots$ are independent, and so (7.25.1) follows from (7.6). □

7.26 A Water Reservoir Model

Let $\{\mathcal{F}_n \mid n \geq 0\}$ be a filter, let X_n be an $\mathcal{F}_n$-measurable random variable for all $n \geq 0$, and let $0 \leq a < b \leq \infty$ such that $a \leq X_0 \leq b$. Let us define W_n by the recursive equations:

(7.26.1) $\quad W_0 = X_0 \quad , \quad W_{n+1} = (W_n + X_{n+1})^+ \wedge b$

Then W_n is $\mathcal{F}_n$-measurable for all $n \geq 0$, and

$$\tau = \inf\{n \geq 0 \mid W_n \leq a\} \qquad (\inf \emptyset = \infty)$$

is a stopping time. Suppose, moreover, that $\mu > 0$, $c \in \mathbf{R}$ and φ is a Borel function from $\overline{\mathbf{R}}$ into $\overline{\mathbf{R}}$ satisfying

(7.26.2) $\quad c \leq \varphi(x) \leq \varphi(a) < \infty \qquad \forall x \in [a, b]$

(7.26.3) $\quad \varphi(x) \geq \varphi(a) \quad \forall x \leq a \quad$ *and* $\quad \varphi(x) \geq \varphi(b) \quad \forall x \geq b$

(7.26.4) $\quad E(\varphi(X_{n+1} + W_n) \mid \mathcal{F}_n) \leq \mu + \varphi(W_n) \qquad \forall n \geq 0$

Then we have

$$E(\tau \mid \mathcal{F}_0) \geq \frac{\varphi(a) - \varphi(W_0)}{\mu} \tag{7.26.5}$$

Remarks: (1): The model described above, applies to a large class of queuing models, storage models, waiting time models and inventory models. Let me describe a water reservoir model. Consider a water reservoir with a finite capacity b . Let I_n denote the inset to the reservoir in the time interval from $(n-1)$ to n , let D_n denote the demand of water in the same time period, and let $X_n = I_n - D_n$. If W_n denotes the water contents at time $n = 0, 1, 2, \ldots$, then $\{W_n\}$ must satisfy (7.26.1) because the water content cannot be negative nor can it exceed the capacity b of the reservoir. Let $0 \leq a < b$ be the minimal acceptable level of the reservoir. Then τ is the first time that the water content drops below the acceptable level, and (7.26.5) gives a lower bound for the (conditional) expectation of τ .

(2): Suppose that φ is decreasing on $]-\infty, b]$ and increasing on $[b, \infty[$ and that $\varphi(a)$ and $\varphi(b)$ are finite, then (7.26.2) and (7.26.3) hold. In this case (7.26.4) expresses that we don't expect X_{n+1} to be too large nor too small. An easy computation shows the following:

(7.26.6) If $E(X_{n+1} \mid \mathcal{F}_n) \leq \alpha$ and $E(e^{-\lambda X_{n+1}} \mid \mathcal{F}_n) \leq 1$ for some $\alpha, \lambda > 0$, then $\varphi(x) = \lambda e^{-\lambda b} x + e^{-\lambda x}$ satisfies (7.26.2)–(7.26.4) with $\mu = \lambda \alpha e^{-\lambda b}$

Proof: By induction, we see that W_n is $\mathcal{F}_n$-measurable. Hence, by (7.2.4) we have that τ is a stopping time. Let $m = (\varphi(a) - c)/\mu$ and $\psi(x) = (\varphi(a) - \varphi(x))/\mu$ for $x \in \overline{\mathbf{R}}$. Then by (7.26.2) and (7.26.3) we have that $\psi(x) \leq \psi(a) = 0$ for all $x \leq a$, $0 \leq \psi(x) \leq m < \infty$ for all $x \in [a, b]$, and $\psi(x) \leq \psi(b)$ for all $x \geq b$. Hence, if

$$\psi_0(x) = \begin{cases} 0 & \text{if } x \leq a \\ \psi(x) & \text{if } a \leq x \leq b \\ \psi(b) & \text{if } x \geq b \end{cases}$$

then $0 \leq \psi_0(x) \leq m$ for all $x \in \overline{\mathbf{R}}$. Since $a \geq 0$, then by definition of ψ_0 we have that $\psi_0(x) = \psi_0(x^+ \wedge b) \geq \psi(x)$ for all $x \in \overline{\mathbf{R}}$. In particular, we have that $\psi_0(W_{n+1}) \geq \psi(X_{n+1} + W_n)$. Hence, by (7.26.4) and (6.8) we find

$$\begin{aligned} E(\psi_0(W_{n+1}) \mid \mathcal{F}_n) &\geq E(\psi(X_{n+1} + W_n) \mid \mathcal{F}_n) \\ &= \mu^{-1}\big(\varphi(a) - E(\varphi(X_{n+1} + W_n) \mid \mathcal{F}_n)\big) \\ &\geq \mu^{-1}\big(\varphi(a) - \mu - \varphi(W_n)\big) = \psi(W_n) - 1 \end{aligned}$$

and since $\psi_0 \geq 0$, $W_n \leq b$ and $(\psi(x)-1)^+ \geq \psi_0(x)-1$ for all $x \leq b$, we get

$$E(\psi_0(W_{n+1}) \mid \mathcal{F}_n) \geq (\psi(W_n) - 1)^+ \geq \psi_0(W_n) - 1$$

Thus, $\{\psi_0(W_n)\}$ is a scaled submartingale, and by rescaling [see (7.22.8)], we see that $\{\psi_0(W_n)+n\}$ is a submartingale. By assumption we have $a \leq W_0 \leq b$ and so $\psi_0(W_0) = \psi(W_0)$ by definition of ψ_0 . Hence, if $F = \{\tau^* < \infty\}$ and $G = \{\tau^* = \infty\}$, where $\tau^* = E(\tau \mid \mathcal{F}_0)$, then $F \in \mathcal{F}_0$ and so by (7.7.3) and (6.8.2) we get

$$\begin{aligned}\psi(W_0) &\leq 1_F\, \psi_0(W_0) + 1_G\, \tau^* \\ &\leq 1_F\, E(\psi_0(W_{\tau\wedge k}) \mid \mathcal{F}_0) + 1_F\, E(\tau \wedge k \mid \mathcal{F}_0) + 1_G\, \tau^* \\ &\leq E(1_F\, \psi_0(W_{\tau\wedge k}) \mid \mathcal{F}_0) + \tau^*\end{aligned}$$

Moreover, τ^* is $\mathcal{F}_0$-measurable and so

$$\int_{\{\tau^* \leq n\}} \tau\, dP = \int_{\{\tau^* \leq n\}} \tau^*\, dP \leq n \qquad \forall n \geq 0$$

Hence, $\tau(\omega) < \infty$ for a.a. $\omega \in F$, and since $W_\tau(\omega) \leq a$ for all $\omega \in F$ we have that $1_F\, \psi_0(W_{\tau\wedge k}) \to 1_F\, \psi_0(W_\tau) = 0$ a.s. Moreover, by (6.10.7) and boundedness of ψ_0 we find

$$\psi(W_0) \leq \tau^* + \lim_{n\to\infty} E(1_F\, \psi_0(W_{\tau\wedge k}) \mid \mathcal{F}_0) = \tau^*$$

Thus, (7.26.5) is proved. □

7.27 Convergence of U-Statistics

Let $X = (X_1, X_2, \cdots)$ be an infinite sequence of exchangeable random functions with values in the measurable space $(T, \mathcal{B})$, let $\mathcal{S}_X^n$ be the permutation invariant σ-algebras [see (6.17.8)], and let $\mathcal{T}_X^n$ be the tail σ-algebras [see (6.19.1+2)] for all $n \in \mathbf{N} \cup \{\infty\}$. Let $Y \in L(P)$ and let $f : T^\infty \to \overline{\mathbf{R}}$ be a measurable function such that $f(X) \in L(P)$ if $U_n := S_n f(X)$ is the U-statistics associated to f [see (6.17.7) and (6.18.1)]. Then we have

(7.27.1) $\quad U_n \to E(f(X) \mid \mathcal{S}_X^\infty)$ *a.s.*

(7.27.2) $\quad U_n \to E(f(X) \mid \mathcal{S}_X^\infty)$ *a.s. and in 1-mean if* $f(X) \in L^1(P)$

(7.27.3) $\quad E(Y \mid \mathcal{T}_X^\infty) = E(Y \mid \mathcal{S}_X^\infty)$

Thus, any $\mathcal{S}_X^\infty$-measurable function equals a $\mathcal{T}_X^\infty$-measurable function a.s. Moreover, if $\varphi : T \to \overline{\mathbf{R}}$ and $\psi : T \times T \to \overline{\mathbf{R}}$ are measurable functions satisfying

$$\varphi(X_1) \in L(P) \quad , \quad \psi(X_1, X_2) \in L(P) \quad , \quad E\sqrt{|\psi(X_1, X_1)|} < \infty \tag{7.27.4}$$

then we have

$$\frac{1}{n}\sum_{j=1}^{n} \varphi(X_n) \to E(\varphi(X_1) \mid \mathcal{S}_X^n) \quad a.s. \tag{7.27.5}$$

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \psi(X_i, X_j) \to E(\psi(X_1, X_2) \mid \mathcal{S}_X^n) \quad a.s. \tag{7.27.6}$$

Finally, suppose that $X_1, X_2, \ldots$ are independent and identically distributed. Then $X_1, X_2, \ldots$ are exchangeable and we have

(7.27.7) $P(F) = 0$ *or* 1 *for all* $F \in \mathcal{S}_X^\infty$ *and every* $\mathcal{S}_X^\infty$*-measurable random variable is degenerated*

$$E(Y \mid \mathcal{S}_X^\infty) = EY \quad a.s. \quad \forall Y \in L(P) \tag{7.27.8}$$

Remark: Recall that $\mathcal{T}_X^\infty \subseteq \mathcal{S}_X^\infty$ [see (6.19.3)] and note that (7.27.3) states that a set in $\mathcal{S}_X^\infty$ a.s. equals a set in $\mathcal{T}_X^\infty$. Thus, the two σ-algebras are essentially equal. Observe that (7.27.5) is a version of the strong law of large numbers for exchangeable random variables, and that (7.27.6) is a quadratic form of the strong law of large numbers for exchangeable random variables. Note that (7.27.7) is a 0-1 law for the permutation invariant σ-algebra $\mathcal{S}_X^\infty$.

Proof: (7.27.1+2): Follows from (6.18.1) and Levy's theorem (7.12.2+3) applied to the filter $\{\mathcal{S}_X^{-n} \mid -\infty < n \leq 1\}$.

(7.27.3): Suppose that there exists $a \in \mathbf{R}_+$ such that $|Y| \leq a$, and let $q \in \mathbf{N}$ be fixed for a moment. If

$$\mathcal{F}_q = \sigma\{X_1, \ldots, X_q\} \quad , \quad V_q = E(Y \mid \mathcal{F}_q)$$
$$g(t_1, \ldots, t_q) = E(Y \mid X_1 = t_1, \ldots, X_n = t_q)$$

then $|g(t)| \leq a$ and $V_q = g(X_1, \ldots, X_q)$ by (6.7.7). Let U_n be the U-statistics associated with g :

$$U_n = \frac{1}{n!}\sum_{\sigma \in \mathcal{P}_n} g(X_{\sigma(1)}, \ldots, X_{\sigma(q)})$$

Let $n > 2q$, and set $\mathcal{R}_{nq} = \{\sigma \in \mathcal{P}_n \mid \sigma(j) > q \ \forall 1 \le j \le q\}$ and

$$U_n^* = \frac{1}{n!} \sum_{\sigma \in \mathcal{R}_{nq}} g(X_{\sigma(1)}, \ldots, X_{\sigma(q)})$$

Then, evidently, we have:

$$\mathcal{P}_n \setminus \mathcal{R}_{nq} = \bigcup_{j=1}^{q} \bigcup_{i=1}^{q} \{\sigma \in \mathcal{P}_n \mid \sigma(i) = j\}$$

and so we see that $\#(\mathcal{P}_n \setminus \mathcal{R}_{nq}) \le q^2(n-1)!$. Moreover, since $|g| \le a$, we get

$$|U_n(\omega) - U_n^*(\omega)| \le \frac{a}{n!} \#(\mathcal{P}_n \setminus \mathcal{R}_{nq}) \le \frac{a\, q^2}{n} \qquad \forall n > 2q \ \forall \omega \in \Omega$$

Hence, $U_n - U_n^* \to 0$ uniformly on Ω and $U_n \to E(V_q \mid \mathcal{S}_X^\infty)$ a.s. and in 1-mean by (7.27.2). Thus, $U_n^* \to E(V_q \mid \mathcal{S}_X^\infty)$ a.s. and in 1-mean. Since U_n^* is $\mathcal{T}_X^q$-measurable for all $n > 2q$, $\mathcal{F}_q \uparrow \mathcal{S}_X^1 = \sigma(X)$ and Y is bounded, then by (6.10.7) and Levy's theorem (7.12.3) we see that $W_q := \liminf_{n\to\infty} U_n^*$ is a $\mathcal{T}_X^q$-measurable random variable such that $W_q = E(V_q \mid \mathcal{S}_X^\infty)$ and $V_q \to E(Y \mid \mathcal{S}_X^1)$ a.s. for $q \to \infty$. Moreover, since $|V_q| \le a$, then by (6.10.7) and (6.8.8) we have

$$\lim_{q\to\infty} W_q = \lim_{q\to\infty} E(V_q \mid \mathcal{S}_X^\infty) = E(E(Y \mid \mathcal{S}_X^1) \mid \mathcal{S}_X^\infty) = E(Y \mid \mathcal{S}_X^\infty) \quad \text{a.s.}$$

Hence, if $W = \liminf W_q$, then W is a $\mathcal{T}_X^\infty$-measurable random variable by (1.40.3) such that $W = E(Y \mid \mathcal{S}_X^\infty)$ a.s. Thus, if Y is bounded, then (7.27.3) follows from (6.8.9) and (6.19.3). If Y is unbounded, then (7.27.3) follows easily from the bounded case by approximating Y by bounded random variables and apply the convergence theorems for conditional expectations (6.10).

(7.27.5+6): Setting $f(t_1, t_2, \ldots) = \varphi(t_1)$ we see that (7.27.5) follows from (7.27.1). Moreover, by (6.18.10) we get

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi(X_i, X_j) = \left(1 - \frac{1}{n}\right) U_n + \frac{1}{n^2} \sum_{j=1}^{n} \psi(X_j, X_j)$$

where U_n is the U-statistics associated with $f(t_1, t_2, t_3 \ldots) = \psi(t_1, t_2)$, and so (7.27.6) follows from (7.27.1) and (4.36.3) with $r = 1$, $q = \frac{1}{2}$ and $Y_n = \psi(X_n, X_n)$.

Finally, note that the last part of the theorem follows from (7.27.3) and the 0–1 law (2.13). □

7.28 References to Chapter 7

Martingales was first, formally, introduced by Doob (see [2]). However, the very first probabilists (Huyghens, the Bernuolli's, de Moivre, etc.) often used the method of constructing a martingale (*a fair game*, in their terminology) from a given sequence of random variables and thereby solve their problem by use of the optional sampling theorem. In the early days of probability, the optional sampling theorem was taken for granted, without realizing that it requires a proof and is not generally true. Later the optional sampling theorem and the martingale convergence theorem was proved in many particular cases. But first with the emergence of the general notion of a martingale, probabilists realized that the many ostensible disparate results could be proved in one stroke. Karlin and Taylor's textbook [5] contains a wealth of applications of martingale theory to daily-life problems. Neveu's textbook [6] contains a wealth of information about martingales with discrete parameter, that is, the kind of martingales studied in this chapter. The notion of a martingale offers many different extensions. For instance, in [1] you will find an extension to martingales with continuous parameter, that is, martingales indexed by a parameter running in an interval; and in [4] you will an extension of the theory to the so-called *amarts*. The martingale convergence theorem may be considered as a law of large numbers for martingales (actually, many of the laws of large numbers derived in Chapter 4 are consequences of the martingale convergence theorem). The central limit theorem can, likewise, be extended to martingales, and it can be shown (see [3]) that a conditional form of the Lindeberg condition (5.22.1) implies asymptotic normality of martingales.

[1] Dellacherie, C. and Meyer, P.-A.(1978): *Probabilities and Potentials*, North-Holland Publishing Co., Amsterdam and New York

[2] Doob, J. L. (1940): *Regularity Properties of Certain Families of Chance Variables*, Trans. Amer. Math. Soc. 47, pp. 455–486

[3] Gaenssler, P., Strobel, J. and Stute, W. (1978): *On Central Limit Theorems for Martingale Triangular Arrays*, Acta Math. Acad. Sci. Hung. 31, pp. 205–216

[4] Edgar, G. A. and Sucheston, L. (1976): *A Class of Asymptotic Martingales. A: Discrete Parameter*, Journ. Multivar. Anal. 6, pp. 193–221

[5] Karlin, S. and Taylor, M. H.(1975): *A First Course in Stochastic Processes*, Academic Press, New York and London

[6] Neveu, J. (1975): *Discrete-Parameter Martingales*, North-Holland Publishing Co., Amsterdam and Oxford; and American Elsevier Publishing Co., Inc., New York

7.29 Exercises to Chapter 7

Exc. 7.1: Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter and let τ be a stopping time for $\{\mathcal{F}_n \mid n \in T\}$. Suppose that $\{X_n \mid n \in T^*\} \subseteq L(P)$ and $\{Y_n \mid n \in T^*\} \subseteq L(P)$ are random variables such that $X_\tau, Y_\tau \in L(P)$ and

$$E(X_n \mid \mathcal{F}_n) \leq E(Y_n \mid \mathcal{F}_n) \qquad \forall n \in T^*$$

Show that $E(X_\tau \mid \mathcal{F}_\tau) \leq E(Y_\tau \mid \mathcal{F}_\tau)$. [Mimic the proof of (7.4.6).]

