

Galen R. Shorack

Probability for Statisticians

With 23 Figures



Springer

Preface

There is a thin self-contained textbook within this larger presentation.

To be sure that this is well understood, I will describe later how I have used this text in the classroom at the University of Washington in Seattle.

Let me first indicate what is different about this book. As implied by the title, there is a difference. Not all the difference is based on inclusion of statistical material. (To begin, Chapters 1–6, provide the mathematical foundation for the rest of the text. Then Chapters 7–8 hone some tools geared to probability theory, while Chapter 9 provides a brief introduction to elementary probability theory right before the main emphasis of the presentation begins.)

The classical weak law of large numbers (WLLN) and strong law of large numbers (SLLN) as presented in Sections 10.2–10.4 are particularly complete, and they also emphasize the important role played by the behavior of the maximal summand. Presentation of good inequalities is emphasized in the entire text, and this chapter is a good example. Also, there is an (optional) extension of the WLLN in Section 10.6 that focuses on the behavior of the sample variance, even in very general situations.

Both the classical central limit theorem (CLT) and its Lindeberg and Liapunov generalizations are presented in two different chapters. They are first presented in Chapter 11 via Stein's method (with a new twist), and they are again presented in Chapter 14 using the characteristic function (chf) methods introduced in Chapter 13. The CLT proofs given in Chapter 11 are highly efficient. Conditions for both the weak bootstrap and the strong bootstrap are developed in Chapter 11, as is a universal bootstrap CLT based on light trimming of the sample. The approach emphasizes a statistical perspective. Much of Section 11.1 and most of Sections 11.2–11.5 are quite unusual. I particularly like this chapter. Stein's method is also used in the treatment of U-statistics and Hoeffding's combinatorial CLT (which applies to sampling from a finite population) in the optional Chapter 17. Also, the chf proofs in Section 14.2 have a slightly unusual starting point, and the approach to gamma approximations in the CLT in Section 14.4 is new.

Both distribution functions (dfs $F(\cdot)$) and quantile functions (qfs $K(\cdot) \equiv F^{-1}(\cdot)$) are emphasized throughout (quantile functions are important to statisticians). In Chapter 7 much general information about both dfs and qfs and the Winsorized variance is developed. The text includes presentations showing how to exploit the inverse transformation $X \equiv K(\xi)$ with $\xi \cong \text{Uniform}(0, 1)$. In particular, Chapter 7 inequalities relating the qf and the Winsorized variance to some empirical process results of Chapter 12 are used in Chapter 16 to treat trimmed means and L-statistics, rank and permutation tests, sampling from finite populations, and bootstrapping. (Though I am very fond of Sections 7.6–7.11, their prominence is minimized in the subsequent parts of the text.)

At various points in the text choices can be voluntarily made that will offer the opportunity for a statistical example or foray. (Even if the instructor does not exercise a particular choice, a student can do so individually.) After the elementary introduction to probability theory in Section 9.1, many of the classical distributions

of statistics are introduced in Section 9.2, while useful linear algebra and the multivariate normal distribution are the subjects of Section 9.3 and Section 9.4. Following the CLT via Stein's method in Section 11.1, extensions in Section 11.2–11.3, and application of these CLTs to the bootstrap in Sections 11.4–11.5, there is a large collection of statistical examples in Section 11.6. During presentation of the CLT via chfs in Chapter 14, statistical examples appear in Sections 14.1, 14.2, and 14.4. Statistical applications based on the empirical df appear in Sections 12.10 and 12.12. The highly statistical optional Chapters 16 and 17 were discussed briefly above. Also, the conditional probability Sections 8.5 and 8.6 emphasize statistics. Maximum likelihood ideas are presented in Section A.2 of Appendix A. Many useful statistical distributions contain parameters as an argument of the gamma function. For this reason, the gamma and digamma functions are first developed in Section A.1. Section A.3 develops cumulants, Fisher information, and other useful facts for a number of these distributions. Maximum likelihood proofs are in Section A.4.

It is my hope that even those well versed in probability theory will find some new things of interest.

