

Contents

Preface	xiii
Index of Statistical Applications and Notable Examples	xv
1 Introduction	1
1.1 Why More Rigor is Needed	1
2 Size Matters	9
2.1 Cardinality	9
2.2 Summary	16
3 The Elements of Probability Theory	17
3.1 Introduction	17
3.2 Sigma-Fields	18
3.2.1 General Case	18
3.2.2 Sigma-fields When $\Omega = R$	20
3.2.3 Sigma-fields When $\Omega = R^k$	22
3.3 The Event That A_n Occurs Infinitely Often	26
3.4 Measures/Probability Measures	27
3.5 Why Restriction of Sets is Needed	33
3.6 When We Cannot Sample Uniformly	35
3.7 The Meaninglessness of Post-Facto Probability Calculations	36
3.8 Summary	36
4 Random Variables and Vectors	39
4.1 Random Variables	39

4.1.1	Sigma-Fields Generated by Random Variables	39
4.1.2	Probability Measures Induced by Random Variables	43
4.2	Random Vectors	46
4.2.1	Sigma-Fields Generated by Random Vectors	46
4.2.2	Probability Measures Induced by Random Vectors	48
4.3	The Distribution Function of a Random Variable	48
4.4	The Distribution Function of a Random Vector	52
4.5	Introduction to Independence	56
4.5.1	Definitions and Results	56
4.5.2	Product Measure	63
4.6	Take $(\Omega, \mathcal{F}, P) = ((0, 1), \mathcal{B}_{(0,1)}, \mu_L)$, Please!	66
4.7	Summary	72
5	Integration and Expectation	75
5.1	Heuristics of Two Different Types of Integrals	75
5.2	Lebesgue-Stieltjes Integration	78
5.2.1	Nonnegative Integrands	78
5.2.2	General Integrands	80
5.3	Properties of Integration	81
5.4	Important Inequalities	87
5.5	Iterated Integrals and More on Independence	90
5.6	Densities	94
5.7	Keep It Simple	97
5.8	Summary	99
6	Modes of Convergence	101
6.1	Convergence of Random Variables	102
6.1.1	Almost Sure Convergence	102
6.1.2	Convergence in Probability	105
6.1.3	Convergence in L^p	109
6.1.4	Convergence in Distribution	110
6.2	Connections between Modes of Convergence	114

6.2.1	Convergence in L^p Implies Convergence in Probability, but Not Vice Versa	115
6.2.2	Convergence in Probability Implies Convergence in Distribution, but Not Vice Versa	116
6.2.3	Almost Sure Convergence Implies Convergence in Probability, but Not Vice Versa	117
6.2.4	Subsequence Arguments	120
6.2.5	Melding Modes: Slutsky's Theorem	124
6.3	Convergence of Random Vectors	126
6.4	Summary	128
7	Laws of Large Numbers	129
7.1	Basic Laws and Applications	129
7.2	Proofs and Extensions	138
7.3	Random Walks	142
7.4	Summary	148
8	Central Limit Theorems	149
8.1	CLT for iid Random Variables and Applications	149
8.2	CLT for Non iid Random Variables	155
8.3	Harmonic Regression	163
8.4	Characteristic Functions	166
8.5	Proof of Standard CLT	173
8.5.1	Proof of CLT for Symmetric Random Variables	174
8.5.2	Proof of CLT for Arbitrary Random Variables	174
8.6	Multivariate Ch.f.s and CLT	176
8.7	Summary	182
9	More on Convergence in Distribution	183
9.1	Uniform Convergence of Distribution Functions	183
9.2	The Delta Method	188
9.3	Convergence of Moments: Uniform Integrability	193
9.4	Normalizing Sequences	196
9.5	Review of Equivalent Conditions for Weak Convergence	197

9.6 Summary	199
10 Conditional Probability and Expectation	201
10.1 When There is a Density or Mass Function	202
10.2 More General Definition of Conditional Expectation	205
10.3 Regular Conditional Distribution Functions	211
10.4 Conditional Expectation As a Projection	217
10.5 Conditioning and Independence	220
10.6 Sufficiency	223
10.6.1 Sufficient and Ancillary Statistics	223
10.6.2 Completeness and Minimum Variance Unbiased Estimation	224
10.6.3 Basu's Theorem and Applications	225
10.6.4 Conditioning on Ancillary Statistics	226
10.7 Expect the Unexpected from Conditional Expectation	229
10.7.1 Conditioning on Sets of Probability 0	230
10.7.2 Substitution in Conditioning Expressions	231
10.7.3 Weak Convergence of Conditional Distributions	235
10.8 Conditional Distribution Functions As Derivatives	237
10.9 Appendix: Radon-Nikodym Theorem	239
10.10 Summary	239
11 Applications	241
11.1 $F(X) \sim U[0, 1]$ and Asymptotics	241
11.2 Asymptotic Power and Local Alternatives	243
11.2.1 T-Test	244
11.2.2 Test of Proportions	245
11.2.3 Summary	246
11.3 Insufficient Rate of Convergence in Distribution	246
11.4 Failure to Condition on All Information	248
11.5 Failure to Account for the Design	249
11.5.1 Introduction and Simple Analysis	249
11.5.2 Problem with the Proposed Method	250
11.5.3 Connection to Fisher's Exact Test	250