Exc. 7.2: Let $\{X_n, \mathcal{F}_n \mid n \in T\}$ be a martingale such that $X_n \in L^2(P)$ for all $n \in T$. Let $n, k \in T$ such that $n \leq k$. Show that

(i) $E(X_n X_k \mid \mathcal{F}_n) = X_n^2 \quad , \quad E(X_n (X_k - X_n) \mid \mathcal{F}_n) = 0$

(ii) $E((X_k - X_n)^2 \mid \mathcal{F}_n) = E(X_k^2 - X_n^2 \mid \mathcal{F}_n)$

(iii) $E(X_n X_k) = EX_n^2 \quad , \quad E(X_n (X_n - X_k)) = 0$

(iv) $E(X_k - X_n)^2 = EX_k^2 - EX_n^2$

Exc. 7.3: Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter where $T = [[\alpha, \beta]]$ is an integer interval, and let $\{D_n \mid n \in T\}$ and $\{S_n \mid n \in T\}$ be two sequences of ***finite*** random variables such that D_n and S_n are $\mathcal{F}_n$-measurable for all $n \in T$ and $D_{n+1} = S_{n+1} - S_n$ for all $n \in [[\alpha, \beta - 1]]$. Show the following [see (7.17)]:

(i) $\{S_n, \mathcal{F}_n \mid n \in T\}$ is a submartingale if and only if $ES_n^+ < \infty$ for all $n \in T$ and $E(D_{n+1} \mid \mathcal{F}_n) \geq 0$ for all $n \in [[\alpha, \beta - 1]]$

(ii) $\{S_n, \mathcal{F}_n \mid n \in T\}$ is a supermartingale if and only if $ES_n^- < \infty$ for all $n \in T$ and $E(D_{n+1} \mid \mathcal{F}_n) \leq 0$ for all $n \in [[\alpha, \beta - 1]]$

(iii) $\{S_n, \mathcal{F}_n \mid n \in T\}$ is a martingale if and only if $E|S_n| < \infty$ for all $n \in T$ and $E(D_{n+1} \mid \mathcal{F}_n) = 0$ for all $n \in [[\alpha, \beta - 1]]$

Exc. 7.4: Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter where $T = [[\alpha, \beta]]$ is an integer interval, and let $\{V_n \mid n \in T\}$ and $\{D_n \mid n \in T\}$ be two sequences of ***finite*** random variables such that D_n is $\mathcal{F}_n$-measurable for all $n \in T$, V_n is $\mathcal{F}_{n-1}$-measurable and $V_n D_n \in L(P)$ for all $n \in T$ with the convention: $\mathcal{F}_{\alpha-1} := \mathcal{F}_\alpha$. Show the following [see (7.17)]:

(i) If $\{D_n\}$ is a martingale difference, then so is $\{V_n D_n\}$

(ii) If $V_n \geq 0$ for all $n \in [[\alpha+1, \beta]]$ and $\{D_n\}$ is a submartingale difference, then so is $\{V_n D_n\}$

(iii) If $V_n \geq 0$ for all $n \in [[\alpha+1, \beta]]$ and $\{D_n\}$ is a supermartingale difference, then so is $\{V_n D_n\}$

Exc. 7.5: A rabbit is jumping around at random in the following way: At time 0, it starts at $(0,0)$; at time $n = 1, 2, \ldots$, it tosses a fair coin if head comes up it jumps one step due north, and if tails comes up it jumps one step due west. Thus, its position at time n equals $Z_n = (X_n, Y_n)$ where X_n is the number of heads in the first n tosses and Y_n is the number of tails in the first n tosses. Let $\mathcal{F}_n = \sigma(Z_0, \ldots, Z_n)$. Show that $\{(X_n - \frac{n}{2}), \mathcal{F}_n | \ n \geq 0\}$ and $\{(Y_n - \frac{n}{2}), \mathcal{F}_n | \ n \geq 0\}$ are martingales.

Suppose that the rabbit is fenced in a bounded region $\Gamma \subseteq \mathbf{N}_0 \times \mathbf{N}_0$, and let τ be the first time that the rabbit hits the fence, that is,

$$\tau = \inf\{n \geq 0 \mid Z_n \notin \Gamma\}$$

Show that τ is a bounded a stopping time for $\{\mathcal{F}_n \mid n \geq 0\}$ satisfying

(i) $EX_\tau = \frac{1}{2} E\tau = EY_\tau$

Thus, no matter what the bounded region Γ looks like, then at the time when the rabbit hits the fence, it has, on average, moved the same number of steps to the north as to the west.

Exc. 7.6: Let $\xi_1, \xi_2, \ldots$ be independent random variables, and let A and B be two Borel subset of $\mathbf{R}$ such that $P(\xi_j \in A) = p$ and $P(\xi_j \in B) = q$ for all $j = 1, 2, \ldots$ where $p, q > 0$ are two given positive numbers. Let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ and let τ be the least $j \geq 1$ such that $\xi_{j-1} \in A$ and $\xi_j \in B$, that is,

$$\tau(\omega) = \inf\{j \geq 2 \mid \xi_{j-1}(\omega) \in A \ , \ \xi_j(\omega) \in B\} \qquad (\inf \emptyset = \infty)$$

Show that τ is a proper a stopping time for $\{\mathcal{F}_n \mid n \geq 1\}$ satisfying

(i) $P(\tau > 2n) \leq (1 - pq)^n \qquad \forall n \geq 1$

(ii) $E\tau = \frac{1}{pq} + \frac{1}{p} P(\xi_\tau \in A)$

(iii) $E\tau = \begin{cases} \frac{1}{pq} & \text{if } A \cap B = \emptyset \\ \frac{1}{pq} + \frac{1}{p} & \text{if } A \supseteq B \end{cases}$

Remark: Suppose that you are tossing a fair coin, let τ_{hh} be the waiting time for the first occurrence of "head-head", and let τ_{ht} be waiting time for the first occurrence of "head-tail" in that order. Then by (iii) we have $E\tau_{hh} = 6$ and $E\tau_{ht} = 4$ in spite of the fact, that the two event "head-head" and "head-tail" we are waiting for have the same probability, namely $\frac{1}{4}$. Can you give a reasonable intuitive argument for this difference? I can't!

Hint: (i): Note that

$$\{\tau > 2n\} \subseteq \bigcap_{j=1}^{n} \{(\xi_{2j-1}, \xi_{2j}) \notin A \times B\}$$

and apply the independence of $\xi_1, \xi_2, \ldots$. (ii)+(iii): Let $F_j = \{\xi_j \in A\}$ for $j \geq 1$ and let $G_1 = \emptyset$ and $G_j = \{\xi_{j-1} \in A\ ,\ \xi_j \in B\}$ for $j \geq 2$. Set

$$M_j = \tfrac{1}{p}\, 1_{F_j}\ ,\quad L_j = \tfrac{1}{pq}\, 1_{G_j}\ ,\quad X_n = M_n - n + \sum_{j=1}^{n} L_j$$

and show that $E(M_{n+1} \mid \mathcal{F}_n) = 1$ and $E(L_{n+1} \mid \mathcal{F}_n) = M_n$. Deduce that $\{X_n, \mathcal{F}_n\}$ is a martingale such that $EX_n = 0$ and

$$X_\tau = \tfrac{1}{p}\, 1_{\{\xi_\tau \in A\}} - \tau + \tfrac{1}{pq}$$

Finally, show that τ is optional for $\{X_n\}$ and apply the optional sampling theorem (7.14.3) with $\sigma \equiv 1$.

Exc. 7.7: Let $X_1, X_2, \ldots$ be independent random variables all having a Poisson distribution $Po(\lambda)$ where $\lambda > 0$ is a given positive number. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and set

$$S_n = \sum_{j=1}^{n} X_j\ ,\quad M_n = S_n - n\lambda\ ,\quad \tau(\omega) = \inf\{n \geq 1 \mid S_n(\omega) \geq 2\}$$

with the usual convention: $\inf \emptyset = \infty$. Show that $\{M_n, \mathcal{F}_n \mid n \geq 1\}$ is a martingale with $EM_n = 0$ and that τ is a proper a stopping time for $\{\mathcal{F}_n \mid n \geq 1\}$ satisfying

(i) $\quad P(\tau > n) = e^{-n\lambda}(1 + n\lambda) \qquad \forall n = 0, 1, 2, \ldots$

(ii) $\quad ES_\tau = \lambda E\tau = \dfrac{\lambda - \lambda e^{-\lambda} + \lambda^2 e^{-\lambda}}{(1 - e^{-\lambda})^2}$

Hint: (i): Note that $\{\tau > n\} = \{S_n \geq 2\}$. (ii): Apply (4.2.9) to show that

$$E\tau = \frac{1 - e^{-\lambda} + \lambda e^{-\lambda}}{(1 - e^{-\lambda})^2} < \infty$$

$$1_{\{\tau > n\}} |M_n| \leq 1 + \lambda\tau \quad , \quad |M_\tau| \leq 1 + X_\tau + \lambda\tau$$

Observe that $\{X_j > n\}$ and $\{\tau > j - 1\}$ are independent for all $n, j \geq 0$. Deduce that

$$EX_\tau = \sum_{n=0}^{\infty} \sum_{j=1}^{\infty} P(X_j > n \, , \, \tau = j) \leq \sum_{n=0}^{\infty} \sum_{j=1}^{\infty} P(X_j > n \, , \, \tau > j - 1)$$

$$= (EX_1) \cdot (E\tau) < \infty$$

Finally, apply criterion (7.16.4) to show that τ is optional for $\{M_n\}$.

Exc. 7.8: Let $\{X_n, \mathcal{F}_n | \, n \geq 0\}$ be a supermartingale, let τ be a stopping time for $\{\mathcal{F}_n \mid n \geq 0\}$, and let

$$W_\tau = \sup_{n \geq 0} X_{\tau \wedge n} \quad , \quad V_\tau = \inf_{n \geq 0} X_{\tau \wedge n}$$

Show that [apply (7.8.1)–(7.8.5) on the submartingale $\{-X_n\}$]

(i) $$\int_{\{\tau < \infty\}} X_\tau^- \, dP \leq \sup_{n \geq 0} EX_{\tau \wedge n}^- \leq \sup_{n \geq 0} EX_n^-$$

(ii) $$\int_{\{\tau < \infty\}} X_\tau^+ \, dP \leq \sup_{n \geq 0} E(X_n^- + X_0)$$

(iii) $$\int_{\{\tau < \infty\}} |X_\tau| \, dP \leq \sup_{n \geq 0} E\big(2\, X_{\tau \wedge n}^- + X_0\big)$$

(iv) $$P(W_\tau \geq t) \leq \frac{1}{t} \sup_{n \geq 0} E\big(X_{\tau \wedge n}^- + X_0\big) \qquad \forall t > 0$$

(v) $$P(V_\tau \leq -t) \leq \frac{1}{t} \sup_{n \geq 0} EX_{\tau \wedge n}^- \qquad \forall t > 0$$

Exc. 7.9: Let $(\Omega, \mathcal{F}, P)$ be a probability space where $\Omega = \mathbf{N}$, $\mathcal{F} = 2^{\mathbf{N}}$ and $p(n) := P(\{n\}) > 0$ for all $n \in \mathbf{N}$. Let $\mathcal{F}_n = \sigma(\{1\}, \ldots, \{n\})$ and

$$r(n) = P(\{n+1, n+2, \ldots\}) = \sum_{j=n+1}^{\infty} p(j) \qquad \forall n \geq 0$$

$$R_X(x, n) = \frac{1}{r(n)} \left\{ x - \sum_{j=1}^{n} p(j)\, X(j) \right\} \qquad \forall n \in \mathbf{N} \, \forall x \in \mathbf{R}$$

whenever $X : \mathbf{N} \to \mathbf{R}$ is a function. Show that the function $\tau : \mathbf{N} \to \mathbf{N} \cup \{\infty\}$ is a stopping time for $\{\mathcal{F}_n \mid n \geq 1\}$ if and only if τ satisfies

$$(*) \qquad \exists k \in \mathbf{N} \cup \{\infty\} \quad \text{so that} \quad \tau(\omega) \wedge \omega = \omega \wedge k \quad \forall \omega \in \mathbf{N}$$

Let $X_1, X_2, \ldots$ be a sequence of functions from $\mathbf{N}$ into $\mathbf{R}$. Show that $\{X_n, \mathcal{F}_n \mid n \geq 1\}$ is a martingale if and only if there exist a real number a and a function $X : \mathbf{N} \to \mathbf{R}$ satisfying

$$(**) \qquad X_n(\omega) = \begin{cases} X(\omega) & \text{if } \omega \leq n \\ R_X(a, n) & \text{if } \omega > n \end{cases}$$

Let $\{X_n, \mathcal{F}_n \mid n \geq 1\}$ be a martingale given by (**) and let τ be a stopping time for $\{\mathcal{F}_n \mid n \geq 1\}$ given by (*). Set $X_\infty = X$ and show that

(i) $\quad X_\tau = X_k$ and $\lim_{n\to\infty} X_n(\omega) = X(\omega) \quad \forall \omega \in \mathbf{N}$

(ii) $\quad k < \infty$ or $E|X| < \infty \quad \Rightarrow \quad E(X_\tau \mid \mathcal{F}_\sigma) = X_{\tau\wedge\sigma} \quad \forall \sigma \in \mathcal{S}$

where $\mathcal{S}$ as usual denotes the set of all stopping times. Thus, all martingales on $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}, P)$ converges everywhere. Finally, give an example of a martingale $\{X_n, \mathcal{F}_n \mid n \geq 1\}$ satisfying

(iii) $\quad \sup_{n\geq 1} EX_n^+ = \infty$, but $\{X_n(\omega)\}$ converges to a finite number $X(\omega)$ for all $\omega \in \Omega$

Hint: (*): Suppose that τ is a stopping time. Show that at most one of the sets $\{\tau = n\}$ can be infinite. If none of the sets $\{\tau = n\}$ are infinite, set $k = \infty$; otherwise let $k \in \mathbf{N}$ be the unique integer such that $\{\tau = n\}$ is infinite. Show that (*) holds with this choice of k. The converse is easy. (**)+(i)+(ii): Let $Y \in L^1(P)$ be a finite random variable with $EY = \mu$ and show that

$$E(Y \mid \mathcal{F}_n)(\omega) = \begin{cases} Y(\omega) & \text{if } \omega \leq n \\ R_Y(\mu, n) & \text{if } \omega > n \end{cases}$$

(iii): Let $X(\omega) = \frac{1}{p(\omega)}$ and define X_n by (**) with $a = 0$. Show that $EX_n^+ = n$ for all $n \geq 1$.

Exc. 7.10: Let $\{\mathcal{F}_n \mid n \in T\}$ be a filter on the probability space $(\Omega, \mathcal{F}, P)$ where $T = [[\alpha, \beta]]$ is a given integer interval. Let $f : T^* \to T^*$ and $g : T^* \times T^* \to T^*$ be functions satisfying

$$f(i) \geq i \quad \text{and} \quad g(i,j) \geq i \vee j \qquad \forall i, j \in T^*$$

Let τ and σ be stopping times. Show that $f(\tau)$ and $g(\tau, \sigma)$ are stopping times.

Exc. 7.11: Let $\{\mathcal{F}_n \mid n \geq 1\}$ be a filter and let $\{X_n\}$ be a sequence of random variables. Let τ be a stopping time for $\{\mathcal{F}_n \mid n \geq 1\}$ such that $EU_\tau < \infty$ where

$$U_n = \sum_{j=1}^{n} E\big(X_j^+ \mid \mathcal{F}_j\big) \quad \forall n = 1, 2, \ldots, \infty$$

Show that τ is optional for $\{S_n^+\}$ and $\{V_n\}$ where

$$S_n = \sum_{j=1}^{n} X_j \quad , \quad V_n = \sum_{j=1}^{n} X_j^+ \quad \forall n = 1, 2, \ldots$$

Hint: Let $V_\infty = \sum_{j=1}^{\infty} X_j^+$. Show that $EV_\tau = EU_\tau < \infty$ [apply Beppo-Levi's theorem] and note that $S_{\tau\wedge n}^+ \leq V_{\tau\wedge n} \leq V_\tau$. Then show that the result follows from (7.16.2).

Exc. 7.12: Let $\{X_n, \mathcal{F}_n \mid n \geq 1\}$ be a martingale and set $X_0 = 0$. Suppose

$$\sum_{n=1}^{\infty} E(X_n - X_{n-1})^2 < \infty$$

Show that $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists for a.a. $\omega \in \Omega$ and that $EX_\infty^2 < \infty$.

Hint: Show that $EX_n^2 = E(X_1 - X_0)^2 + \cdots E(X_n - X_{n-1})^2$ [see Exc.7.2.(iv)] and apply the martingale convergence theorem (7.10).

Exc. 7.13: Let $\varepsilon_1, \varepsilon_2, \ldots$ be independent identically distributed random variables with $P(\varepsilon_1 = 1) = p$ and $P(\varepsilon_1 = -1) = q$ where $\frac{1}{2} < p \leq 1$ and $p + q = 1$. Let $0 < u \leq 1 \leq t < \infty$ such that $u\,(tp + \frac{q}{t}) = 1$ and let $a \in \mathbf{N}$ be a given integer. Set $\mathcal{F}_n = \sigma(\varepsilon_1, \ldots \varepsilon_n)$ and

$$S_n = \varepsilon_1 + \cdots + \varepsilon_n \quad , \quad M_n = u^n\, t^{S_n} \quad , \quad \tau = \inf\{\, n \geq 1 \mid S_n \geq a \,\}$$

Show that $\{M_n, \mathcal{F}_n \mid n \geq 1\}$ is a martingale and that τ is a proper stopping time. Moreover, show that τ is optional for $\{M_n\}$ and that $q_\tau(u) = t^{-a}$, where q_τ denotes the generating function of τ . Let $0 < u \leq 1$ be given and show that the equation: $u\,(tp + \frac{q}{t}) = 1$ in the unknown t has two solutions t_1, t_2 satisfying $0 < t_1 \leq 1 \leq t_2$. Deduce that

(i) $$q_\tau(u) = \left(\frac{2pu}{1 + \sqrt{1 - 4pqu}}\right)^a \quad \forall\, 0 < u \leq 1$$

Hint: Note that $E\varepsilon_1 = p - q > 0$ and apply the law of large numbers (4.12) to conclude that $S_n \to \infty$ a.s., and argue that implies that τ is proper. Moreover, observe that

$$M_\tau = u^\tau t^a \le t^a \quad \text{and} \quad \{\tau > n\} \subseteq \{0 \le M_n \le t^a\}$$

Apply (7.16.4) to prove optionality of τ. Finally, apply the optional sampling theorem (7.14) to deduce that $q_\tau(u) = t^{-a}$.