I have learned much through my association with David Mason, and I would like to acknowledge that here. Especially (in the context of this text), Theorem 12.4.3 is a beautiful improvement on Theorem 12.4.2, in that it still has the potential for necessary and sufficient results. I really admire the work of Mason and his colleagues. It was while working with David that some of my present interests developed. In particular, a useful companion to Theorem 12.4.3 is knowledge of quantile functions. Sections 7.6–7.11 present what I have compiled and produced on that topic while working on various applications, partially with David.

Jon Wellner has taught from several versions of this text. In particular, he typed an earlier version and thus gave me a major critical boost. That head start is what turned my thoughts to writing a text for publication. Sections 8.6, 19.2, and the Hoffman–Jorgensen inequalities came from him. He has also formulated a number of exercises, suggested various improvements, offered good suggestions and references regarding predictable processes, and pointed out some difficulties. My thanks to Jon for all of these contributions. (Obviously, whatever problems may remain lie with me.)

My thanks go to John Kimmel for his interest in this text, and for his help and guidance through the various steps and decisions. Thanks also to Lesley Poliner, David Kramer, and the rest at Springer-Verlag. It was a very pleasant experience.

This is intended as a textbook, not as a research manuscript. Accordingly, the main body is lightly referenced. There is a section at the end that contains some discussion of the literature.

Contents

Use of this Text	xiii
Definition of Symbols	xviii

Chapter 1. Measures

1. Basic Properties of Measures	1
2. Construction and Extension of Measures	12
3. Lebesgue–Stieltjes Measures	18

Chapter 2. Measurable Functions and Convergence

1. Mappings and σ -Fields	21
2. Measurable Functions	24
3. Convergence	29
4. Probability, RVs, and Convergence in Law	33
5. Discussion of Sub σ -Fields	35

Chapter 3. Integration

1. The Lebesgue Integral	37
2. Fundamental Properties of Integrals	40
3. Evaluating and Differentiating Integrals	44
4. Inequalities	46
5. Modes of Convergence	51

Chapter 4. Derivatives via Signed Measures

1. Decomposition of Signed Measures	61
2. The Radon–Nikodym Theorem	66
3. Lebesgue’s Theorem	70
4. The Fundamental Theorem of Calculus	74

Chapter 5. Measures and Processes on Products

1. Finite-Dimensional Product Spaces	79
2. Random Vectors on (Ω, \mathcal{A}, P)	84
3. Countably Infinite Product Probability Spaces	86
4. Random Elements and Processes on (Ω, \mathcal{A}, P)	90

Chapter 6. General Topology and Hilbert Space

1. General Topology	95
2. Metric Spaces	101
3. Hilbert Space	104

Chapter 7. Distribution and Quantile Functions

1. Character of Distribution Functions	107
2. Properties of Distribution Functions	110
3. The Quantile Transformation	111
4. Integration by Parts Applied to Moments	115
5. Important Statistical Quantities	119
6. Infinite Variance	123
7. Slowly Varying Partial Variance	127
8. Specific Tail Relationships	134
9. Regularly Varying Functions	137
10. Some Winsorized Variance Comparisons	140
11. Inequalities for Winsorized Quantile Functions	147

Chapter 8. Independence and Conditional Distributions

1. Independence	151
2. The Tail σ -Field	155
3. Uncorrelated Random Variables	157
4. Basic Properties of Conditional Expectation	158
5. Regular Conditional Probability	168
6. Conditional Expectations as Projections	174

Chapter 9. Special Distributions

1. Elementary Probability	179
2. Distribution Theory for Statistics	187
3. Linear Algebra Applications	191
4. The Multivariate Normal Distribution	199

Chapter 10. WLLN, SLLN, LIL, and Series

0. Introduction	203
1. Borel–Cantelli and Kronecker Lemmas	204
2. Truncation, WLLN, and Review of Inequalities	206
3. Maximal Inequalities and Symmetrization	210
4. The Classical Laws of Large Numbers, LLNs	215
5. Applications of the Laws of Large Numbers	223
6. General Moment Estimation	226
7. Law of the Iterated Logarithm	235
8. Strong Markov Property for Sums of IID RVs	239
9. Convergence of Series of Independent RVs	241
10. Martingales	246
11. Maximal Inequalities, Some with \nearrow Boundaries	247
12. A Uniform SLLN	252