11.5.4	A Flawed Argument that $P(Z_{n1} \leq z_1 Z_{n2} = z_2) \rightarrow \Phi(z_1)$	251
11.5.5	Fixing the Flaw: Polya's Theorem to the Rescue	251
11.5.6	Conclusion: Asymptotics of the Hypergeometric Distribution	252
11.6	Validity of Permutation Tests: I	253
11.7	Validity of Permutation Tests: II	255
11.7.1	A Vaccine Trial Raising Validity Questions	255
11.7.2	Assumptions Ensuring Validity	255
11.7.3	Is a Z-Test of Proportions Asymptotically Valid?	257
11.8	Validity of Permutation Tests III	258
11.8.1	Is Adaptive Selection of Covariates Valid?	258
11.8.2	Sham Covariates Reduce Variance: What's the Rub?	259
11.8.3	A Disturbing Twist	260
11.9	A Brief Introduction to Path Diagrams	261
11.10	Estimating the Effect Size	267
11.10.1	Background and Introduction	267
11.10.2	An Erroneous Application of Slutsky's Theorem	268
11.10.3	A Correct Application of Slutsky's Theorem	268
11.10.4	Transforming to Stabilize The Variance: The Delta Method	269
11.11	Asymptotics of An Outlier Test	270
11.11.1	Background and Test	270
11.11.2	Inequality and Asymptotics for $1 - \Phi(x)$	271
11.11.3	Relevance to Asymptotics for the Outlier Test	272
11.12	An Estimator Associated with the Logrank Statistic	274
11.12.1	Background and Goal	274
11.12.2	Distributions of $(X_i n_{Ti}, n_{Ci})$ and $X_i - E(X_i n_{Ti}, n_{Ci})$	275
11.12.3	The Relative Error of $\hat{\tau}_n$	276
A	Whirlwind Tour of Prerequisites	279
A.1	A Key Inequality	279
A.2	The Joy of Sets	279
A.2.1	Basic Definitions and Results	279
A.2.2	Sets of Real Numbers: Inf and Sup	281

A.3	A Touch of Topology of R^k	282
A.3.1	Open and Closed Sets	282
A.3.2	Compact Sets	286
A.4	Sequences in R^k	288
A.5	Series	291
A.6	Functions	294
A.6.1	Mappings	294
A.6.2	Limits and Continuity of Functions	295
A.6.3	The Derivative of a Function of k Variables	298
B	Common Probability Distributions	301
B.1	Discrete Distributions	301
B.2	Continuous Distributions	303
B.3	Relationships between Distributions	305
C	References	309
D	Mathematical Symbols and Abbreviations	313
	Index	317

Preface

Biostatistics and statistics departments are struggling with how much probability and measure theory to include in their curricula. The traditional statistics department model of a full year of probability using the texts of Billingsley or Chung, for example, is losing favor as students are in a rush to get through a graduate program and begin their careers. Some biostatistics departments have gone to the extreme of eliminating graduate-level probability altogether. Consequently, their students are left with a background that does not prepare them to make rigorous arguments. We wrote this book as a compromise between the two unpalatable extremes: overloading statistics students with extensive and mathematically challenging measure theory versus leaving them unprepared to prove their results. Rather than offering a comprehensive treatment of all of probability and measure theory, replete with proofs, we present the essential probability results that are used repeatedly in statistics applications. We also selectively present proofs that make repeated use of mathematical techniques that we continue to use in our statistical careers when rigor is needed. As biostatisticians, we have encountered numerous applications requiring careful use of probability. We share these in this book, whose emphasis is on being able to understand rigorously the meaning and application of probability results. While traditional graduate probability books are sometimes written for mathematics students with no knowledge of elementary statistics, our book is written with statisticians in mind. For example, we motivate characteristic functions by first discussing harmonic regression and its usefulness in understanding circadian rhythm of biological phenomena. Another example of our use of statistical applications to help understand and motivate probability is the study of permutation tests. Permutation tests provide fertile ground for understanding conditional distributions and asymptotic arguments. For example, it is both challenging and instructive to try to understand precisely what people mean when they assert the asymptotic equivalence of permutation and t-tests. In summary, we believe that this book is ideal for teaching students essential probability theory to make rigorous probability arguments.