Exc. 7.14: Let $\xi_1, \xi_2, \ldots$ be independent non-negative random variables with the same finite mean $\mu := E\xi_n > 1$ for all $n \ge 1$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and let X_n and M_n be defined recursively as follows: $X_1 = \xi_1$ and

$$X_n = \xi_n X_{n-1} + \xi_n \quad , \quad M_n = \left(1 - \tfrac{1}{\mu}\right) \mu^{-n} X_n + \mu^{-n} - 1$$

Show that X_n and M_n are $\mathcal{F}_n$-measurable and that

(i) $\quad X_n \ge 0 \quad , \quad E(X_{n+1} \mid \mathcal{F}_n) = \mu X_n + \mu \quad$ and $\quad EX_n = \mu \dfrac{\mu^n - 1}{\mu - 1}$

(ii) $\quad \{M_n, \mathcal{F}_n\}$ is a martingale with $EM_n = 0$ and $E|M_n| \le 2$

(iii) $\quad \lim\limits_{n\to\infty} M_n = M_\infty$ exists a.s. and $E|M_\infty| \le 2$

Exc. 7.15: Let $\xi_n = (\xi_{1n}, \ldots, \xi_{dn})$ for $n = 1, 2, \ldots$ be a sequence of independent d-dimensional random vectors and suppose that there exists $0 < q < 1$ satisfying

$$P(\xi_n = (0, \ldots, 0)) = q \quad , \quad \xi_{jn} \ge 0 \quad \text{and} \quad E\xi_{jn} = 1 \quad \forall n \ge 1 \, \forall 1 \le j \le d$$

Let $f(x) = (f_1(x), \ldots, f_d(x))$ be a Borel function from $\mathbf{R}$ into $\mathbf{R}^d$ satisfying

$$f_j(0) = 0 \quad , \quad f_j(x) > 0 \quad \text{and} \quad \sum_{i=1}^{d} f_i(x) = x \quad \forall x > 0 \, \forall 1 \le j \le d$$

Set $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ and

$$X_1 = \sum_{j=1}^{d} \xi_{j1} \quad , \quad X_{n+1} = \sum_{j=1}^{d} f_j(X_n)\, \xi_{j,n+1} \quad , \quad \tau = \inf\{n \mid X_n = 0\}$$

with the usual convention: $\inf \emptyset = \infty$. Show that $\{X_n, \mathcal{F}_n\}$ is a non-negative martingale with $EX_n = d$ for all $n \ge 1$ and that τ is a proper stopping time satisfying

(i) $\quad P(\tau > n) = (1 - q)^n \qquad \forall n = 0, 1, 2, \ldots$

(ii) $X_n(\omega) = 0 \qquad \forall n \geq \tau(\omega)$

(iii) $\lim_{n\to\infty} X_n = 0$

(iv) τ is **not** optional for $\{X_n\}$

Exc. 7.16: Let $X_1, X_2, \ldots$ be independent random variables which only assumes the values -1 , 0 or 1 with the following probabilities:

$$P(X_n = 0) = 1 - \tfrac{1}{n} \quad , \quad P(X_n = 1) = \tfrac{1}{2n} \quad , \quad P(X_n = -1) = \tfrac{1}{2n}$$

for all $n = 1, 2, \ldots$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and

$$M_1 = X_1 \quad , \quad M_{n+1} = 1_{\{M_n = 0\}}\, X_{n+1} + (n+1)\, M_n\, |X_{n+1}|$$

$$\tau = \inf\{\, n \geq 1 \mid M_n = 0 \,\} \quad (\, \inf \emptyset = \infty \,)$$

Show that M_n is integer-valued and that

(i) $\{M_n, \mathcal{F}_n \mid n \geq 1\}$ is a martingale with $E|M_n| \leq \sum_{j=1}^{n} \frac{1}{j}$ for all $n \geq 1$

(ii) $\{\, M_n \neq 0 \,\} = \{\, |M_n| \geq 1 \,\} = \{\, |X_n| = 1 \,\}$

(iii) τ is a proper stopping time

(iv) $M_n \to 0$ in *pr*, but $\limsup_{n\to\infty} |M_n| \geq 1$ a.s.

Hint: (iii): Note that $\tau = \infty$ if and only if $|X_n| \geq 1$ for all $n \geq 1$. Deduce that $P(\tau = \infty) = 0$. (iv): Apply (ii) and the second Borel-Cantelli lemma (2.11.2) noting that the events $\{\, |X_n| = 1 \,\}$ are independent for $n = 1, 2, \ldots$.

Exc. 7.17: Let $\varepsilon_0, \eta_1, \varepsilon_1, \eta_1, \ldots$ be random variables such that ε_{n+1} and $(\varepsilon_0, \eta_0, \ldots, \varepsilon_n, \eta_n)$ are independent for all $n \geq 0$, and let $0 < p_n < 1$ be given numbers satisfying

$$P(\varepsilon_n = 1) = P(\varepsilon_n = -1) = \tfrac{1}{2} \quad , \quad P(\eta_n = 1) = p_n \quad , \quad P(\eta_n = 0) = 1 - p_n$$

for all $n \geq 0$. Let $\mathcal{F}_n = \sigma(\varepsilon_0, \eta_0, \ldots \varepsilon_n, \eta_n)$ and

$$M_0 = 0 \quad , \quad M_n = M_{n-1}\,(1 + \varepsilon_n) + \varepsilon_n\, \eta_{n-1} \quad \forall n \geq 1$$

Show that $\{M_n, \mathcal{F}_n \mid n \geq 0\}$ is an integer-valued martingale satisfying

(i) $EM_n = 0$ and $E|M_n| \leq \sum_{j=1}^{n} p_j \quad \forall n \geq 1$

(ii) $\sum_{j=1}^{\infty} p_j < \infty \quad \Rightarrow \quad \lim_{n\to\infty} M_n = 0$ a.s.

Exc. 7.18: Let $Y_1, Y_2, \dots$ be independent random variables, let $0 < a_j \le 1$ be numbers such that $P(Y_j = a_j) = P(Y_j = -a_j) = \frac{1}{2}$ for all $j = 1, 2, \dots$, and let $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ and

$$M_n = \prod_{j=1}^{n} (1 + Y_j) \quad , \quad \tau = \inf\{\, n \ge 1 \mid Y_j = -a_j \,\} \quad (\inf \emptyset = \infty)$$

Show that $\{M_n, \mathcal{F}_n \mid n \ge 1\}$ is a non-negative martingale satisfying

(i) $EM_n = 1$ and $M_\infty = \lim_{n\to\infty} M_n$ exists a.s. and $EM_\infty \le 1$

(ii) τ is a proper stopping time with $P(\tau > n) = 2^{-n}$

(iii) $EM_{\tau\wedge n} = 1$ and $EM_\tau \le 1$

(iv) $\lim_{n\to\infty} 2^{-n} \prod_{j=1}^{n} (1 + a_j) = 0 \quad \Rightarrow \quad EM_\tau = 1$

Exc. 7.19: Let $\{(X_n, Y_n) \mid n \ge 0\}$ be independent 2-dimensional random vectors, and let $\alpha, \beta \in \mathbf{R}$ be real numbers satisfying $EX_n = \alpha$ and $EY_n = \beta$ for all $n \ge 0$. Let $\mathcal{F}_n = \sigma(X_0, Y_0, \dots, X_n, Y_n)$ and

$$S_n = \sum_{j=1}^{n} X_{j-1} Y_j \quad , \quad M_n = S_n + \beta X_n - n\alpha\beta \quad \forall n \ge 1$$

$$\tau = \inf\{\, n \ge 1 \mid X_{n-1} Y_n \ne 0 \,\} \quad (\inf \emptyset = \infty)$$

Show that $\{M_n, \mathcal{F}_n \mid n \ge 1\}$ is a martingale and that τ is a stopping time satisfying

(i) $EM_n = \alpha\beta$ and $M_n(\omega) = \beta X_n(\omega) - n\alpha\beta \quad \forall \omega \in \{\tau > n\}$

Suppose, moreover, that there exist $0 < p < 1$ such that $P(X_{n-1} Y_n = 0) \le p$ for all $n \ge 1$, and show the following:

(i) τ is proper and $E\tau < \infty$

(ii) If $E|Y_\tau| < \infty$ and $|X_n| \le c$ a.s. for all $n \ge 0$ and some number $0 < c < \infty$, then τ is optional for $\{M_n \mid n \ge 1\}$ and $EM_\tau = \alpha\beta$

Exc. 7.20: Let $X_1, X_2, \ldots$ be independent non-negative random variables with $EX_j = \mu_j$ and $EX_j^2 = 1$ for all $j = 1, 2, \ldots$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and

$$L_n = \prod_{j=1}^{n} X_j \quad , \quad \tau = \inf\{\, n \geq 1 \mid L_n \leq a \,\} \quad (\inf \emptyset = \infty)$$

where $0 < a < \infty$ is a given number. Show that $\{L_n, \mathcal{F}_n \mid n \geq 1\}$ is a uniformly integrable supermartingale and that $\{L_n^2, \mathcal{F}_n \mid n \geq 1\}$ is a martingale. Moreover, show that

(i) $L_\infty = \lim_{n\to\infty} L_n$ exists a.s. and $EL_\infty = \prod_{j=1}^{\infty} \mu_j$

(ii) τ is stopping time such that $\prod_{j=1}^{\infty} \mu_j \leq EL_\tau \leq \mu_1$

(iii) $\prod_{j=1}^{\infty} \mu_j = 0 \quad \Rightarrow \quad P(\tau = \infty) = 0$

Hint: Show that $\mu_j \leq 1$ [see (4.3.9)] and note that $EL_n^2 = 1$.

Exc. 7.21: Let $(X_0, Y_0), (X_1, Y_1), \ldots$ be independent 2-dimensional random variables such that $X_n \geq 0$, $EX_n = a_n < \infty$, $Y_n \geq 0$ and $EY_n = 1$ for all $n \geq 0$. Let $\mathcal{F}_n = \sigma(X_0, Y_0, \ldots, X_n Y_n)$ and

$$U_n = X_{n-1} Y_n \quad , \quad M_n = X_n + \sum_{j=1}^{n} (U_j - a_j) \quad , \quad \tau = \inf\{\, n \geq 1 \mid U_n = 0 \,\}$$

for $n = 1, 2, \ldots$ with the usual convention: $\inf \emptyset = \infty$. Show that τ is a stopping time and that $\{M_n, \mathcal{F}_n \mid n \geq 1\}$ is a martingale satisfying

(i) $EM_n = a_0$ and $E|M_n| \leq 2 \sum_{j=1}^{n} a_j$

(ii) $\sum_{j=1}^{\infty} a_j < \infty \quad \Rightarrow \quad M_\infty = \lim_{n\to\infty} M_n$ exist a.s. and $E|M_\infty| \leq 2 \sum_{j=1}^{\infty} a_j$

Moreover, suppose that there exist numbers $0 < C < \infty$ and $0 < \beta < 1$ satisfying

$$0 \leq X_n \leq C \quad , \quad 0 \leq Y_n \leq C \quad , \quad P(U_n = 0) \geq 1 - \beta \quad \forall n \geq 1$$

Show that τ is a proper stopping time satisfying

(iii) $E\tau < \infty$ and τ is optional for $\{M_n \mid n \geq 1\}$

(iv) $EM_\tau = a_0$

Hint: (iii): Show that $\{\tau > 2n\} \subseteq \{U_2 \neq 0, U_4 \neq 0, \ldots, U_{2n} \neq 0\}$ and that $U_2, U_4, U_6, \ldots$ are independent. Then conclude that $P(\tau > 2n) \leq \beta^n$. Thus, τ is proper and $E\tau < \infty$. Note that $|M_n| \leq C + (C^2 + C)n$ and argue that this implies optionality of τ.

Exc. 7.22: *(Scaled submartingales)* Let $\{\mathcal{F}_n \mid n \geq 0\}$ be a filter on the probability space $(\Omega, \mathcal{F}, P)$. Let X_n, α_n and β_n be $\mathcal{F}_n$-measurable random variables for $n = 0, 1, 2 \ldots$ such that $EX_n^+ < \infty$ and $\beta_n > 0$ a.s. for all $n \geq 0$ and [cf. (7.22), but observe that α_n and β_n are random variables]

(a) $\quad \alpha_n + \beta_n X_n \leq E(X_{n+1} \mid \mathcal{F}_n) \qquad \forall n \geq 1$

Let us define

$$B_n = \prod_{j=0}^{n-1} \frac{1}{\beta_j} \quad , \quad A_n = \sum_{j=0}^{n-1} \frac{\alpha_j B_{j+1}}{B_n} \quad , \quad L_n = B_n(X_n - A_n) \quad \forall n \geq 1$$

Show that A_n and B_n are $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$ and the following:

(i) $\quad \beta_n B_{n+1} = B_n \quad , \quad A_{n+1} = \alpha_n + \beta_n A_n \qquad \forall n \geq 1$

(ii) If $E(B_n X_n^+) < \infty$ and $E(B_n X_n^-) < \infty$, then $\{L_n, \mathcal{F}_n \mid n \geq 1\}$ is a submartingale

(iii) $\quad L_n^+ \leq B_n (X_n^+ + A_n^-) \qquad \forall n \geq 1$

Suppose, moreover, that the following condition holds:

(b) $\quad \sup_{n \geq 1} E(B_n X_n^+) < \infty \quad$ and $\quad \sup_{n \geq 1} E(B_n A_n^-) < \infty$

Set

$$C = \left\{ \omega \;\middle|\; \sum_{n=1}^{\infty} |1 - \beta_n(\omega)| < \infty \;,\; \sum_{n=1}^{n} |\alpha_n(\omega)| < \infty \right\}$$

and show that $X_\infty(\omega) = \lim_{n \to \infty} X_n(\omega)$ exists and is finite for a.a. $\omega \in C$.

Exc. 7.23: Let $T = [[\alpha, \beta]]$ and $S = [[\gamma, \delta]]$ be integer intervals, and let $\{X_n, \mathcal{F}_n \mid n \in T\}$ be a submartingale. Let $\{\tau_j \mid j \in S\}$ be a sequence of stopping times such that τ_j is optional for $\{X_n^+ \mid n \in T\}$ and $\tau_i \leq \tau_j$ for all $i, j \in S$ with $i \leq j$. Show that $\{X_{\tau_j}, \mathcal{F}_{\tau_j} \mid j \in S\}$ is a submartingale.

PREREQUISITES

P.1 Events and Subsets

In formal probability, events are considered as subsets of a given set Ω . If $\{A_j \mid j \in J\}$ is a family of subsets of Ω , then their *intersection*

$$\text{(P.1.1)} \qquad \bigcap_{j \in J} A_j := \{\omega \in \Omega \mid \omega \in A_j \quad \text{for all} \quad j \in J\}$$

is interpreted as the event that all the events A_j occurs. Their *union*

$$\text{(P.1.2)} \qquad \bigcup_{j \in J} A_j := \{\omega \in \Omega \mid \text{there exists some } j \in J \text{ , so that } \omega \in A_j\}$$

is interpreted as the event that at least one of the events A_j occurs. If A and B are subsets of Ω , then their *difference*

$$\text{(P.1.3)} \qquad A \setminus B := \{\omega \in \Omega \mid \omega \in A \quad \text{but} \quad \omega \notin B\}$$

is interpreted as the event that A occurs, but B does not. Their *symmetric difference*

$$\text{(P.1.4)} \qquad A \triangle B := (A \setminus B) \cup (B \setminus A)$$

is interpreted as the event, that exactly one of the events A or B occurs. The *complement*

$$\text{(P.1.5)} \qquad A^c := \Omega \setminus A = \{\omega \in \Omega \mid \omega \notin A\}$$

is interpreted as the event that A does not occur. Let $A_1, A_2, \ldots$ be a sequence of subsets of Ω . Then their *limes superior*

$$\text{(P.1.6)} \qquad \limsup A_n := \bigcap_{j=1}^{\infty} \bigcup_{n=j}^{\infty} A_n = \{\omega \in \Omega \mid \omega \in A_j \text{ for infinitely many } j\}$$

is interpreted as the event that infinitely of the events A_j occurs. Their *limes inferior*

$$\text{(P.1.7)} \qquad \liminf A_n := \bigcup_{j=1}^{\infty} \bigcap_{n=j}^{\infty} A_n = \{\omega \in \Omega \mid \exists j \text{ so that } \omega \in A_n \ \forall n \geq j\}$$

is interpreted as the event that the events A_j will occur from a certain number and onwards. We write

$$\text{(P.1.8)} \qquad A_n \uparrow A \quad \text{if} \quad A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq \ldots \quad \text{and} \quad A = \bigcup_{n=1}^{\infty} A_n$$

$$\text{(P.1.9)} \qquad A_n \downarrow A \quad \text{if} \quad A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots \quad \text{and} \quad A = \bigcap_{n=1}^{\infty} A_n$$

If (P.1.8) holds, we say that $\{A_n\}$ is *increasing* and that $\{A_n\}$ *increases to* A . Similarly, if (P.1.9) holds, we say that $\{A_n\}$ is *decreasing* and that $\{A_n\}$ *decreases to* A .

A family $\{A_j \mid j \in J\}$ is said to be *disjoint* if $A_i \cap A_j = \emptyset$ whenever $i, j \in J$ and $i \neq j$ åwhere $\emptyset$, as usual, denotes *the empty set.* The set of all subsets of the set Ω is called *the power set* of Ω , and it is denoted

$$\text{(P.1.10)} \qquad 2^{\Omega} := \{A \mid A \subseteq \Omega\}$$

Let $\Omega_1, \ldots, \Omega_n$ be a finite number of sets. Then the *product set* is defined by

$$\text{(P.1.11)} \qquad \Omega_1 \times \cdots \times \Omega_n = \prod_{j=1}^{n} \Omega_j := \{(\omega_1, \ldots, \omega_n) \mid \omega_1 \in \Omega_1, \ldots, \omega_n \in \Omega_n\}$$

Formally speaking, the n-tuple: $(\omega_1, \ldots, \omega_n)$ is defined to be the function f from $\{1, \ldots, n\}$ into $\Omega_1 \cup \cdots \cup \Omega_n$ given by: $f(j) = \omega_j$ for all $j = 1, \ldots, n$. In general, if $\{\Omega_j \mid j \in J\}$ is any collection of sets, then we define their *product* in a similar manner

$$\text{(P.1.12)} \qquad \prod_{j \in J} \Omega_j := \left\{ f : J \to \bigcup_{j \in J} \Omega_j \;\middle|\; f(j) \in \Omega_j \quad \forall j \in J \right\}$$

P.2 The Calculus of Sets

The events of probability are often formed by applying a series of the operations defined in (P.1) to certain basic events. So it is important to know the elementary calculus of these operations. In this section you will find the some of the most important rules of the calculus of sets.

The laws of de Morgan express how complementation acts on the other operations of set theory and they read as follows:

(P.2.1) $$\Big(\bigcup_{j\in J} A_j\Big)^c = \bigcap_{j\in J} A_j^c \quad , \quad \Big(\bigcap_{j\in J} A_j\Big)^c = \bigcup_{j\in J} A_j^c$$

(P.2.2) $$(A \setminus B)^c = A^c \cup B \quad , \quad (A\triangle B)^c = (A\cap B)\cup(A\cup B)^c$$

(P.2.3) $$(A^c)^c = A \quad , \quad \Omega^c = \emptyset \quad , \quad \emptyset^c = \Omega$$

(P.2.4) $$(\limsup A_n)^c = \liminf A_n^c \quad , \quad (\liminf A_n)^c = \limsup A_n^c$$

The distributive laws express how intersection and union act on the other operations of set theory and they read as follows:

(P.2.5) $$B\cap\Big(\bigcup_{j\in J} A_j\Big) = \bigcup_{j\in J}(B\cap A_j) \quad , \quad B\cap\Big(\bigcap_{j\in J} A_j\Big) = \bigcap_{j\in J}(B\cap A_j)$$

(P.2.6) $$B\cup\Big(\bigcup_{j\in J} A_j\Big) = \bigcup_{j\in J}(B\cup A_j) \quad , \quad B\cup\Big(\bigcap_{j\in J} A_j\Big) = \bigcap_{j\in J}(B\cup A_j)$$

(P.2.7) $$B\cap A^c = B\setminus A \quad , \quad B\cap(C\triangle D) = (B\cap C)\triangle(B\cap D)$$

(P.2.8) $$B\cup A^c = (A\setminus B)^c \quad , \quad B\cap(C\setminus D) = (B\cap C)\setminus(B\cap D)$$

(P.2.9) $$B\cap\limsup A_n = \limsup(B\cap A_n)$$

(P.2.10) $$B\cap\liminf A_n = \liminf(B\cap A_n)$$

(P.2.11) $$B\cup\limsup A_n = \limsup(B\cup A_n)$$

(P.2.12) $$B\cup\liminf A_n = \liminf(B\cup A_n)$$

(P.2.13) $$\bigcap_{i\in I}\Big(\bigcup_{j\in J} A_{ij}\Big) = \bigcup_{\sigma\in J^I}\Big(\bigcap_{i\in I} A_{i\sigma(i)}\Big) \quad , \quad \bigcup_{i\in I}\Big(\bigcap_{j\in J} A_{ij}\Big) = \bigcap_{\sigma\in J^I}\Big(\bigcup_{i\in I} A_{i\sigma(i)}\Big)$$

where J^I denotes the set of all functions $\sigma : I \to J$ from I into J. We have the following *distributive laws* for products:

(P.2.14) $$\bigcap_{j\in J}\Big(\prod_{i\in I} A_{ij}\Big) = \prod_{i\in I}\Big(\bigcap_{j\in J} A_{ij}\Big)$$

(P.2.15) $$\Big(\prod_{i=1}^{n} A_i\Big)\setminus\Big(\prod_{j=1}^{n} B_j\Big) = \bigcup_{j=1}^{n}\Big(\prod_{i=1}^{n} C_{ij}\Big)$$

where $C_{ij} = A_i \cap B_i$ if $i < j$, $C_{ii} = A_i \setminus B_i$, and $C_{ij} = A_i$ if $i > j$.

