

Contents

<i>List of Tables and Figures</i>	<i>page</i> xii
<i>Introduction</i>	xv
PART I PROBABILITY	1
1 Probability Theory	3
1.1 Random Events	4
1.1.1 What Is Probability?	4
1.1.2 Fundamentals of Probability Theory	5
1.1.3 Joint and Conditional Probabilities	9
1.1.4 Independence of Events	14
1.2 Random Variables	15
1.2.1 What Is a Random Variable?	16
1.2.2 Discrete Random Variables	18
1.2.3 Cumulative Distribution Functions	21
1.2.4 Continuous Random Variables	24
1.2.5 Support	30
1.3 Bivariate Relationships	31
1.3.1 Discrete Bivariate Distributions	32
1.3.2 Discrete Marginal and Conditional Distributions	33
1.3.3 Jointly Continuous Random Variables	36
1.3.4 Independence of Random Variables	38
1.4 Multivariate Generalizations	39
1.5 Further Readings	43
	vii

viii	<i>Contents</i>
2	Summarizing Distributions 44
2.1	<i>Summary Features of Random Variables</i> 45
2.1.1	<i>Expected Values</i> 45
2.1.2	<i>Moments, Variances, and Standard Deviations</i> 50
2.1.3	<i>Mean Squared Error</i> 56
2.2	<i>Summary Features of Joint Distributions</i> 59
2.2.1	<i>Covariance and Correlation</i> 59
2.2.2	<i>Covariance, Correlation, and Independence</i> 64
2.2.3	<i>Conditional Expectations and Conditional Expectation Functions</i> 67
2.2.4	<i>Best Predictors and Best Linear Predictors</i> 75
2.2.5	<i>CEFs and BLPs under Independence</i> 82
2.3	<i>Multivariate Generalizations</i> 84
2.4	<i>Further Readings</i> 88
PART II	STATISTICS 89
3	Learning from Random Samples 91
3.1	<i>I.I.D. Random Variables</i> 91
3.1.1	<i>Random Sampling</i> 92
3.1.2	<i>I.I.D. as Approximation</i> 94
3.2	<i>Estimation</i> 96
3.2.1	<i>Sample Means</i> 96
3.2.2	<i>Estimation Theory</i> 102
3.2.3	<i>Variance Estimators</i> 105
3.2.4	<i>The Central Limit Theorem for Sample Means</i> 108
3.2.5	<i>Asymptotic Estimation Theory</i> 111
3.2.6	<i>Estimating Standard Errors of Sample Means</i> 114
3.3	<i>The Plug-In Principle</i> 116
3.3.1	<i>The Usual Plug-In Regularity Conditions</i> 120
3.3.2	<i>Kernel Estimation</i> 121
3.4	<i>Inference</i> 124
3.4.1	<i>Confidence Intervals</i> 124
3.4.2	<i>Hypothesis Testing</i> 128
3.4.3	<i>The Bootstrap</i> 130
3.4.4	<i>Micronumerosity</i> 132
3.5	<i>Cluster Samples</i> 135

<i>Contents</i>	ix
3.5.1 <i>Estimation with Clustering</i>	136
3.5.2 <i>Inference with Clustering</i>	140
3.6 <i>Further Readings</i>	141
4 <i>Regression</i>	143
4.1 <i>Regression Estimation</i>	143
4.1.1 <i>Bivariate Case</i>	144
4.1.2 <i>OLS Regression</i>	145
4.1.3 <i>Regression with Matrix Algebra</i>	147
4.2 <i>Inference</i>	151
4.2.1 <i>Standard Errors and Inference</i>	151
4.2.2 <i>Inference with Robust Standard Errors and the Bootstrap</i>	152
4.2.3 <i>Classical Standard Errors</i>	155
4.3 <i>Estimation of Nonlinear Conditional Expectation Functions</i>	156
4.3.1 <i>Polynomials</i>	158
4.3.2 <i>Overfitting</i>	161
4.3.3 <i>Interactions</i>	164
4.3.4 <i>Summarizing Partial Derivatives of the CEF</i>	165
4.3.5 <i>Sieve Estimation</i>	168
4.3.6 <i>Penalized Regression</i>	169
4.4 <i>Application: Access to Clean Water and Infant Mortality</i>	170
4.5 <i>Further Readings</i>	177
5 <i>Parametric Models</i>	178
5.1 <i>Models and Parameters</i>	178
5.1.1 <i>The Classical Linear Model</i>	180
5.1.2 <i>Binary Choice Models</i>	182
5.2 <i>Maximum Likelihood Estimation</i>	185
5.2.1 <i>The Logic of Maximum Likelihood Estimation</i>	185
5.2.2 <i>Maximum Likelihood Estimation when the Parametric Model Is True</i>	193
5.2.3 <i>Maximum Likelihood Estimation when the Parametric Model Is Not True</i>	194
5.2.4 <i>Maximum Likelihood Plug-In Estimation</i>	197
5.2.5 <i>Mixture Models</i>	198
5.2.6 <i>Penalized Maximum Likelihood Regression</i>	201
5.2.7 <i>Inference</i>	202