Chapter 11. Convergence in Distribution

1. Stein's Method for CLTs 255
2. Winsorization and Truncation 264
3. Identically Distributed RVs 269
4. Bootstrapping 274
5. Bootstrapping with Slowly Increasing Trimming 276
6. Examples of Limiting Distributions 279
7. Classical Convergence in Distribution 288
8. Limit Determining Classes of Functions 292

Chapter 12. Brownian Motion and Empirical Processes

1. Special Spaces 295
2. Existence of Processes on (C, \mathcal{C}) and (D, \mathcal{D}) 298
3. Brownian Motion and Brownian Bridge 302
4. Stopping Times 305
5. Strong Markov Property 308
6. Embedding a RV in Brownian Motion 311
7. Barrier Crossing Probabilities 314
8. Embedding the Partial Sum Process 318
9. Other Properties of Brownian Motion 323
10. Various Empirical Processes 325
11. Inequalities for the Various Empirical Processes 333
12. Applications 338

Chapter 13. Characteristic Functions

1. Basic Results, with Derivation of Common Chfs 341
2. Uniqueness and Inversion 346
3. The Continuity Theorem 350
4. Elementary Complex and Fourier Analysis 352
5. Esseen's Lemma 358
6. Distributions on Grids 361
7. Conditions for ϕ to Be a Characteristic Function 363

Chapter 14. CLTs via Characteristic Functions

0. Introduction 365
1. Basic Limit Theorems 366
2. Variations on the Classical CLT 371
3. Local Limit Theorems 380
4. Gamma Approximation 383
5. Edgeworth Expansions 390
6. Approximating the Distribution of $h(\bar{X}_n)$ 396

Chapter 15. Infinitely Divisible and Stable Distributions

- | | |
|---|-----|
| 1. Infinitely Divisible Distributions | 399 |
| 2. Stable Distributions | 407 |
| 3. Characterizing Stable Laws | 410 |
| 4. The Domain of Attraction of a Stable Law | 412 |

Chapter 16. Asymptotics via Empirical Processes

- | | |
|---|-----|
| 0. Introduction | 415 |
| 1. Trimmed and Winsorized Means | 416 |
| 2. Linear Rank Statistics and Finite Sampling | 426 |
| 3. The Bootstrap | 432 |
| 4. L-Statistics | 437 |

Chapter 17. Asymptotics via Stein's Approach

- | | |
|----------------------------------|-----|
| 1. U-Statistics | 449 |
| 2. Hoeffding's Combinatorial CLT | 458 |

Chapter 18. Martingales

- | | |
|---|-----|
| 1. Basic Technicalities for Martingales | 467 |
| 2. Simple Optional Sampling Theorem | 472 |
| 3. The Submartingale Convergence Theorem | 473 |
| 4. Applications of the S-mg Convergence Theorem | 481 |
| 5. Decomposition of a Submartingale Sequence | 487 |
| 6. Optional Sampling | 492 |
| 7. Applications of Optional Sampling | 499 |
| 8. Introduction to Counting Process Martingales | 501 |
| 9. Doob–Meyer Submartingale Decomposition | 511 |
| 10. Predictable Processes and $\int_0^\cdot H dM$ Martingales | 516 |
| 11. The Basic Censored Data Martingale | 522 |
| 12. CLTs for Dependent RVs | 529 |

Chapter 19. Convergence in Law on Metric Spaces

- | | |
|---|-----|
| 1. Convergence in Distribution on Metric Spaces | 531 |
| 2. Metrics for Convergence in Distribution | 540 |

Appendix A. Distribution Summaries

- | | |
|---|-----|
| 1. The Gamma and Digamma Functions | 546 |
| 2. Maximum Likelihood Estimators and Moments | 551 |
| 3. Examples of Statistical Models | 555 |
| 4. Asymptotics of Maximum Likelihood Estimation | 563 |

References 568

Index 575