If $\omega \in \Omega$ is an element of Ω , then $\{\omega\}$ is the subset of Ω having ω as its only element. A subset of this form is called a *singleton.* You should not confuse the singleton $\{\omega\}$ with the element ω ; they are never equal. Note that $\omega \in \{\omega\}$, but by the so-called *axiom of foundation* of set theory [see (P.12) in the Appendix to Volume II], we have that $\omega \notin \omega$. For instance, the singleton $\{\emptyset\}$ consisting of the single element $\emptyset$ is not empty, since $\emptyset \in \{\emptyset\}$. Actually, in set theory one usually define the natural numbers in the following way:

$$0 := \emptyset \ , \ 1 := \{0\} \ , \ 2 := \{0,1\} \ , \ldots , \ n := \{0,1,\ldots,n-1\}$$

which means that the natural numbers can be constructed from nothing, that is, from the empty set and the empty set alone.

P.3 Images and Co-images

Let f be a function from the set Ω into the set Λ . If A is a subset of Ω and B is a subset of Λ , then we define *the image* $f(A)$ and the *the co-image* [also called *the inverse image*] $f^{-1}(B)$ as follows:

$$f(A) := \{f(\omega) \mid \omega \in A\} \quad , \quad f^{-1}(B) := \{\omega \in \Omega \mid f(\omega) \in B\}$$

Thus, $f(A)$ is the set all values in Λ that f assumes on A , and $f^{-1}(B)$ is the set of values in Ω that are mapped into B by f . In particular, if $\lambda \in \Lambda$, then $f^{-1}(\lambda)$ is the set of all solutions $\omega \in \Omega$ to the equation: $f(\omega) = \lambda$, or

$$f^{-1}(\lambda) := \{\omega \in \Omega \mid f(\omega) = \lambda\}$$

If $f(\Omega) = \Lambda$, we say that f is *surjective*; and if $f(\omega') \neq f(\omega'')$, whenever $\omega' \neq \omega''$, we say that f is *injective.* If f is injective and surjective, we say that f is *bijective.* Note that f is surjective if and only if $f^{-1}(\lambda) \neq \emptyset$ for all $\lambda \in \Lambda$, and f is injective if and only if $f^{-1}(\lambda)$ is either empty or a singleton for all $\lambda \in \Lambda$, and f is bijective if and only if $f^{-1}(\lambda)$ is a singleton for all $\lambda \in \Lambda$

The co-image respects all the operations of set theory that we have introduced in (P.1). More precisely we have the following equalities:

(P.3.1) $$f^{-1}\Big(\bigcup_{j\in J} B_j\Big) = \bigcup_{j\in J} f^{-1}(B_j) \quad , \quad f^{-1}\Big(\bigcap_{j\in J} B_j\Big) = \bigcap_{j\in J} f^{-1}(B_j)$$

(P.3.2) $$f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2) \quad , \quad f^{-1}(B^c) = f^{-1}(B)^c$$

(P.3.3) $$f^{-1}(\Lambda) = \Omega \quad , \quad f^{-1}(\emptyset) = \emptyset$$

But, in general, images only respects unions:

$$f\Big(\bigcup_{j\in J} A_j\Big) = \bigcup_{j\in J} f(A_j) \quad , \quad f(\emptyset) = \emptyset \tag{P.3.4}$$

$$f\Big(\bigcap_{j\in J} A_j\Big) \subseteq \bigcap_{j\in J} f(A_j) \quad , \quad f(A_1 \setminus A_2) \supseteq f(A_1) \setminus f(A_2) \tag{P.3.5}$$

However, if f is injective, then we have equality in (P.3.5). It is a common mistake to believe that image and co-image are inverse operations; this is **not** true, but we have

$$f\big(f^{-1}(B)\big) = B \cap f(\Omega) \subseteq B \tag{P.3.6}$$

$$f^{-1}(f(A)) \supseteq A \tag{P.3.7}$$

However, if f is surjective, then we have equality in (P.3.6), and if f is injective, then we have equality in (P.3.7).

Random variables are defined to be certain functions defined on the set Ω. However, it is often unnatural to consider random variables as functions. We shall, therefore, often drop the arguments of the function and we shall adopt the following shorthand notation:

$$\{f \in B\} := \{\omega |\, f(\omega) \in B\} = f^{-1}(B)$$

$$\{f \in B\ ,\ g \in C\} := \{\omega |\, f(\omega) \in B\ ,\ g(\omega) \in C\} = f^{-1}(B) \cap g^{-1}(C)$$

$$\{g \le a\} := \{\omega |\, g(\omega) \le a\}$$

$$\{f \le g\} := \{\omega |\, f(\omega) \le g(\omega)\}$$

etc.

P.4 The Real and Complex Numbers

The set of all *real numbers* is denoted by $\mathbf{R}$. We let $\mathbf{N} = \{1, 2, \ldots\}$, $\mathbf{N}_0 = \{0, 1, 2, \ldots\}$ and $\mathbf{Z} = \{0, \pm 1, \pm 2, \ldots\}$ denote the set of all *positive natural numbers*, all *non-negative numbers* and all *integers*, respectively. We let $\mathbf{R}_+ = \{x \in \mathbf{R} \mid x \ge 0\}$ denote the set of all *non-negative real numbers* and $\mathbf{C} = \{x + iy \mid x, y \in \mathbf{R}\}$ denotes the set of all *complex numbers*. If $z = x + iy \in \mathbf{C}$ is a complex number, then $\Re z := x$ and $\Im z := y$ denote *the real* and *the imaginary* part of z, $|z| := \sqrt{x^2 + y^2}$ denotes *the modulus* of z, and $\bar{z} := x - iy$ denotes *the complex conjugate* of z. We let $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$

denote *the extended real line*, and we extend the usual arithmetic on $\mathbf{R}$ to $\overline{\mathbf{R}}$ as follows:

$$a + \infty = \infty \quad \text{if} \quad -\infty < a \leq \infty$$

$$a + (-\infty) = a - \infty = -\infty \quad \text{if} \quad -\infty \leq a < \infty$$

$$a \cdot \infty = \infty \quad , \quad a \cdot (-\infty) = -\infty \quad \text{if} \quad 0 < a \leq \infty$$

$$a \cdot \infty = -\infty \quad , \quad a \cdot (-\infty) = \infty \quad \text{if} \quad -\infty \leq a < 0$$

$$0 \cdot \infty = 0 \cdot (-\infty) = 0$$

These conventions defines all arithmetic expressions except $\infty - \infty$, which we leave undefined. If $a, b \in \overline{\mathbf{R}}$, then we define *the open interval, the closed interval* and *the half-open intervals* as follows:

$$]a,b[:= \{x \in \overline{\mathbf{R}} \mid a < x < b\} \quad , \quad [a,b] := \{x \in \overline{\mathbf{R}} \mid a \leq x \leq b\}$$

$$]a,b] := \{x \in \overline{\mathbf{R}} \mid a < x \leq b\} \quad , \quad [a,b[:= \{x \in \overline{\mathbf{R}} \mid a \leq x < b\}$$

We let $\overline{\mathbf{R}}_+ = [0,\infty] = \mathbf{R}_+ \cup \{\infty\}$ denote *the extended positive real line.* Let x, y and $x_1, \ldots, x_n$ belong $\overline{\mathbf{R}}$. Then we define:

$$x \vee y := \max\{x,y\} \quad , \quad x_1 \vee \cdots \vee x_n := \max\{x_1, \ldots, x_n\}$$

$$x \wedge y := \min\{x,y\} \quad , \quad x_1 \wedge \cdots \wedge x_n := \min\{x_1, \ldots, x_n\}$$

$$x^+ := x \vee 0 \quad , \quad x^- := (-x) \vee 0 = (-x)^+$$

If f, g and $f_1, \ldots, f_n$ are functions from Ω into $\overline{\mathbf{R}}$, then we define the functions: $f \vee g$, $f_1 \vee \cdots \vee f_n$, $f \wedge g$, $f_1, \wedge \cdots \wedge f_n$, f^+ , f^- and $|f|$ in the following way:

$$(f \vee g)(\omega) := f(\omega) \vee g(\omega) \quad , \quad (f_1 \vee \cdots \vee f_n)(\omega) := f_1(\omega) \vee \cdots \vee f_n(\omega)$$

$$(f \wedge g)(\omega) := f(\omega) \wedge g(\omega) \quad , \quad (f_1 \wedge \cdots \wedge f_n)(\omega) := f_1(\omega) \wedge \cdots \wedge f_n(\omega)$$

$$f^+(\omega) := f(\omega)^+ \quad , \quad f^-(\omega) := f(\omega)^- \quad , \quad |f|(\omega) := |f(\omega)|$$

for all $\omega \in \Omega$. Clearly, we have the following relations:

(P.4.1) $\quad x = x^+ - x^- \quad , \quad |x| = x^+ + x^- = 2x^+ - x = x - 2x^-$

(P.4.2) $\quad x^- = (-x)^+ \quad , \quad x^+ = (-x)^-$

(P.4.3) $\quad x^+ = x \quad \text{if} \quad x \geq 0 \quad , \quad x^+ = 0 \quad \text{if} \quad x \leq 0$

(P.4.4) $\quad x^- = 0 \quad \text{if} \quad x \geq 0 \quad , \quad x^- = -x \quad \text{if} \quad x \leq 0$

whenever $x \in \overline{\mathbf{R}}$.

Let Ω be a set. Then we say that Ω is *countable* if there exists a surjective function f from $\mathbf{N}$ onto Ω . Setting $\omega_j = f(j)$, we see that Ω is countable if and only if Ω can be *enumerated*: $\Omega = \{\omega_1, \omega_2, \ldots\}$. It is well-known that the set $\mathbf{Q}$ of all rational numbers is countable, but the set $\mathbf{R}$ of all real numbers is not countable. Clearly, every subset of a countable set is itself countable. Moreover, we have the following:

(P.4.5) A countable union of countable sets is countable.

(P.4.6) A finite product of countable sets is countable.

Let $a_1, a_2, \ldots$ be a sequence in $\overline{\mathbf{R}}$. If the partial sums $s_k = a_1 + \cdots + a_k$ is well-defined for all $k \in \mathbf{N}$ [that is, if either $a_j \neq \infty$ for all j , or if $a_j \neq -\infty$ for all j] and converge to a number s in $\overline{\mathbf{R}}$, then we say that *the series* $\sum_1^\infty a_j$ *is convergent with sum* s and we write $s = \sum_1^\infty a_j$, that is,

$$\sum_{n=1}^{\infty} a_n = s \iff s = \lim_{k\to\infty} (a_1 + \cdots + a_k)$$

If the series $\sum_1^\infty a_j$ is convergent to a finite sum, then we say that $\sum_1^\infty a_j$ is *summable*.

Let $\{a_j \mid j \in J\}$ be an arbitrary family of non-negative extended real numbers, that is, $a_j \in \overline{\mathbf{R}}_+$ for all $j \in J$. Then we define *the unordered sum* of $\{a_j \mid j \in J\}$ in the following way:

$$\sum_{j\in J} a_j = \sup_{\gamma\in\Gamma}\{ \sum_{j\in\gamma} a_j \}$$

where Γ is the set of all finite non-empty subsets of J . Similarly, if $\{a_j \mid j \in J\}$ is an arbitrary family of numbers in $\overline{\mathbf{R}}$, then we define *the unordered sum* of $\{a_j \mid j \in J\}$ in the following way:

$$\sum_{j\in J} a_j = \sum_{j\in J} a_j^+ - \sum_{j\in J} a_j^-$$

provided that the right-hand side is well-defined (that is, not of the form $\infty - \infty$). It is easily seen that the unordered sum $s = \sum_{j\in J} a_j$ exists and is finite if and only if

(P.4.7) $\forall \varepsilon > 0\ \exists \sigma \in \Gamma$ so that $|s - \sum_{j\in\gamma} a_j| < \varepsilon \quad \forall \gamma \in \Gamma$ with $\gamma \supseteq \sigma$

where Γ is the set of all finite non-empty subsets of J .

Let Ω be a set, and let A be a subset of Ω . Then we define *the indicator function* of A , denoted 1_A , in the following way:

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in \Omega \setminus A \end{cases}$$

Then it is easily checked that we have:

(P.4.8) $\quad 1_{A\cup B} = 1_A \vee 1_B = 1_A + 1_B - 1_A \cdot 1_B$

(P.4.9) $\quad 1_{A\cup B} = 1_A + 1_B \quad \text{if} \quad A \cap B = \emptyset$

(P.4.10) $\quad 1_{A\cap B} = 1_A \wedge 1_B = 1_A \cdot 1_B$

(P.4.11) $\quad 1_{A\setminus B} = (1_A - 1_B)^+ = 1_A - 1_A \cdot 1_B$

(P.4.12) $\quad 1_{A^c} = 1 - 1_A \quad , \quad 1_{A\triangle B} = |1_A - 1_B|$

(P.4.13) $\quad 1_{A_1\times\cdots\times A_n}(\omega_1, \cdots, \omega_n) = 1_{A_1}(\omega_1) \cdot 1_{A_2}(\omega_2) \cdots 1_{A_n}(\omega_n)$

whenever A and B are subsets of Ω , and A_j is a subset of the set Ω_j for all $j = 1, \ldots, n$. Moreover, if $A_1, A_2, \ldots$ are subsets of Ω and

$$S = \bigcup_{n=1}^{\infty} A_n \quad , \quad T = \bigcap_{n=1}^{\infty} A_n \quad , \quad U = \limsup A_n \quad , \quad V = \liminf A_n$$

then we have

(P.4.14) $\quad 1_S = \sup_{n\geq 1} 1_{A_n}$

(P.4.15) $\quad 1_S = \sum_{n=1}^{\infty} 1_{A_n} \quad \text{if} \quad A_i \cap A_j = \emptyset \quad \forall 1 \leq i \neq j < \infty$

(P.4.16) $\quad 1_T = \prod_{n=1}^{\infty} 1_{A_n} = \inf_{n\geq 1} 1_{A_n}$

(P.4.17) $\quad 1_U = \limsup_{n\to\infty} 1_{A_n}$

(P.4.18) $\quad 1_V = \liminf_{n\to\infty} 1_{A_n}$

Note that (P.4.17+18) show the connection between lim sup and lim inf for real numbers and lim sup and lim inf for sets. Recall that lim inf and lim sup of a sequence $a_1, a_2, \ldots \in \overline{\mathbf{R}}$ are defined by

$$\liminf_{n\to\infty} a_n := \sup_{j\geq 1} \inf_{n\geq j} a_n \quad , \quad \limsup_{n\to\infty} a_n := \inf_{j\geq 1} \sup_{n\geq j} a_n$$

and that we have the following rules of calculus for $\limsup$ and $\liminf$:

$$\text{(P.4.19)} \quad \begin{aligned} \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n &\le \liminf_{n\to\infty} (a_n + b_n) \le \liminf_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \\ &\le \limsup_{n\to\infty} (a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \end{aligned}$$

$$\text{(P.4.20)} \quad \liminf_{n\to\infty} (-a_n) = -\limsup_{n\to\infty} a_n \quad , \quad \limsup_{n\to\infty} (-a_n) = -\liminf_{n\to\infty} a_n$$

$$\text{(P.4.21)} \quad \liminf_{n\to\infty} a_n \le \limsup_{n\to\infty} a_n$$

$$\text{(P.4.22)} \quad a = \lim_{n\to\infty} a_n \iff \liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = s$$

$$\text{(P.4.23)} \quad a = \lim_{n\to\infty} a_n \iff \liminf_{n\to\infty} a_n \ge s \ge \limsup_{n\to\infty} a_n$$

where it is assumed the relevant sums in (P.4.19) exist, that is, not of the form $\infty - \infty$.

P.5 Metric Spaces

Recall that a *metric space* (S, d) is a set S and a function d from $S \times S$ into $\mathbf{R}_+$ satisfying

$$\text{(P.5.1)} \quad d(x,y) = d(y,x) \quad \text{and} \quad d(x,x) = 0 \qquad \forall x, y \in S$$

$$\text{(P.5.2)} \quad d(x,y) \le d(x,z) + d(z,y) \qquad \forall x, y, z \in S$$

$$\text{(P.5.3)} \quad d(x,y) > 0 \qquad \forall x, y \in S \quad \text{with} \quad x \ne y$$

A function $d : S \times S \to \mathbf{R}_+$ satisfying (P.5.1)–(P.5.3) is called a *metric*. The number $d(x,y)$ is interpreted as *the distance* between x and y . Hence (P.5.2) is just the well-known *triangle inequality* dating back to Euclid ($\approx$ 300 B.C.): "The length of one side in a triangle is less the sum of the lengths of the two others." If $d : S \times S \to \overline{\mathbf{R}}_+$ satisfies (P.5.1) and (P.5.2), i.e. if two different points may have distance 0 or ∞ , then we say that d is a *pseudo-metric* and we say that (S, d) is a *pseudo-metric space*. In Section (5.22) we shall meet a series of pseudo-metric spaces, which are not metric spaces.

Suppose that (S, d) is a pseudo-metric space. Then we define *the open ball* respectively *the closed ball* with *center* at $x_0 \in S$ and *radius* $r \in \mathbf{R}_+$ to be the sets

$$b(x_0, r) := \{x \in S \mid d(x, x_0) < r\} \quad , \quad b[x_0, r] := \{x \in S \mid d(x, x_0) \le r\}$$

A set $G \subseteq S$ is said to be *open* if for every $x_0 \in G$ there exists $r > 0$ such that $b(x_0, r) \subseteq G$. A set $F \subseteq S$ is said to be *closed* if the complement F^c is open. A sequence $\{x_n\}$ in S is said to be *convergent* with *limit* x if $\lim_{n\to\infty} d(x, x_n) = 0$ and if so, we write:

$$x_n \to x \quad \text{or} \quad x = \lim_{n\to\infty} x_n$$

It is well-known that $F \subseteq S$ is closed if and only if F satisfies the following condition:

$$x_n \in F \quad \forall n \geq 1 \quad \text{and} \quad x_n \to x \quad \Rightarrow \quad x \in F$$

In a metric space a sequence $\{x_n\}$ has at most one limit, but in a pseudo-metric space it may have several limits. More precisely if $x_n \to x$, then $x_n \to y$ if and only if $d(x, y) = 0$. If $A \subseteq S$, then $\mathrm{cl}(A)$ denotes *the closure* of the set A , that is, $x \in \mathrm{cl}(A)$ if and only if there exists a sequence $\{x_n\} \subseteq A$ such that $x_n \to x$. Then $\mathrm{cl}(A)$ is the smallest closed set containing A . Similarly, we let $\mathrm{int}(A)$ denote *the interior* of the set A , that is, $x \in \mathrm{int}(A)$ if and only if there exists $r > 0$ such that $b(x, r) \subseteq A$. Then $\mathrm{int}(A)$ is the largest open set contained in A . We let $\mathrm{bd}(A) := \mathrm{cl}(A) \setminus \mathrm{int}(A)$ denote *the boundary* of the set A . If A and B are subset of S , then we say that B is *dense* in A if $B \subseteq A \subseteq \mathrm{cl}(B)$, and we say that A is *separable* if there exists a countable set $Q \subseteq A$ such that Q is dense in A . It is easy to see that we have

(P.5.4) Every subset of a separable set is separable. The closure of a separable set is separable. A countable union of separable sets is separable

In $\mathbf{R}^k$ we have the usual *euclidean metric*:

$$\|x - y\| = \left\{ \sum_{j=1}^{k} (x_j - y_j)^2 \right\}^{1/2}$$

for $x = (x_1, \ldots, x_n) \in \mathbf{R}^k$ and $y = (y_1, \ldots, y_n) \in \mathbf{R}^k$. Since the set $\mathbf{Q}^k$ of all vectors with rational coordinates is a countable dense subset of $\mathbf{R}^k$, we see that $\mathbf{R}^k$ is separable.