The book is organized as follows. The first chapter is intended as a broad introduction to why more rigor is needed to take that next step to graduate-level probability. Chapter 1 makes use of intriguing paradoxes to motivate the need for rigor. The chapter assumes knowledge of basic probability and statistics. We return to some of these paradoxes later in the book. Chapter 2, on countability, contains essential background material with which some readers may already be familiar. Depending on the background of students, instructors may want to begin with this chapter or have students review it on their own and refer to it when needed. The remaining chapters are in the order that we believe is most logical. Chapters 3 and 4 contain the backbone of probability: sigma-fields, probability measures, and random variables and vectors. Chapter 5 introduces and contrasts Lebesgue-Stieltjes integration with the more familiar Riemann-Stieltjes integration, while Chapter 6 covers different modes of convergence. Chapters 7 and 8 concern laws of large numbers and central

limit theorems. Chapter 9 contains additional results on convergence in distribution, including the delta method, while Chapter 10 covers the extremely important topic of conditional probability, expectation, and distribution. Chapter 11 contains many interesting applications from our actual experience. Other applications are interspersed throughout the book, but those in Chapter 11 are more detailed. Many examples in Chapter 11 rely on material covered in Chapters 9-10, and would therefore be difficult to present much earlier. The book concludes with two appendices. Appendix A is a brief review of prerequisite material, while Appendix B contains useful probability distributions and their properties. Each chapter contains a chapter review of key results, and exercises are intended to constantly reinforce important concepts.

We would like to express our extreme gratitude to Robert Taylor (Clemson University), Jie Yang (University of Illinois at Chicago), Wlodek Byrc (University of Cincinnati), and Radu Herbei (Ohio State University) for reviewing the book. They gave us very helpful suggestions and additional material, and caught typos and other errors. The hardest part of the book for us was constructing exercises, and the reviewers provided additional problems and suggestions for those as well.

Index of Statistical Applications and Notable Examples

In this text, we illustrate the importance of probability theory to statistics. We have included a number of illustrative examples that use key results from probability to gain insight on the behavior of some commonly used statistical tests, as well as examples that consider implications for design of clinical trials. Here, we highlight our more statistical applications and a few other notable examples.

Statistical Applications

1. Clinical Trials
 - i. ECMO Trial: Example 10.45
 - ii. Interim Monitoring: Example 8.51
 - iii. Permuted block randomization: Example 1.1, 5.33, 7.3, 7.19, 10.20
 - iv. Two-stage study design: Example 10.44
2. Test statistics
 - i. Effect size: Section 11.10
 - ii. Fisher's Exact test: Section 11.5.3
 - iii. Goodness of fit test: Example 8.49
 - iv. Likelihood ratio test: Section 11.3
 - v. Logrank statistic: Section 11.12
 - vi. Maximum test statistic: Example 9.1
 - vii. Normal scores test: Example 4.58
 - viii. Outlier test: Section 11.11
 - ix. P-value: Example 9.1, 9.3
 - x. Permutation test: Example 8.10, 8.20, 10.20, 10.45, 10.53; Sections 11.6, 11.7, 11.8
 - xi. Pitman's test : Example 10.35
 - xii. Sample mean: Example 4.44, 6.6, 10.34
 - xiii. Sample median: Example 6.10, 7.4; Exercise: Section 7.1, #3

- xiv. Sample variance: Example 4.44, 6.53, 7.6, 8.5, 9.7; Exercise: Section 7.1, #4
 - xv. Sample covariance: Example 8.6
 - xvi. Test of proportions: Section 11.2.2, 11.5.3, 11.7.3
 - xvii. t-test: Example 7.6, 8.3, 8.4, 10.54; Section 11.2.1
3. Multiplicity and hypothesis testing
 - i. Bonferroni correction: Example 6.25; Section 11.4, 11.11.1
 - ii. Dunnett's test: Exercise: Section 10.3, #3
 - iii. Fisher's least significance difference procedure: Exercise: Section 10.3, #5
 4. Regression and analysis of variance
 - i. Correlation coefficient: Example 7.7
 - ii. Harmonic regression: Section 8.3
 - iii. Mixed model: Example 10.34, 10.36
 - iv. Regression: Example 7.7, 8.7; Section 11.8; Exercise: Section 6.2.3, #7, Section 7.1, #9
 5. Other statistical applications
 - i. Bland Altman plot: Example 10.35
 - ii. Causal (Path) diagrams: Section 11.7, 11.9
 - iii. Copulas: Example 4.59
 - iv. Missing at random: Example 10.37
 - v. Regression to the mean: Example 10.38
 - vi. Relative risk: Example 6.6, 9.12, 9.16
 - vii. Spline fitting: Example 7.9
 - viii. Survival function: Example 6.9, 6.26, 7.4
 - ix. Variance stabilization: Example 9.11, 11.10.4

Other Notable Examples

- i. Axiom of choice: Example 2.14
- ii. Balls in box paradox: Example 1.5
- iii. Bible code: Example 3.1
- iv. Binary expansion: Example 1.1, 4.38, 5.33,
- v. Borel's paradox: Example 1.3, 10.48
- vi. Dyadic rationals: 6.22
- vii. Face on Mars: Example 3.37, 4.51
- viii. No density: Example 1.1, 1.2, 7.3
- ix. Quincunx: Example 8.10, 8.11, 8.12, 8.13
- x. Two envelope problem: Example 1.4, 10.46