A sequence $\{x_n\}$ in the pseudo-metric space (S, d) is called a *Cauchy sequence* if $d(x_n, x_m) \to 0$ as n and m tends to infinity, that is, if

$$\forall \varepsilon > 0 \; \exists n_0 \quad \text{so that} \quad d(x_n, x_m) < \varepsilon \quad \forall n, m \geq n_0$$

It is well-known that every convergent sequence is a Cauchy sequence, but the converse is false, in general. If $F \subseteq S$, then we say that F is *complete* if every Cauchy sequence in F converges to a point in F . A complete separable

metric space is called a *polish space*. It is well-known that $\mathbf{R}^k$ with its euclidean metric is a polish space.

The notion of a Cauchy sequence was invented by Augustin Louis Cauchy (1789–1857) in his famous book *Cours d'Analyse*. Cauchy shows with painstaking care that every convergent sequence of real numbers is a Cauchy sequence, and then he writes in a clause: "le contraire est èvidente" (the contrary is evident). However, the contrary is by no means evident, moreover, it was beyond the powers of Cauchy because it depends on a precise definition of the real numbers which simply was not available at the time Cauchy wrote his book.

Let S_0 be a separable subset of the pseudo-metric space (S, d) and let D be a countable dense subset of S_0 . If $\mathcal{G}$ denotes the paving of all finite unions of open balls with center in D radius in the positive rational numbers, then $\mathcal{G}$ is a countable $(\cup f)$-stable paving consisting of open subsets of S such that for every open set $U \subseteq S$ there exists an increasing sequence $\{G_n\} \subseteq \mathcal{G}$ satisfying:

$$\text{(P.5.5)} \qquad S_0 \cap U = \bigcup_{n=1}^{\infty} (G_n \cap S_0) \subseteq \bigcup_{n=1}^{\infty} G_n \subseteq U$$

A subset C of the pseudo-metric space (S, d) is called *precompact* (some textbooks uses the terminology *totally bounded*) if for every $\varepsilon > 0$ there exists finitely many balls $B_j = b(x_j, \varepsilon)$, $1 \leq j \leq n$, of radius ε such that $C \subseteq B_1 \cup \cdots \cup B_n$. Thus, C is precompact if we can cover C by finitely many balls of radius ε no matter how small ε may be. A set $K \subseteq S$ is called *compact* if for every family $\{G_j \mid j \in J\}$ of **open** subsets of S with $K \subseteq \cup_{j \in J} G_j$, there exists a **finite** set $\gamma \subseteq J$ such that $K \subseteq \cup_{j \in \gamma} G_j$. Thus, K is compact if every open covering of K contains a finite subfamily which covers K . Since the set of all open balls of radius ε is a covering of S , we have that every compact set is precompact. It is easy to verify that every precompact is separable. One of the fundamental theorems of analysis states that the following three statements are equivalent:

(P.5.6) K is compact

(P.5.7) For every sequence $\{x_n\} \subseteq K$, there exists $x \in K$ and integers $1 \leq \sigma(1) < \sigma(2) < \ldots$ such that $x_{\sigma(n)} \to x$

(P.5.8) K is precompact and complete

In a metric space every complete set is closed and, thus, every compact subset of a metric space is closed. However, in a pseudo-metric space there may exist compact subsets which are **not** closed.

Suppose that d_0 and d_1 are pseudo-metrics on S . Then we say that d_0 and d_1 are *topologically equivalent* if for every sequence $\{x_n\} \subseteq S$ and every

$x \in S$, we have

$$\lim_{n\to\infty} d_0(x_n, x) = 0 \quad \Leftrightarrow \quad \lim_{n\to\infty} d_1(x_n, x) = 0$$

Suppose that d_0 and d_1 are topologically equivalent. Then d_0 and d_1 induces the same convergent sequences, the same open sets, the same closed sets and the same compact sets on S , but not necessarily the same Cauchy sequences nor the same precompact sets. For instance, the two metrics

$$d_0(x, y) = |x - y| \quad , \quad d_1(x, y) = |\arctan x - \arctan y|$$

on $\mathbf{R}$ are topologically equivalent such that $(\mathbf{R}, d_0)$ is complete and non-precompact, but $(\mathbf{R}, d_1)$ is precompact and non-complete. Let S_0 be a subset of the pseudo-metric space (S, d) , then the restriction d_0 of d to $S_0 \times S_0$ is evidently a pseudo-metric on S_0 which is called *the induced d-metric* on S_0 . Let (L, ρ) and (S, d) be two pseudo-metric spaces. Then we say that (L, ρ) is *imbedded* in (S, d) if $L \subseteq S$ and the induced d-metric on L is topological equivalent to ρ . Note that if (L, ρ) is imbedded in (S, d) , then a set $A \subseteq L$ is open [respectively closed] in (L, ρ) if and only if there exists an open [respectively a closed] set B in (S, d) such that $A = L \cap B$. If (L, ρ) is imbedded in a compact [respectively complete] pseudo-metric space (M, δ) , then we say that (M, δ) is a *compactification* [respectively a *completion*] of (L, ρ) . One of the fundamental theorems of metric spaces states that

(P.5.9) Every pseudo-metric space admits a completion, and every separable pseudo-metric space admits a compactification.

For instance, let $\overline{\mathbf{R}}^k$ be the k-fold product of the extended real line and define the metric ρ_k by

(P.5.10) $$\rho_k(x, y) = \max_{1\le j\le k} |\arctan x_j - \arctan y_j|$$

whenever $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$ belongs to $\overline{\mathbf{R}}^k$ where we set $\arctan(\pm\infty) = \pm\frac{\pi}{2}$. Then $(\overline{\mathbf{R}}^k, \rho_k)$ is a compact metric space, and if $S \subseteq \mathbf{R}^k$ has the induced euclidean metric, then $(\overline{\mathbf{R}}^k, \rho_k)$ is a compactification of S .

Let $(S_1, d_1), \ldots, (S_k, d_k)$ be pseudo-metric spaces. Then we define *the product metric*, denoted d^k , on the product space $S_1 \times \cdots \times S_k$ in the following way:

(P.5.11) $$d^k(x, y) = \sqrt{\sum_{j=1}^{k} d_j(x_j, y_j)^2}$$

whenever $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$ belong to the product space $S_1 \times \cdots \times S_k$. Similarly, if $(S_1, d_1), (S_2, d_2), \ldots$ is a sequence of pseudo-metric spaces, then we define *the product metric*, denoted d^∞ , on the infinite product space $\prod_{j=1}^\infty S_j$ in the following way:

$$\text{(P.5.12)} \qquad d^\infty(x, y) = \sum_{j=1}^{\infty} 2^{-j} \arctan d_j(x_j, y_j)$$

whenever $x = (x_j)$ and $y = (y_j)$ belong to the infinite product space $\prod_{j=1}^\infty S_j$. It is easy checked that a sequence in a finite or countable product of pseudo-metric spaces converges in the product metric if and only if all the coordinates converges in their respective spaces. Moreover, it can be shown that we have

(P.5.13) A finite or countable product of separable, respectively complete, respectively compact spaces is itself separable, respectively complete, respectively compact.

The separable and complete case is straightforward to prove. The compact case, which is called *Tychonov's theorem*, follows easily from (P.5.7) and the well-known *Cantor's diagonalization argument.*

Let (S, d) and (T, δ) be pseudo-metric spaces and let $f : S \to T$ be a function. If $a \in S$, then we say that f is *continuous* at a if

$$\text{(P.5.14)} \qquad \forall \varepsilon > 0 \; \exists \rho > 0 \quad \text{so that} \quad \delta(f(x), f(a)) < \varepsilon \qquad \forall x \in b(a, \rho)$$

or, equivalently if $f(x_n) \to f(a)$, whenever $x_n \to a$. We say that f is *continuous* if f is continuous at a for all $a \in S$. It is well-known that f is continuous if and only if $f^{-1}(B)$ is open [respectively closed] in S whenever B is open [respectively closed] in T .

Let (S, d) be a metric space and let $A \subseteq S$. Then we say that A is *analytic* if there exists a polish space (T, δ) and a continuous function $f : T \to S$ such that $A = f(T)$ [recall that a polish space is a separable complete metric space]. In other words, A is analytic if and only if A is a continuous image of a polish space. Evidently, we have that every polish space is analytic, and it is easily verified that every analytic space is separable. Let (S, d) and (T, δ) be two metric spaces, let A be an analytic subset of S , and let $f : S \to T$ be a Borel function [see (1.39)]. Then it can be shown that we have the following:

(P.5.15) $A \cap B$ is analytic for all $B \in \mathcal{B}(S)$

(P.5.16) $f(A)$ is an analytic subset of T

(P.5.17) If S is analytic, then $B \in \mathcal{B}(S)$ if and only if B and B^c are analytic.

(P.5.18) A is μ-measurable for every Borel measure μ on S .

(P.5.19) If S is analytic, then every σ-finite Borel measure μ on S is a Radon measure.

where $\mathcal{B}(S)$ is the Borel σ-algebra on S [see (1.8), (1.22) and (1.34)].

The notion of analytic sets dates back to an idea of M.M. Souslin (1916) [see (1.47)]. Note that if B is a Borel subset of $\mathbf{R}^2$, then B is analytic by (P.5.15) and so its projection on the first axis is analytic by (P.5.16), but not necessarily a Borel set. However, by (P.5.18) it is μ-measurable for any Borel measure μ on $\mathbf{R}$ which, from a measure theoretic point of view, is just as good. Thus, the mistake of Lebesgue [see (1.47)] is not really serious. Also note that (P.5.16) and (P.5.18) show that the image catastrophe is after all no catastrophe. By (P.5.15) we have that any Borel subset of a polish space is analytic, and the analyticity condition is so weak that it is very hard to construct a separable non-analytic metric space. But they do exist; for instance, the set $F(D)$ defined in (1.17) is analytic, but $\mathbf{R} \setminus F(D)$ is non-analytic.

Let F be a non-empty subset of the pseudo-metric space (S, d) . Then $d(x, F)$ denotes *the minimal distance* from x to a point in F , that is,

$$d(x, F) := \inf_{y \in F} d(x, y) \qquad \forall x \in S \, \forall F \subseteq S \, , \; F \neq \emptyset$$

Let $x, y \in S$ and let $z \in F$. Then by the triangle inequality we have $d(x, F) \leq d(x, z) \leq d(x, y) + d(y, z)$, taking infimum over $z \in F$, we get

(P.5.20) $$|d(x, F) - d(y, F)| \leq d(x, y) \qquad \forall x, y \in S \, \forall F \subseteq S \, , \; F \neq \emptyset$$

In particular, we see that the function: $x \rightsquigarrow d(x, F)$ is continuous and by definition of the closure $\mathrm{cl}(F)$ we find

(P.5.21) $$\mathrm{cl}(F) = \{x \in S \mid d(x, F) = 0\}$$

(P.5.22) $$F \text{ is closed} \quad \Leftrightarrow \quad F = \{x \in S \mid d(x, F) = 0\}$$

Let (S, d) be a pseudo-metric space, and let $f : S \to \overline{\mathbf{R}}$ be a function. Then we say that f is *lower semicontinuous* at $a \in S$ if

(P.5.23) $$\forall t < f(a) \; \exists \delta > 0 \quad \text{so that} \quad f(x) > t \qquad \forall x \in b(a, \delta)$$

or, equivalently if

(P.5.24) $$\liminf_{n \to \infty} f(x_n) \geq f(a) \qquad \text{whenever} \quad x_n \to a$$

Similarly, we say that f is *upper semicontinuous* at a if

(P.5.25) $\forall t > f(a)\ \exists\, \delta > 0$ so that $f(x) < t \quad \forall x \in b(a, \delta)$

or, equivalently if

(P.5.26) $\limsup_{n\to\infty} f(x_n) \le f(a)$ whenever $x_n \to a$

If f is lower respectively upper semicontinuous at all points $a \in S$, then we say that f is *lower* respectively *upper semicontinuous*. It is easily checked that we have the following:

(P.5.27) f is lower [respectively upper] semicontinuous if and only if $\{f > t\}$ [respectively $\{f < t\}$] is open in S for all $t \in \mathbf{R}$

(P.5.28) f is lower semicontinuous at a if and only if $(-f)$ is upper semicontinuous at a

(P.5.29) f is continuous at a if and only if f is lower and upper semicontinuous at a

(P.5.30) 1_A is lower [respectively upper] semicontinuous if and only if A is open [respectively closed] in S. Moreover if $K \subseteq S$ is compact and $f : S \to \overline{\mathbf{R}}$ is lower [respectively upper] semicontinuous, then f attains its minimum [respectively maximum] on the set K

Comparing definitions (P.5.24) and (P.5.26) with the definition of continuity you see that lower and upper semicontinuity are simply the lower and upper half of the continuity notion. In our discussion of convergence in law [see (9.2)] we shall need the following approximation result for lower semicontinuous functions:

Approximation of lower semicontinuous functions. *Let (S, d) be a pseudometric space and let $g : S \to \mathbf{R}$ be a function which is bounded below, that is, $m = \inf_{x\in S} g(x) > -\infty$. Let*

(P.5.31) $g_k(x) = \inf\ \{k \wedge g(y) + k \cdot d(x, y) \mid y \in S\} \quad \forall x \in S\ \forall k \ge 1$

Then we have

(P.5.32) $m \wedge k \le g_{k-1}(x) \le g_k(x) \le k \wedge g(x) \quad \forall x \in S\ \forall k \ge 2$

(P.5.33) g_k *is bounded for all* $k \ge 1$

(P.5.34) $|g_k(x) - g_k(y)| \le k\, d(x, y) \quad \forall x, y \in S\ \forall k \ge 1$

(P.5.35) $g_k(a) \uparrow g(a)$ *if* g *is lower semicontinuous at* a

Remark: The functions g_k as defined above may look rather peculiar. However, there is a perfectly good reason for defining g_k by (P.5.31) which arises from the so-called *railroad track problem.* Suppose that we want to build a railroad track between two points in a landscape, and suppose that the landscape is given as the graph of the function g . There is a maximal height, say $\alpha > 0$, at which the engine works properly, and there is a maximal slope, say $\beta > 0$, that the train may ascend. Thus, in order to build the railroad track we have reshape to the landscape, let the reshaped landscape be given as the graph of the function f , then the requirement of maximal height α and maximal slope β leads to the following two requirements on the function f :

(P.5.36) $f(x) \leq \alpha \wedge g(x) \qquad \forall x \in S$

(P.5.37) $|f(x) - f(y)| \leq \beta\, d(x,y) \qquad \forall x, y \in S$

Suppose that we want to minimize the construction work for our railroad track. Then we seek the largest possible function f which satisfies (P.5.36) and (P.5.37). In the proof below it will be shown that the function

(P.5.38) $f_{\alpha\beta}(x) = \inf\, \{\alpha \wedge g(y) + \beta \cdot d(x,y) \mid y \in S\} \qquad \forall x \in S$

satisfies (P.5.36) and (P.5.37). Moreover, it is easy to verify that for any function f satisfying (P.5.36) and (P.5.37), we have that $f(x) \leq f_{\alpha\beta}(x)$ for all $x \in S$. Thus, the function g_k is simply the solution to the railroad track problem with $\alpha = \beta = k$.

Proof: Clearly, we have that $m \wedge k \leq g_{k-1}(x) \leq g_k(x)$ and choosing $y = x$ in the infimum defining g_k , we see that $g_k(x) \leq k \wedge g(x)$. Hence, (P.5.32) and (P.5.33) are proved. Let $x, y, z \in S$. Then

$$g_k(x) \leq k \wedge g(z) + k\, d(x,z) \leq k \wedge g(z) + k\, d(y,z) + k\, d(x,y)$$

Taking infimum over $z \in S$, we get $g_k(x) \leq g_k(y) + k\, d(x,y)$. Interchanging x and y , we obtain (P.5.34). Now suppose that g is lower semicontinuous at a and let $t < g(a)$. Then there exists $\delta > 0$ such that $g(y) > t$ for all $y \in b(a,\delta)$, and so we find

$$k \wedge g(y) + k\, d(x,y) \geq k \wedge t \qquad \text{if } x \in b(a,\delta)$$

$$k \wedge g(y) + k\, d(x,y) \geq m + k\,\delta \qquad \text{if } x \notin b(a,\delta)$$

Hence, if $k \geq |t| + \frac{|t-m|}{\delta}$, then $g_k(a) > t$. Thus, (P.5.35) follows from (P.5.32). □

P.6 Function Spaces

Let Ω and S be sets. Then S^{Ω} denotes the set of all functions from Ω into S. Let $\{f_n\}$ be sequence of functions from Ω into the pseudo-metric space (S, d). Then we say that f_n *converges pointwise* to the function $f \in S^{\Omega}$ and we write $f_n \to f$ if $\{f_n(\omega)\}$ converges to $f(\omega)$ for all $\omega \in \Omega$, that is,

$$\text{(P.6.1)} \qquad f_n \to f \quad \Leftrightarrow \quad f(\omega) = \lim_{n\to\infty} f_n(\omega) \quad \forall \omega \in \Omega$$

If $f, g \in S^{\Omega}$ and $F \subseteq \Omega$, then we define the *sup-metric* over F as

$$\text{(P.6.2)} \qquad d_F(f, g) := \sup_{\omega \in F} d(f(\omega), g(\omega))$$

If $\{f_n\}$ is a sequence in S^{Ω}, then we say that f_n *converges uniformly* on F and we write $f_n \rightrightarrows f$ *uniformly on* F if $d_F(f_n, f) \to 0$ as $n \to \infty$; that is,

$$\text{(P.6.3)} \qquad f_n \rightrightarrows f \text{ uniformly on } F \quad \Leftrightarrow \quad \lim_{n\to\infty} d_F(f_n, f) = 0$$

Let $\{f_n\}$ be a sequence of functions from Ω into $\overline{\mathbf{R}}$. Then we say that $\{f_n\}$ is *increasing* [respectively *decreasing*] if $f_1(\omega) \le f_2(\omega) \le \dots$ [respectively if $f_1(\omega) \ge f_2(\omega) \ge \dots$] for all $\omega \in \Omega$ and we write

$$\text{(P.6.4)} \qquad f_n \uparrow f \text{ if } \{f_n\} \text{ is increasing and } f_n \to f$$

$$\text{(P.6.5)} \qquad f_n \downarrow f \text{ if } \{f_n\} \text{ is decreasing and } f_n \to f$$

If $\{f_n\}$ is either increasing or decreasing, then we say that $\{f_n\}$ is *monotone*. Let $F \subseteq \Omega$, then we set:

$$\|f\|_F := \sup_{\omega \in F} |f(\omega)| \quad \forall f \in \overline{\mathbf{R}}^{\Omega} \quad , \quad B(\Omega) := \left\{ f \in \mathbf{R}^{\Omega} \;\middle|\; \|f\|_{\Omega} < \infty \right\}$$

Thus, $B(\Omega)$ is the set of all bounded functions from Ω into $\mathbf{R}$. Note that $\|f - g\|_{\Omega}$ defines a metric on $B(\Omega)$ such that convergence in the metric space $(B(\Omega), \|\cdot\|_{\Omega})$ coincides with uniform convergence on Ω. It is well-known and easy to prove that we have

$$\text{(P.6.6)} \qquad (B(\Omega), \|\cdot\|_{\Omega}) \text{ is a complete metric space.}$$

It can be shown that $(B(\Omega), \|\cdot\|_{\Omega})$ is separable if and only if Ω is a finite set.

Let (S, d) be a pseudo-metric space, then $C(S)$ denotes the set of all bounded continuous functions from S into $\mathbf{R}$:

$$\text{(P.6.7)} \qquad C(S) := \{f \in B(S) \mid f \text{ is continuous}\}$$

Since a uniform limit of continuous functions is continuous, we see that $C(S)$ is a closed subset of $B(S)$. Hence, we have

(P.6.8) $(C(S), \|\cdot\|_S)$ is a complete metric space.

It can be shown that $(C(S), \|\cdot\|_S)$ is separable if and only if S is compact. A standard compactness argument gives *Dini's theorem* which you probably know for continuous functions:

(P.6.9) If $f_n \in \mathbf{R}^S$ is upper semicontinuous and $f_n \downarrow 0$, then $f_n \rightrightarrows 0$ uniformly on K for all compact sets $K \subseteq S$

Recall that by Tychonov's theorem [see (P.5.13)] we have that S^Ω is compact in the product metric whenever S is compact and Ω is countable or finite. Thus, we have the following:

(P.6.10) If S is compact and Ω is countable or finite, then for every sequence $\{f_n\} \subseteq S^\Omega$ there exist integers $\sigma(1) < \sigma(2) < \ldots$ and a function $f \in S^\Omega$ such that $f_{\sigma(n)} \to f$

P.7 Relations and Orderings

A *relation* on the set Ω is a subset R of the product set $\Omega \times \Omega$. If R is a relation on Ω , then we write $\omega' R \omega''$ [read: ω' *is in relation* R *to* ω''] if $(\omega', \omega'') \in R$ and we say that

(P.7.1) R is *reflexive* if $\omega R \omega$ for all $\omega \in \Omega$

(P.7.2) R is *symmetric* if $\omega' R \omega''$ implies $\omega'' R \omega'$

(P.7.3) R is *transitive* if $\omega' R \omega''$ and $\omega'' R \omega'''$ implies $\omega' R \omega'''$

(P.7.4) R is *antisymmetric* if $\omega' R \omega''$ and $\omega'' R \omega'$ implies $\omega' = \omega''$

If R is reflexive, symmetric, and transitive, then we say that R is an *equivalence relation*; if R is reflexive and transitive, then we say that R is a *preordering*; and if R is reflexive, transitive and antisymmetric, then we say that R is a *partial ordering*. If R is a partial ordering such that either $\omega' R \omega''$ or $\omega'' R \omega'$ for every $\omega', \omega'' \in \Omega$, then we say that R is a *linear ordering*. For instance, if $f : \Omega \to \mathbf{R}$ and

$$\omega' R \omega'' \quad \Leftrightarrow \quad f(\omega') = f(\omega'')$$

$$\omega' Q \omega'' \quad \Leftrightarrow \quad f(\omega') \leq f(\omega'')$$

then R is an equivalence relation and Q is a preordering. Moreover, if f is injective, then Q is a linear ordering.

A partial ordering is usually denoted $\preceq$. Let $\preceq$ be a partial ordering on Ω, and let $\Upsilon \subseteq \Omega$. Then $\preceq$ or rather $(\Upsilon \times \Upsilon) \cap \preceq$ is a partial ordering on Υ. If $\preceq$ is a linear ordering on Υ, then we say that Υ is a *linearly ordered* subset of Ω. If Υ is a linearly ordered subset of Ω such that $\Upsilon = \Gamma$ whenever Γ is a linear ordered subset of Ω with $\Gamma \supseteq \Upsilon$, then we say that Υ is a *maximal linearly ordered* subset of Ω. An element $\theta \in \Upsilon$ is called a *minimum* of Υ if $\theta \preceq v$ for all $v \in \Upsilon$, and an element $\tau \in \Upsilon$ is called a *minimal element* of Υ if $v \in \Upsilon$ and $v \preceq \tau$ implies $v = \tau$. We define a *maximum* and a *maximal element* of Υ in the same way. Clearly, every minimum is a minimal element, but the converse is false in general. A linear ordering $\preceq$ is called a *well-ordering* if every non-empty subset admits a minimum. For instance, usual ordering on $\mathbf{N}$ is a well-ordering, but the usual ordering on $\mathbf{R}$ is not.

Let R be an equivalence relation. Then we often write $\omega' \equiv \omega''(\text{mod } R)$ [read: ω' *is equivalent to* ω'' *modulo* R] rather than $\omega' R \omega''$. We define *the equivalence class* containing ω_0, denoted $R[\omega_0]$, to be the set

$$R[\omega_0] := \{\omega \in \Omega \mid \omega_0 \equiv \omega(\text{mod } R)\} \tag{P.7.5}$$

The set of all equivalence classes is a paving on Ω which is called *the quotient space*, and is denoted Ω/R. It is well-known and easy to prove that we have

$$R[\omega_0] \cap R[\omega_1] \neq \emptyset \quad \Leftrightarrow \quad \omega_0 \equiv \omega_1(\text{mod } R) \tag{P.7.6}$$

$$R[\omega_0] \cap R[\omega_1] \neq \emptyset \quad \Leftrightarrow \quad R[\omega_0] = R[\omega_1] \tag{P.7.7}$$

$$\omega \in R[\omega] \qquad \forall \omega \in \Omega \tag{P.7.8}$$

Thus, the set of equivalence classes forms a disjoint partition of the set Ω.

P.8 Linear Algebra

The n-dimensional euclidean space is denoted $\mathbf{R}^n$ and the set of all $(n \times k)$-*matrices* is denoted $\mathbf{R}^{n \times k}$. If $\alpha = \{\alpha_{ij}\} \in \mathbf{R}^{n \times k}$ and $\beta = \{\beta_{ij}\} \in \mathbf{R}^{k \times q}$, then $\alpha\beta = \{\sum_\nu \alpha_{i\nu}\beta_{\nu j}\} \in \mathbf{R}^{n \times q}$ denotes the *matrix product* of α and β and $\alpha^t = \{\alpha_{ji}\} \in \mathbf{R}^{k \times n}$ denotes *the transposed matrix* of α. We shall always consider vectors in $\mathbf{R}^n$ as *row vectors*, that is, as $(1 \times n)$-matrices. Thus, we identify $\mathbf{R}^n$ with $\mathbf{R}^{1 \times n}$. Hence, if $x \in \mathbf{R}^n$, then $x^t \in \mathbf{R}^{n \times 1}$ is a *column vector*. If $\alpha \in \mathbf{R}^{n \times k}$, then $x \curvearrowright x\alpha$ is a *linear function* from $\mathbf{R}^n$ into $\mathbf{R}^k$.

If $\alpha = \{\alpha_{ij}\} \in \mathbf{R}^{n\times n}$ is a *square matrix*, then $\det(\alpha)$ denotes *the determinant* of α, and $\mathrm{tr}(\alpha)$ denotes *the trace* of α, that is,

$$\text{(P.8.1)} \qquad \det(\alpha) := \sum_{\sigma\in\mathcal{P}_n} \left\{\mathrm{sign}(\sigma) \prod_{i=1}^{n} \alpha_{i\sigma(i)}\right\} \quad , \quad \mathrm{tr}(\alpha) := \sum_{i=1}^{n} \alpha_{ii}$$

where $\mathcal{P}_n$ is the set of all permutations of $\{1,\ldots,n\}$ and $\mathrm{sign}(\sigma) = \pm 1$ if σ is an even/odd permutation. Let $\alpha, \beta \in \mathbf{R}^{n\times n}$, $\gamma \in \mathbf{R}^{n\times p}$ and $\eta \in \mathbf{R}^{p\times n}$ be given, then it is well-known that we have

$$\text{(P.8.2)} \qquad \det(\alpha^t) = \det(\alpha) \quad , \quad \det(\alpha\beta) = \det(\alpha)\det(\beta)$$

$$\text{(P.8.3)} \qquad \mathrm{tr}(\alpha^t) = \mathrm{tr}(\alpha) \quad , \quad \mathrm{tr}(\gamma\eta) = \mathrm{tr}(\eta\gamma)$$

If $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ are vectors in $\mathbf{R}^n$, then $x \bullet y$ denotes the usual *inner product* of x and y, that is,

$$\text{(P.8.4)} \qquad x \bullet y := \sum_{j=1}^{n} x_j\, y_j = xy^t = yx^t = \mathrm{tr}(x^t y) = \mathrm{tr}(y^t x)$$

If $x = (x_1,\ldots,x_n) \in \mathbf{R}^n$ and $\alpha = \{\alpha_{ij}\} \in \mathbf{R}^{n\times k}$, then $||x||$ and $||\alpha||$ denote *the euclidean norms* of x and α, that is,

$$\text{(P.8.5)} \qquad ||x|| := \sqrt{\sum_{j=1}^{n} x_j^2} = \sqrt{(x \bullet x)}$$

$$\text{(P.8.6)} \qquad ||\alpha|| := \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{k} \alpha_{ij}^2} = \sqrt{\mathrm{tr}(\alpha\alpha^t)} = \sqrt{\mathrm{tr}(\alpha^t\alpha)}$$

If $x, y \in \mathbf{R}^n$, $z \in \mathbf{R}^k$, $\alpha = \{\alpha_{ij}\} \in \mathbf{R}^{n\times k}$, $\beta \in \mathbf{R}^{k\times q}$ and $\gamma \in \mathbf{R}^{k\times p}$, then it is well-known that we have

$$\text{(P.8.7)} \qquad |(x \bullet y)| \le ||x|| \cdot ||y|| \quad , \quad ||\alpha\beta|| \le ||\alpha|| \cdot ||\beta||$$

$$\text{(P.8.8)} \qquad \max_{i,j} |\alpha_{ij}| \le ||\alpha|| = ||\alpha^t|| \le \sqrt{nk}\, \max_{i,j} |\alpha_{ij}|$$

$$\text{(P.8.9)} \qquad (\alpha\beta)^t = \beta^t\alpha^t \quad , \quad (x\alpha) \bullet z = x \bullet (z\alpha^t) \quad , \quad |\mathrm{tr}(\alpha\gamma)| \le ||\alpha|| \cdot ||\gamma||$$

Let L be a linear subspace of $\mathbf{R}^n$. Then $p_L \in \mathbf{R}^{n\times n}$ denotes *the orthogonal projection* of $\mathbf{R}^n$ onto L, and $L^\perp$ denotes *the orthogonal complement* of L; that is,

$$\text{(P.8.10)} \qquad L^\perp := \left\{x \in \mathbf{R}^n \;\middle|\; x \bullet y = 0 \quad \forall y \in L\right\}$$

Let $\alpha \in \mathbf{R}^{n\times k}$. Then we define *the range* of α , denoted $\mathcal{R}_\alpha$, *the null space* of α , denoted $\mathcal{N}_\alpha$, and *the rank* of α , denoted $\operatorname{rank}(\alpha)$, as follows:

(P.8.11) $\mathcal{R}_\alpha := \{x\alpha \mid x \in \mathbf{R}^n\} \subseteq \mathbf{R}^k \quad , \quad \mathcal{N}_\alpha := \{x \in \mathbf{R}^n \mid x\alpha = 0\} \subseteq \mathbf{R}^n$

(P.8.12) $\operatorname{rank}(\alpha) := \dim \mathcal{R}_\alpha$

We let p_α denote *the orthogonal projection* of $\mathbf{R}^k$ onto the range $\mathcal{R}_\alpha$. It is well-known that we have

(P.8.13) $\dim \mathcal{R}_\alpha + \dim \mathcal{N}_\alpha = n \quad , \quad \operatorname{rank}(\alpha) = \operatorname{rank}(\alpha^t)$

(P.8.14) $\mathcal{N}_{\alpha^t} = \mathcal{R}_\alpha^{\perp} \quad , \quad \mathcal{R}_{\alpha^t} = \mathcal{N}_\alpha^{\perp}$

(P.8.15) $\mathbf{R}^n = L + L^\perp \quad , \quad n = \dim L + \dim L^\perp \quad , \quad L = L^{\perp\perp}$

(P.8.16) $\operatorname{rank}(\alpha_1\alpha_2\cdots\alpha_p) \le \min\limits_{1\le i\le p} \{\operatorname{rank}(\alpha_i)\}$

whenever $\alpha \in \mathbf{R}^{n\times k}$, L is a linear subspace of $\mathbf{R}^n$ and $\alpha_1, \ldots, \alpha_p$ are matrices such that the product $\alpha_1\alpha_2\cdots\alpha_p$ is well-defined.

List of Symbols

Words beginning with a symbol or a Greek letter are listed at the beginning under its Latin transliteration. For instance, "κ_X , cumulant transform" is listed under K. Moreover, due to the limitations of the automatic indexing, caligraphic letters such as $\mathcal{M}$ or $\mathcal{L}$ will be written as bold italic letters $\boldsymbol{M}$ or $\boldsymbol{L}$, for instance $\mathcal{M}(\mu)$ or $\mathcal{B}(\mathbf{R})$ will be listed as $\boldsymbol{M}(\mu)$ or $\boldsymbol{B}(\mathbf{R})$.

, counting measure 16
$\downarrow$, decreasing convergence, functions . . 563
$\downarrow$, decreasing convergence, sets 548
$\uparrow$, increasing convergence, functions . . 563
$\uparrow$, increasing convergence, sets 548
$\rightarrow$, pointwise convergence 563
$\perp$, singular 176
<< , absolutely continuous 175
$\equiv$ (mod G) , atomic equivalence 441
$\perp\!\!\!\perp$, independent 104
$\|x-y\|$, euclidean distance 556
$\|\alpha\|$, euclidean norm, matrix 566
$\|x\|$, euclidean norm, vector 566
$[[\alpha,\beta]]$, integer interval 484
$\|f\|_F$, sup-norm 563
(C)-stable 2
$(\cap c)$-stable 2
(Σc)-stable 2
$(\cup c)$-stable 2
$(\downarrow)$-stable 2
$(/)$-stable 2
$(\cap f)$-stable 2
(Σf)-stable 2
$(\cup f)$-stable 1
$(\uparrow)$-stable 2, 55
(-)-stable 2
$|f|_\phi$, ϕ a function 184
$\|f\|_q$, q-norm $q\geq 0$ 185
$\|X\|_q$, q-norm $q\geq 0$ 253
$x\vee y$, maximum 552
$x\wedge y$, minimum 552
$1_A(\ldots)$, indicator function 554
2^Ω , power set 548

$\xrightarrow{\mathcal{L}}$, convergence in law 355
$\sim$, identically distributed 111

A

$\alpha(\boldsymbol{F})$, the algebra generated by $\boldsymbol{F}$ 8
$\alpha_0(\boldsymbol{F})$, ring generated by $\boldsymbol{F}$ 68
$\boldsymbol{A}(M)$, analytic sets 97
$A\Delta B$, symmetric difference 547
A^c , complement 547

B

$\boldsymbol{B}(\Omega)$, Borel σ-algebra 12, 46
$B(\Omega)$, bounded functions 563
$\boldsymbol{B}(\mathbf{R})$, the Borel σ-algebra on $\mathbf{R}$ 13
$\boldsymbol{B}(\mathbf{R}^k)$, the Borel σ-algebra on $\mathbf{R}^k$ 14
$b(s,r)$, open ball 555
$b[s,r]$, closed ball 555
$\boldsymbol{B}^d$, d-fold product σ-algebra 464
bd(A) , boundary 556
$\boldsymbol{B}_I(\boldsymbol{F})$, bi-Souslin $\boldsymbol{F}$-sets 96
$B_n(f,t)$, Bernstein polynomial 333

C

$\mathbf{C}$, the complex numbers 551
$C(\mu)$, μ-a.s. continuous functions . . . 356
$C(S)$, bounded continuous functions 355, 563
C_F , continuity points 372
cl(A) , closure 556
$con_\sigma(\boldsymbol{W})$, $(\uparrow)$-stable convex cone 55
cov(X,Y) , covariance 256
Cov(X,Y) , covariance matrix 257

D

δ_c , Dirac measure 17
$\boldsymbol{D}(\Gamma_\mu)$, domain 223
$d(A,B)$, minimal distance 41
$\boldsymbol{D}(L_X)$, domain 283
$\boldsymbol{D}(m_X)$, domain 253
$\boldsymbol{D}(q_X)$, domain 289
$\boldsymbol{D}(r_X)$, domain 288
$\boldsymbol{D}(\mu_X)$, domain 254

$D(\Gamma_X)$, domain 276
$d(x,F)$, minimal distance 560
$d(x,y)$, metric 555
$d\nu/d\mu$, density function 175
$d\nu=\phi\, d\mu$, indefinite integral 175
$\det(\alpha)$, determinant 566
$d_F(f,g)$, sup-metric 563
$\mathrm{diam}(A)$, diameter 41
$Dis(\mathbf{R})$, the set of distribution functions . 146
$d^k(x,y)$, product metric 558
$d^{oo}(x,y)$, product metric 559
$D_r{}^*f(x)$, right upper derivative 244
$D_rf(x)$, right derivative 244
$D_r{}^of(x)$, right lower derivative 244

E

$E(F|G)$, conditional probability 450
$E(M_g)$, domain 214
$E(X|G)$, conditional expectation 448
$E(X|T)$, conditional expectation 448
$E(X|T=t)$, conditional expectation 449
$E^*(M_g)$, extended domain 214
ess inf X 316
ess sup X 316
EX , mean matrix 257
EX , mean value 250
EX , mean vector 257
extended
domain
$D^*(r_X)$, real generating function . 288

F

$\phi_X(t)$, characteristic function 280
F_τ , information at time τ 487
f^- , negative part 156
$f*g$, discrete convolution 203
$f*g$, Lebesgue convolution 203
$f*F$, Lebesgue-Stieltjes convolution . . 202
$f(A)$, image set 550
$\hat{F}(p)$, smallest fractile 414
f^+ , positive part 156
$f^{-1}(A)$ co-image set 550
$F_{-\infty}$, initial information 485
F_{∞} , ultimate information 485
$F_X(x)$, distribution function . 109, 110, 279
$\check{F}(p)$, largest fractile 414

G

$\Gamma(\alpha,\beta)$, gamma distribution 295
$\Gamma(\lambda)$, gamma function 216
$\Gamma_\mu(t)$, Γ-transform 223
$\Gamma_X(t)$, Γ-transform 276
$\gamma(X)$, Gaussian subvariance 304
$\gcd(p,q)$, greatest common divisor . . . 139
$Ge(p)$, geometric distribution 148

H

$H\exp(\mu,\alpha,\beta,\gamma)$, hyper-exponential distribution 299
$H\exp^+(\mu,\alpha,\beta,\gamma)$, hyper-exponential distribution 299

I

$\mathrm{int}(A)$, interior 556
$int(x)$, integer part 53

K

$\kappa_X(t)$, cumulant transform 283

L

λ , Lebesgue measure 37
λ_k , Lebesgue measure 38
λ_F , Lebesgue-Stieltjes measure 37, 38
$L(\mu)$, $L(\mu)$-integrable functions 157
$L(\mu)$, Lebesgue μ-measurable sets 30
$L^\perp$, orthogonal complement 566
$L^1(\mu)$, integrable functions 157
$L^1(\mu,\mathbf{C})$, complex integrable functions . 210
$Lev(\mu,\sigma,q,\nu)$, Lévy-Khinchine distribution 384
lim inf A_n , limes inferior of numbers . 554
lim inf A_n , limes inferior of sets 548
lim sup a_n , limes superior of numbers . 554
lim sup A_n , limes superior of sets . . . 548
$\lim_{n\to oo} x_n$, limit 556
$\mathrm{Lip}(S)$, Lipschitz' functions 357
$\mathrm{Lip}_1{}^+(S)$, Lipschitz' functions 358
$\log N(\mu,\sigma^2)$, log-normal distribution . . 297
$L^q(\mu)$, q-integrable functions 186
$L^q(\mu,\mathbf{C})$, complex q-integrable functions 210
$L^q(P)$, q-integrable random variables . . 253
$L^q{}_k(\mu)$, q-integrable vector functions . . 225
$L^q{}_{n\times k}(\mu)$, q-integrable matrix functions 225

Lsc(μ) , μ-a.s. lower semicontinuous functions 356
$L_X(t)$, Laplace transform 283

M

$\mu_X{}^0(n)$, central moment 255
μ_f , image measure 51
μ_* , inner measure 23
$\mu_X(n)$, moment 254
μ^* , outer measure 23
$M(\mu)$, μ-measurable set, μ a set function 23
$M(\mu)$, μ-measurable sets, μ a measure . . 30
$M(\Omega,F,\mathbf{C})$, complex measurable functions 210
$M(S)$, bounded measurable functions . . 356
$M(S)$, measurable functions 356
$M_b(X_n)$, L^1-bounded functions 357
$M_g(p)$, Mellin transform 214
$M_u(\nu_n)$, uniformly integrable functions 361
$M_u(X_n)$, uniformly integrable functions 357
$Mu_k(p_1, \ldots ,p_k)$, multinomial distribution 146
$m_X(q)$, absolute moment 253
$m_X{}^0(q)$, central absolute moment 255

N

$\mathbf{N}$, the positive integers 551
$\nu_X(n)$, factorial moment 287
$N(\mu)$, the μ-null sets 20
$N(\mu,\sigma^2)$, normal distribution 290
$\mathbf{N}^{(\mathbf{N})}$, finite multiindices 93
$\mathbf{N}_0$, the non-negative integers 551
$N_k(\mu,\Sigma)$, normal distribution 292
$\mathbf{N}^{\mathbf{N}}$, infinite multiindices 93

O

$\Omega_1 \times \ldots \times \Omega_n$, product set 548
$O(g)$, big O-functions 356
$o(g)$, small o-functions 356

P

$\Pi_{j\in J}\, \Omega_j$, product set 548
$P(A|B)$, conditional probability 101
$P(F|T)$, conditional probability 450
$P(F|T=t)$, conditional probability 450
$Par(\alpha,q)$, Pareto distribution 325
P_n , permutations 465
$Po(\lambda)$, Poisson distribution 297
$Po[\lambda]$, compound Poisson distribution . 383
$P_X(B)$, distribution law 109
$p_X(x)$, probability mass function . 116, 279

Q

$q_X(t)$, generating function 289

R

$\mathbf{R}$, the real numbers 551
$\rho(X,Y)$, correlation coefficient 256
$\rho_o(f)$, lower tail functional 387
$\rho_{oo}(\mathrm{f})$, upper tail functional 387
$R(a,b)$, rectangular distribution 129
$R_\mu(f)$, f a function 218
$R(L_X)$, real domain 283
$\mathbf{R}_+$, the non-negative real numbers . . . 551
$\bar{\mathbf{R}}_+$, the non-negative extended real line 552
$\bar{\mathbf{R}}$, the extended real line 551
rank(α) , rank 567
$\mathbf{R}^n$, the euclidean space 565
$\mathbf{R}^{n\times k}$, set of matrices 565
$R^o(L_X)$, interior domain 283
$r_X(t)$, real generating function 288

S

$\sigma(f)$, σ-algebra generated by one function 48
$\sigma(F)$, the σ-algebra generated by F 8
$\sigma(W)$, σ-algebra generated by many functions 49
$\sigma^2(X|G)$, conditional variance 450
$\sigma^2(X|T)$, conditional variance 450
$\sigma_0(F)$, ring generated by F 68
$\Sigma_{j\in J}\, a_j$, sum 553
$\sigma(X)$, standard deviation 256
$S(A)$, Souslin operation 94
$S(F)$, F-Souslin sets 94
S^n , permutation invariant σ-algebra . . 465
$S_n f(t)$, symmetrization 466
$sp(\mu)$, spectrum 83
$sp_+(\mu)$, point spectrum 82
$sp_+(X)$, discrete support 116
$sp_c(\mu)$, continuous spectrum 84
$Stab(\mu,\sigma,p,\gamma)$, stable distribution 386
$S_X{}^n$, permutation invariant σ-algebra . . 466

T

T^d , d-fold product 464
tr(α) , trace 566
tr$(F|\Omega_0)$, trace σ-algebra 69
$Tri(\mu,\lambda)$, triangular distribution 413

U

$\Upsilon_\Gamma \Delta\tau(\omega)$, occurrence time after τ . . 487
$\Upsilon_\Gamma(\omega)$, occurrence time 487
Usc(μ) , μ-a.s. upper semicontinuous functions 356

V

$V(X)$, covariance matrix 257
var(X) , variance 256

W

$W_f(s)$, global oscillation 333

X

X_τ , composite random variable 490
x^- , negative part 552
$x^{(n)}$, factorial 287, 335
x^+ , positive part 552
$x_n \to x$, convergence 556

Z

$\mathbf{Z}$, the integers 551

Author Index

A

Abramowitz, M., 227
Adams, J., xl
Adrain, R., 294, 319
Alexandrov, A. D., 418
Andræ, C. C. G., 518
Aristotle, xxviii

B

Banach, S., 133
Bandarina, L., xxi
Bayes, T., 102
Benford, F., 324
Berg, C., 301, 327
Bernoulli, D., 320
Bernoulli, J., xxxiii, 263, 305
Bernstein, S. N., 333, 335
Berry, A. C., 399
Bienaymé, J., 331
Billingsly, P., 418
Blackwell, D., 445
Bochner, S., 89
Borel, É., xxxv, 6, 11, 13, 18, 46, 152, 275
Boškovich, R. J., 318
Brahe, T., 317
Buffon, G.-L. L., 140

C

Cantelli, F. P., 6, 275
Cantor, G., 78
Carathéodory, C., 23, 152
Cardano, A, xxii
Cardano, C., xxii
Cardano, G., xxii, 103, 117, 315
Cauchy, A. L., 557
Cavalieri, B., 236
Cesàro, E., 267
Chebyshev, L. P., xxxiv, 256, 379
Chow, Y. S., 131
Christensen, J. P. R., 327
Chung, K. L., 477

D

David, F. N., xl
Dellacherie, C., 535
Diophantus, xxv
Dirac, P. A. M., 17
Dirichlet, G. P. L., 130
Doob, J. L., 477, 535
Dudley, R. M., 131
Dunford, N., 189, 227

E

Edgar, G. A., 535
Egoroff, D. T., 232
Einstein, A., xxxv
Erlang, A. K., 165
Essén, C., 399
Euclid, E., 555
Euler, L., 138, 214, 316, 318

F

Falkner, N., 477
Fatou, P. J. L., 160
Feller, W., 131, 380
Fermat, P. de, xxiv, 149
Fourier, J. B. J., 224, 282
Fournival, R. de, xviii
Fubini, G., 183, 184

G

Gaenssler, P., 535
Gauss, K. F., 294, 317, 319
Gnedenko, B. V., 419
Gradsheyn, I. S., 227
Graunt, J., xxix

H

Haagerup, U., 305, 327
Hahn, H., 81
Hald, A., xl
Halley, E., xxxi
Halmos, P. R., 61
Hardy, G. H., 240, 241
Hilbert, D., xxxv, 241
Hoel, P. G., 131
Hoffmann-Jørgensen, J., 418, 477
Hölder, O. L., 166, 243
Huyghens, C., xxvi, 249, 264

J

Jensen, J. L. W. V., 165, 242, 254
Jordan, P., 81

K

Karlin, S., 535
Kelley, J. L., 61
Kemney, J. G., 477
Kestelman, H., 228
Khinchine, A. Y., 305, 307, 352, 380, 382, 424
Knapp, A. W., 477
Kolmogorov, A. N., xxxvi, 61, 419
Kronecker, H., 16, 267

L

L'Hospital, G.-F.-A. de, 304, 424
Lambert, J. H., 316, 318
Laplace, P.-S. de, xxxiv, 102, 224, 283, 286, 300, 379
Lebesgue, H. L., 30, 36, 37, 53, 99, 152, 161, 175, 198, 201, 228, 248
Legendre, A.-M., 319
Leibniz, G. W., 152, 424
Levi, B., 162
Lévy, P., 380, 382, 424, 494, 505
Lindeberg, J. W., 380, 393, 419
Lipschitz, R. O. S., 357
Littlewood, J., 240, 241
Lloyd, E., xxx
Loeve, M., 132
Lucas, F. É.-A., 118
Lukacs, E., 327
Lusin, N. N., 53, 233
Lyapounov, A. M., 82, 164, 380, 393

M

Maistrov, E., xl
Marczewski, E., 91
Markov, A. A., xxxv, 164, 254, 330, 331, 380
Mellin, H., 214, 302
Meré, C. de, xxiv
Méré, C. de, 149
Meyer, P.-A., 535
Minkowski, H., 167, 240, 243
Moivre, A. de, xxxiii, 293, 354, 379

N

Neveu, J., 535
Newton, I., 152
Nikodym, O. M., 175, 178

O

Ore, O., xl

P

Pacioli, L., xix, 148
Parseval, M.-A., 224
Pascal, B., xx, xxiv, xxv, 148, 149
Pascal, G. R., 134
Peano, G., 245
Petrov, V. V., 419
Pettis, B. J., 189
Peverone, G. F., xx, 148
Piazza, G., 317
Poisson, S. D., 297
Polya, G., 240, 241
Port, S. C., 131

R

Radon, J., 39, 175, 178
Raikov, D. A., 380, 393
Rasmussen, N. C., 124
Ressel, P., 327
Révész, P., 327
Revuz, D., 477
Rhyzhik, I. M., 227
Riemann, B., 138, 152, 198, 248
Riesz, F., 234, 239
Riesz, M., 241
Rogers, C. A., 61
Rudin, W., 61

S

Saks, S., 228
Schur, I., 241
Schwartz, J. T., 227
Segun, I. A., 227
Sierpińsky, W., 8, 54
Snell, J. L., 477
Souslin, M. M., 53, 94, 99
Srinvasan, T. P., 61
Stieltjes, T. J., 36, 37, 201
Stirling, J., 354
Stone, C. J., 131
Stone, H. M., 55
Strobel, J., 535
Stute, W., 535
Sucheston, L., 535
Suttel, B., 134

T

Talagrand, M., 88
Tartaglia, N., xx, 148
Taylor, M. H., 535
Teicher, H., 131
Thorin, G. O., 239
Tonelli, L., 183, 184
Topsøe, F., 61

W

Wald, A., 496
Weierstrass, K. T. W., 333
Westergaard, H., xl
Wibold of Cambray, xviii
Ẃidder, D. V., 327
Wiener, N., xxxvi
Wiles, A., xxv

Subject Index

Index words beginning with a Greek letter are listed at the beginning under its Latin transliteration, for instance "σ-algebra" is listed under S, and "μ-density" is listed under M. Subindex words beginning with a Greek letter are listed at the beginning under the main index word.

0-1 law
  pavings 114
  permutation invariant σ-algebra . . . 533
  random functions 114
0-1 measure 17

A

a posteriori probability 102
a priori probability . . . . . xxxviii, 102
a.a. (almost all) 20
a.s. (almost sure) 20
absolute moment 253
  convergence 375
  transform 253, 286
absolutely
  μ -continuous 175
  μ -continuous part 180
  μ -continuous, random function . . . 250
  continuous 201, 202
    random vector 251
additive set function 22
algebra . . . . . xxxvii, 2, 7
  construction 17
  generated by sets 8
almost all (a.a.) 20
almost sure (a.s.) 20
analytic
  complex function 285
  set 445, 559
anomal distribution 323
antisymmetric 564
approximately continuous 228
approximating paving 23
arithmetic mean 316
astragalus xvi
atom 81, 441
  calculus 442
atom-free 81
atomic 81
  decomposition 81, 82
average 316, 319
axiom of foundation 550

B

ballot theorem 516, 517
Banach's problem 133
Bayes' rule 102
Bayesian school of statisticians . xxxviii, 102
Beppo Levi's Theorem 162
Bernoulli variable 305
Bernstein
  inequality 335
  polynomial 333
Berry-Essén estimate 399
Bessel function of the first kind, modified 339
beta distribution 130
Bienaymé's inequality 331
big O 357
bijective 550
binary expansion 90
binomial coefficient 63
binomial distribution 128
  normal approximation 411
  Poisson approximation 411
biquadratic mean 315
bi-Souslin set 96
Blackwell σ-algebra 445
Bochner measurable 89
Borel
  σ -algebra 11, 12, 18
  σ -algebra on $\mathbf{R}$ and $\mathbf{R}^k$ 13
  σ -algebra, generation 12
  content 12
  function 46
  measurable function 46
  measurable set 12

measure 12
set 12
Borel-Cantelli lemma
first 6, 113
second 113, 275
boundary 556
Buffon's needle problem 140

C

Cantor set 78
Carathéodory's theorem 23
Cardano's formula 315
Cauchy
distribution 386
sequence 556
Cauchy-Schwarz inequality 166
Cavalieri's principle 236
center 555
centered average 264
centering constants 264
central
absolute moment 255
transform 255
limit theorem 379, 381, 398
general 389
normal 379, 382, 392, 394, 395
Poisson 396
stable 397
moment 255
transform 255
Cesàro's lemma 267
characteristic function 280
convergence 373
infinitely divisible 402
inversion formulas 399
stable 403
Chebyshev's inequality 256
for conditional expectation 455
Chevalier de Méré 's problem 149
closed 556
ball 555
closure 556
co-image 550
column vector 565
compact
metric space 557
paving 31
compactification 558
complement 547
orthogonal 566
complementary operation 8
complementation 1
complete
measure 30
metric space 556
completion
σ -algebra 30
measure 30
metric space 558
complex
conjugated 211, 551
domain 283
number 551
composite random variable 490
compound Poisson distribution 383
condensation point 285
conditional
density 458
distribution 450
function 477
expectation 447
calculus 451
Chebyshev's inequality 455
convergence 456
given a σ-algebra 447
given a function 448
given $T=t$ 449
inequalities 455
Jensen's inequality 455
Markov's inequality 455
probability 101
given a σ-algebra 450
given a function 450
given $T=t$ 450
variance
given a σ-algebra 450
given a function 450

conditionally independent 460, 463
given a σ-algebra 450
given a function 451
conjugated, of a complex number 211
consistency theorem xxxvi
content 4
Borel 12
finite-co-finite 16
continuity
of integrals 208
theorem 370
continuous 559
along a set 284
approximately 228
spectrum 84
convergence
μ -a.s. uniformly 184
absolute moments 375
characteristic functions 373
conditional expectations 456
distribution functions 372
empirical fractiles 416
fractiles 415
generating functions 378
in ϕ-mean, ϕ a function 184
in α-mean, $\alpha \geq 0$ 185
in μ-measure 184
in law 354, 355
calculus 364
criterion 360
in *pr* 250, 355
in probability 250, 355
metric space 556
moments 376
partial sums 262, 310
probability mass functions 374
real series 553
relations between 194
U-statistics 532
uniformly 563
converges
μ -a.s. uniformly 184
in ϕ-mean, ϕ a function 184
in α-mean, $\alpha \geq 0$ 185
in μ-measure 184
in law 355
in *pr* 250, 355
in probability 250, 355
pointwise 563
real series 553
uniformly 563
converse two-three-series theorem 351
convex
cone 55
($\uparrow$)-stable 55
function 241
set of distribution functions 147
strictly 243
convolution
discrete 203, 261
Lebesgue 203, 261
Lebesgue-Stieltjes 202, 261
coordinate function 50
correlation coefficient 256
countable 553
disjoint union 1
intersection 1
union 1
countably
generated, σ-algebra 66
separating, σ-algebra 91
counting measure 16
covariance 256
matrix, of a random vector 257
matrix, of two random vectors 257
of two random variables 256

D

de Morgan's laws 549
De Vetula xviii
decimal expansion 90

decomposition
atomic 81, 82
Hahn 81
Jordan 81
Jordan-Hahn 80
point spectral 82, 83
set 80
theorem 80
vector measure 80
decreasing
intersection 1
sequence of functions 563
sets 548
degenerated
random function 115, 116
random variable 256
dense 556
density 201, 202
conditional 458
random vector 251
stable distribution 409
derivative
n^{th} 423
Radon-Nikodym 176, 178
right 244
right lower 244
right upper 244
determinant 566
diagonal 92
diameter 41
difference 1, 547
proper 1
symmetric 547
differentiability of integrals 208
differentiable
n-times 423
Radon-Nikodym 176
right 244
twice 229
diffuse
measure 82
random function 117
Dini's theorem 564
Dirac measure 17, 386
Dirichlet distribution 130
discrete
σ -algebra 89
convolution 203, 261
function 52, 89
measure 82
metric 16
random function 116
support 116
disjoint 548
countable union 1
finite union 1
disjungation 3
distance
between sets 41
measure 318
distribution
anomal 323
beta 130
binomial 128
Cauchy 386
compound Poisson 383
conditional 450
Dirac 386
Dirichlet 130
exponential 126
function 109–111, 279
conditional 477
convergence 372
empirical 127
fractile 414
infinitely divisible 402
largest fractile 415
marginal 111
simultaneous 111
smallest fractile 415
stable 403
gamma 128, 295, 384
Gauss 294
geometric 125, 148
hyper-exponential 299
one-sided 299
two-sided 299
infinitely divisible 402
Laplace 300
law 109

Lévy-Khinchine 382, 384
log-normal 297
negative binomial 384
normal 290
normal, k-dimensional 291
order statistics 469
Pareto 325
Poisson 297, 384
rectangular 129
spherical 130
stable 385, 386, 403
triangular 413
uniform 129
distributive laws, sets 549
domain
Γ^* -transform 223, 277
Γ -transform 223, 276
absolute moment transforms 253
complex, Laplace transform 283
extended, Laplace transform 283
interior Laplace transform 283
Mellin transform 214
moment transforms 254
real generating function 288
real, Laplace transform 283
dominated
convergence theorem 161
L^1, set of functions 188
Dunford-Pettis' theorem 189

E

Egoroff's theorem 232
empirical
distribution function 127
fractile 417
convergence 416
empty set 548
enumerated 553
epidemic 523, 524
equivalence
class 565
relation 564
estimator 467
euclidean
metric 556
norm 566
Euler
formula 138, 139
integral 214
eventual tightness theorem 367
eventually tight 367
criterion for on $\mathbf{R}^k$ 371
exchangeable 466, 495
events 138, 482
random variables 332
expectation . . . . . xxvi, 250
computation of 250
conditional 447
given a σ-algebra 447
given a function 448
given $T=t$ 449
independent random functions 258
exponential
distribution 126
inequality 473
extended
domain
Laplace transform 283
Mellin transform 214
Minkowski inequality 240
extension
content 72
maximal 72
measure 72
set function 22
theorem 29

F

factorial 287
moment 287
transform 287
n^{th} 335
fake infection 123
Fatou's lemma 160
filter 484
finite
disjoint union 1
intersection 1
set function 22
union 1
finite-co-finite content 16

finitely
μ -concentrated 187
founded 74, 171
first
Borel-Cantelli lemma 6, 113
occurrence time 488
after τ 488
transformation theorem 171
formal power series 135
Fourier
kernel 224
transform 224, 282, 373
transform, of a function 282
fractile 414
convergence 415
empirical 417
largest 415
smallest 415
Fubini's theorem 184
function
μ -integrable 157
μ -integral, complex-valued function 210
μ -integral, matrix function 226
μ -integral, non-negative function 152, 153
μ -integral, real-valued function 156, 157
μ -integral, vector function 226
μ -measurable 45
bijective 550
Borel 46
Borel measurable 46
convex 241
coordinate 50
discrete 52, 89
distribution 109–111
gamma 216
graph 93
incomplete hypergeometric 407
indicator 554
injective 550
Lipschitz 357
Marczewski 91
measurable 45
approximation 52
negative part 156
partial 54
positive part 156
probability mass 116, 279
random 108
right continuous 36, 37
set 21, 22
simple 52, 89, 116
surjective 550
symmetric 465
symmetrization 466
tame 465
functional Sierpińsky lemma 54
future information 462

G

Γ-determined 225
Γ-determined, random function 278
Γ^*-transform 223
Γ-transform 223
gamma
distribution 128, 295, 384
function 216
Gauss distribution 294
Gaussian
martingale 529
subvariance 304
general
central limit theorem 389
multiplication law 102
generating function 288, 289
convergence 378
real 288
geometric
distribution 125, 148
mean 316
sample 320
method 319
sample mean 320
global oscillation 333
graph, of a function 93
greatest common divisor 139

H

Hahn decomposition 81
Hardy's inequality 240
Hardy-Littlewood-Polya's inequality 240, 241
Hilbert-Schur-Hardy-Riesz' inequality . 241
Hölder's inequality 166, 243
hounds and jackals xvi
hull 482
hyper-exponential distribution 299
hypergeometric function, incomplete . . 407

I

identically distributed 111
image 550
  catastrophe 54
  measure 51
imaginary part 210, 551
imbedded 558
in ϕ-mean, ϕ a function 184
in α-mean, $\alpha \geq 0$ 185
in μ-measure 184
inclusion-exclusion principle 65, 137
incomplete hypergeometric function . . . 407
increasing
  sequence of functions 563
  set function 22
  sets 548
  union 1
increment 38
  positive 38
indefinite integral 173, 174
independent
  conditional 460, 463
    given a σ-algebra 450
    given a function 451
  events 102, 103
  pairwise 118
  pavings 103
    criteria 104, 106, 107
  random functions 112
    criteria 112
  random variables
    criterion 281
  triangular scheme 380
indicator function 554
induced metric 558
infection rate 525
infinitely divisible 402
  characteristic function 402
  distribution function 402
  random variable 402
information
  σ -algebra 440, 441
  at a stopping time 487
  at time a fixed time 485
  future 462
  initial 485
  past 462
  present 462
  ultimate 485
initial information 485
injective 550
inner
  μ -measure 23
  product 566
  restriction of a measure 71
  tight 27
    criteria 28
insurance xxx
  paradox 322
integer
  interval 484
  part 53
integrable 157, 210
  uniformly 187
integral
  calculus 158
  complex function 210
  continuity 168, 208
  differentiability 208
  discrete function 163
  exists 157
  indefinite 173, 174
  Lebesgue 198
    integration by parts 204
    substitution 205
  Lebesgue-Stieltjes 201, 202
    integration by parts 203
    substitution 204
  matrix function 225

non-negative function 152, 153
over a set 157
real-valued function 156, 157
Riemann 198
simple function 163
vector function 225
integration by parts 203, 204
interior 556
domain, Laplace transform 283
intersection 547
countable 1
decreasing 1
finite 1
inverse image 550
inversion formula 399

J

Jensen's inequality 165, 242, 254
for conditional expectation 455
Jordan decomposition 81
Jordan-Hahn decomposition 80

K

kernel 223
Fourier 224
Laplace 224
measurable 223
multiplicative 223, 277
transposed 223
Khinchine's inequality 305, 307
Kolmogorov's inequality 472
Kronecker
lemma 267
metric 16

L

l'Hospital's rule 424
L^1-dominated 188
Laplace
distribution 300
kernel 224
transform 224, 283, 286
of a function 286
largest fractile 415
lattice
difference 85
infimum 84, 85
supremum 84, 85
law 111
convergence 354, 355
of large numbers
263, 264, 269, 272, 274, 308
exchangeable random variables . 533
quadratic 533
strong 265
weak 264, 265
of small numbers 412
of total probability 102
Lebesgue
μ -measurable 30, 73
convolution 203, 261
dominated convergence theorem . . . 161
integral 198
integration by parts 204
substitution 205
measure 37, 38
Lebesgue-Stieltjes
convolution 202, 261
integral 201, 202
integration by parts 203
substitution 204
measure 36–38
Leibniz' multiplication rule 424
length 37
Lévy-Khinchine
distribution 382, 384
pair 384
representation 424
Lévy
inequality 473
martingale 494
measure 384
theorem 505
limes inferior
calculus 555
numbers 554
sets 548

limes superior
calculus 555
numbers 554
sets 548
limit 556
Lindeberg's condition 393–395
linear
function 565
ordering 564
linearly ordered 565
Lipschitz' condition 357
Lloyd's xxxi
log-normal distribution 297
lower
semicontinuous 560
tail functional 387
L^q-space 186
Lusin's theorem 233
Lyapounov's
condition 393, 394, 396
theorem 82

M

μ-atom 81
μ-continuous 169
μ-density 175, 259
μ-density, of a random function 250
μ-integrable 157
μ-integral, of a complex function 210
μ-integral, of a matrix function 226
μ-integral, of a positive function 153
μ-integral, of a real function 157
μ-integral, of a vector function 226
μ-measurable 23
μ-measurable function 45
μ-measurable set 23, 30
μ-saturated 169
μ-singular set 176
main
logarithm 422
root 216
of a function 422
Marczewski function 91
marginal distribution function 111
Markov
chain, weak 462, 464
property, weak 461, 462
Markov's inequality 164, 254
for conditional expectation 455
one-sided 330, 331
martingale 491, 492
convergence theorem 502, 503
convex function 515
difference 511
examples 494
Gaussian 529
inequality 472, 498
Lévy 494
rescaling 518
reversed 494
scaled 519
scaled sub 518
scaled super 519
stopped 497
Wald 496
maximal
element 565
extension of a content 72
inequality 119, 120, 308
conditional 470
linearly ordered set 565
maximum 565
mean 250
matrix 257
sample 319
vector 257
measurable
μ -measurable 23, 45
Borel
function 46
set 12
function 45
approximation 52
kernel 223
Lebesgue μ-measurable set 30
space 45

measure 6
σ -finite 30, 171
τ -smooth 39
0-1 17
Borel 12
complete 30
completion 30
counting 16
Dirac 17
extension 29
finitely founded 74, 171
image 51
inner 23
inner restriction 71
Lebesgue 37, 38
Lebesgue-Stieltjes 36–38
Lévy 384
metric outer measure 41
generated by 41
operations 21
outer 23
outer restriction 71
probability 100
product 181, 182
projective limit 58, 59
Radon 39
signed 79
space 51
sum-finite 171
support 83
transform 222
vector 79
regular 80
singular 80
median 315, 331
method 318
sample 319
Mellin Transform 214, 302
method of least squares 319
metric 555
discrete 16
euclidean 556
induced 558
Kronecker 16
outer measure 41
generated by 41
product 50, 558
pseudo 555
space 555
sup 563
midrange 316
sample 318
minimal element 565
minimax method 318
minimum 565
Minkowski's inequality 167, 243
extended 240
modified Bessel function of the first kind 339
modular truncation 350
modulus 551
moment 253, 254
absolute 253, 286
absolute central 255
central 255
convergence 376
factorial 287
n-th moment 254
transform 254, 286
monotone
convergence theorem 159
sequence of function 563
moral value 321
multiindex 93
multinomial distribution 146
multiplicative kernel 223, 277

N

negative
binomial distribution 384
part 156
non-degenerated 146

normal
central limit theorem . . 382, 392, 394, 395
distribution 290
density 290, 293
k-dimensional 291
n^{th} factorial 335
null
set 20
space 567

O

observable random function 441
occurrence time 487
first 488
first after τ 488
k'th 488
one-sided
hyper-exponential distribution 299
Markov's inequality 330, 331
open 556
ball 555
opt an asset 521
option contract 521
optional 506
criteria 509
sampling theorem . . 490, 497, 508, 509
stopping time 506
T-time 506
times 506
order statistics 127
distribution 469
ordering 564
orthogonal
complement 566
projection 566
oscillation, global 333
Ottaviani's inequality 472
outer
μ -measure 23
measure
metric 41
metric, generated by 41
restriction of a measure 71
tight 28
criteria 28

P

p-adic expansion 90
pairwise independent 118
Pareto distribution 325
Parseval's equation 224
partial
function 54
ordering 564
past information 462
paving 1
approximating 23
compact 31
trace 69
Peano curve 245
permutation 465
invariant σ-algebra 464–466
Petersburg paradox 320
physical value 321
pigeonhole principle 475
plague xxvii
point
spectral decomposition 82, 83
spectrum 82
Poisson
central limit theorem 396
distribution 297, 384
compound 383
normal approximation 412
polish space 88, 557
polytope 284
portmanteau lemma 358
position parameters 519
positive
μ -atom 81
increment 38
part 156
type 146
power set 1, 548
pr 250
convergence 250
summable 262
precompact 557
preordering 564
present information 462

probability
- a posteriori 102
- a priori . . . . . xxxviii, 102
- conditional 101
  - given a σ-algebra 450
  - given a function 450
  - given $T=t$ 450
- content xxxvii
- frequential xxxviii
- law of total probability 102
- mass function 116, 279
  - convergence 374
- measure . . . . . xxxvi, 100
- space . . . . . xxxvi, 100
  - finitely additive xxxvii
- subjective xxxix

problem of points 148

product
- σ -algebra 49, 464
- inner 566
- measure 181, 182
- metric 50, 558
- set 49, 548

projection 49, 464
- orthogonal 566

projective limit measure 58, 59

proper
- difference 1
- stopping time 486
- T-time 485

pseudo-metric 555
- space 555

Q

quotient space 565

R

radius 555

Radon measure 39

Radon-Nikodym
- derivative 176, 178
- differentiable 176

Radon-Nikodym-Lebesgue's Theorem . 175

Raikov's condition 393

railroad track problem 562

random
- function 108
  - degenerated 115
  - exchangeable 466
  - independent 112
  - observable 441
- variable . . . . . xxxvi, 108
  - biquadratic mean 315
  - composite 490
  - exchangeable 332, 495
  - geometric mean 316
  - independent 112
  - infinitely divisible 402
  - median 315
  - midrange 316
  - range 316
  - stable 403
  - subgaussian 303
  - symmetric 306
- vector 108
  - independent 112
  - sign-symmetric 306

range 567
- of a random variable 316

rank 567

rapidly increasing 266

real
- domain 283
- generating function 288
- part 210, 551

rectangular distribution 129

reflexive 564

regular, vector measure 80

relation 564

reversed
- martingale 494
  - convergence theorem 503
- submartingale 494
  - convergence theorem 503
- supermartingale 494
  - convergence theorem 503

Riemann
- integral 198
- zeta function 138

Riemann-Lebesgue's lemma 248

Riesz' convexity theorem 234, 239
Riesz-Thorin's convexity theorem 239
right
  continuous 36, 37
  derivative 244
  differentiable 244
  lower derivative 244
  upper derivative 244
ring 68
rolling of the bones xvi
row vector 565

S

σ-additive set function 22
σ-algebra . . . . . xxxvi, 2, 7
σ-algebra, atoms 441
σ-algebra, Blackwell 445
σ-algebra, Borel 11, 12
σ-algebra, completion 30
σ-algebra, construction 19
σ-algebra, countably separating 91
σ-algebra, discrete 89
σ-algebra, generated by functions 48
σ-algebra, generated by sets 8
σ -algebra, information 440, 441
σ-algebra, permutation invariant . . 464–466
σ-algebra, product 49, 464
σ-algebra, separable 66, 445
σ-algebra, separates points 89, 442
σ-algebra, tail 469
σ-algebra, trivial 89
σ-convex, set of distribution functions . 147
σ-finite measure 30, 171
σ-ring 68
σ-smooth 76
σ-subadditive set function 22
(S)-stable 94
sample
  covariance, general 468
  mean 319, 467
  median 319
  midrange 318
  variance
    with known mean 467
    with unknown mean 468
scale parameters 519
scaled
  martingale 519
  submartingale 518, 546
  supermartingale 519
second Borel-Cantelli lemma 113
semi-contraction 482
semi-finite measure 74
semialgebra 75
separable
  σ -algebra 66, 445
  metric space 556
separates points 89, 442
set
  function 21, 22
    σ -additive 22
    σ -subadditive 22
    additive 22
    extension 22
    finite 22
    increasing 22
    subadditive 22
    superadditive 22
  null 20
  operation 1
  power 1
  product 49
Sierpińsky's stability lemma 8
  functional 54
sign-symmetric random vector . . . 306, 353
signed measure 79
simple function 52, 89, 116
simultaneous distribution function 111
singleton 15, 550
singular
  measures 176
  part 180
  vector measure 80
small o 357
smallest fractile 415
Souslin
  operation 93, 94
  scheme 93
  set 94
  stable 94
space, measurable 45

spectrum
  continuous 84
  measure 83
  point 82
spherical distribution 130
stable
  (C) , paving 2
  (Σc) , paving 2
  (Σf) , paving 2
  ($\cap c$) , paving 2
  ($\cap f$) , paving 2
  ($\cup c$) , paving 2
  ($\cup f$) , paving 1
  ($\downarrow$) , paving 2
  ($\uparrow$) , paving 2
  ($\uparrow$), set of functions 55
  (-) , paving 2
  (/) , paving 2
  central limit theorem 397
  characteristic function 403
  density 409
  distribution 385, 386, 403
    characterization 404
  distribution function 403
  random variable 403
  (S) 94
  Souslin 94
standard
  deviation 256
  metric space 88
  normal distribution 148
  proof 170
  specification 491
statistics 467
Stirling's formula 64
Stone's condition 55
stopped
  martingale 497
  submartingale 497
  supermartingale 497
stopping time 484, 486
  calculus 488
  information 487
  optional 506
  proper 486
strictly convex 243
strong law of large numbers 265
  exchangeable random variables . . . 533
  quadratic 533
subadditive set function 22
subgaussian random variable 303
submartingale 491
  convergence theorem 502, 503
  convex function 515
  difference 511
  inequality 498
  rescaling 518
  reversed 494
  scaled 518
  stopped 497
substitution
  Lebesgue integrals 205
  Lebesgue-Stieltjes integrals 204
sum-finite 171
summable
  a.s. 262
  in pr 262
  in q-mean 262
  real series 553
sup-metric 563
superadditive set function 22
supermartingale 491
  concave function 515
  convergence theorem 502, 503
  difference 511
  inequality 498
  non-negative 515
  rescaling 518
  reversed 494
  scaled 519
  stopped 497
support, of a measure 83
surjective 550
symmetric
  difference 547
  function 465
  random variable 144, 306
  relation 564
symmetrization 466

T

τ-smooth 39
τ^*-smooth 83
T-time 485
optional 506
tail
σ -algebra 469
function 252
functional
lower 387
upper 387
tame function 465
three-series theorem 349
converse 351
tight
eventually 367
inner 27
criteria 28
outer 28
criteria 28
time set 484, 491, 492
Tonelli's theorem 184
Tonelli-Fubini's Theorem 183
tontine xxxi
topological sum 97
topologically equivalent 557
total variation 324
vector measure 79
totally bounded 557
trace
matrix 566
paving 69
transform
Γ^* -transform 277
Γ -transform 276
absolute central moment 255
absolute moment 253, 286
central moment 255
characteristic function 280
cumulant 283
distribution function 279
domain 223
Euler integral 214
factorial moment 287
Fourier 224
function 282
measure 282
generating function 288, 289
Laplace 224, 283
function 286
measure 286
measure 222
Mellin 214, 302
moment 254, 286
probability mass function 279
random function 276
real generating function 288
transformation theorem, first 171
transitive 564
transposed
kernel 223
domain 223
matrix 565
triangle inequality 555
triangular
distribution 413
scheme 380
independent 380
trivial σ-algebra 89
truncation 349
modular 350
technique 380
twice differentiable 229
two-series theorem 350
converse 351
two-sided hyper-exponential distribution
. 299
Tychonov's theorem 559

U

U-statistics 466, 467
convergence 532
uan condition 380
ultimate information 485
uncorrelated 258
uniform
asymptotic negligibility condition . . 380
distribution 129

uniformly
μ -a.s. convergence 184
bounded 188
distributed 129, 130
integrable 187
criteria 191
summable 237
union 547
countable 1
countable disjoint 1
finite 1
finite disjoint 1
increasing 1
unit
ball 130
simplex 130
unordered sum 553
upcrossing
inequality 500, 501
number 500
upper
semicontinuous 561
tail functional 387

V

variance 256
conditional
given a σ-algebra 450
given a function 450
vector measure 79
version 111
volume 38

W

waiting times 125
continuous model 126
discrete model 125
Wald martingale 496
water reservoir model 530
weak
law of large numbers 264, 265
Markov
chain 462, 464
property 461, 462
Weierstrass-Bernstein's theorem 333
well-ordering 565

Y

Yersina Pestis xxvii