5.3	<i>A Note on Models as Approximations</i>	203
5.4	<i>Further Readings</i>	204
PART III IDENTIFICATION		205
6	Missing Data	207
6.1	<i>Identification with Missing Data [7.1]</i>	208
6.1.1	<i>Bounds [7.1.3]</i>	209
6.1.2	<i>Missing Completely at Random [7.1.4]</i>	213
6.1.3	<i>Missing at Random [7.1.5]</i>	215
6.1.4	<i>The Role of the Propensity Score for Missing Data [7.1.6]</i>	217
6.2	<i>Estimation with Missing Data under MAR [7.2]</i>	219
6.2.1	<i>Plug-In Estimation [7.2.1]</i>	219
6.2.2	<i>Regression Estimation [7.2.2]</i>	222
6.2.3	<i>Hot Deck Imputation [7.2.4]</i>	224
6.2.4	<i>Maximum Likelihood Plug-In Estimation of Propensity Scores [7.2.5]</i>	225
6.2.5	<i>Weighting Estimators [7.2.6]</i>	226
6.2.6	<i>Doubly Robust Estimators [7.2.7]</i>	228
6.3	<i>Application: Estimating the Cross-National Average of Clean Energy Use</i>	231
6.4	<i>Further Readings</i>	234
7	Causal Inference	235
7.1	<i>Identification with Potential Outcomes [6.1]</i>	236
7.1.1	<i>Framework</i>	236
7.1.2	<i>Ties to Missing Data</i>	238
7.1.3	<i>Bounds [6.1.1]</i>	240
7.1.4	<i>Random Assignment [6.1.2]</i>	244
7.1.5	<i>Ignorability [6.1.3]</i>	247
7.1.6	<i>The Role of the Propensity Score for Causal Inference [6.1.4]</i>	250
7.1.7	<i>Post-Treatment Variables</i>	252
7.1.8	<i>Generalizing Beyond Binary Treatments</i>	254
7.2	<i>Estimation of Causal Effects under Ignorability [6.2]</i>	256
7.2.1	<i>Plug-In Estimation [6.2.1]</i>	256
7.2.2	<i>Regression Estimation [6.2.2]</i>	258
7.2.3	<i>Maximum Likelihood Plug-In Estimation of Causal Effects</i>	261

<i>Contents</i>	xi
7.2.4 <i>Matching</i> [6.2.3]	262
7.2.5 <i>Maximum Likelihood Plug-In Estimation of Propensity Scores</i> [6.2.4]	264
7.2.6 <i>Weighting Estimators</i> [6.2.5]	264
7.2.7 <i>Doubly Robust Estimators</i> [6.2.6]	267
7.2.8 <i>Placebo Testing</i>	270
7.3 <i>Overlap and Positivity</i>	271
7.3.1 <i>Changing the Target Population</i>	273
7.3.2 <i>Empirical Overlap, Micronumerosity, and Weighting Estimators</i>	273
7.4 <i>Further Extensions</i>	276
7.5 <i>Application: The Effect of Gender on Swiss Citizenship Approval Votes</i>	276
7.6 <i>Further Readings</i>	280
Glossary of Mathematical Notation	282
Glossary of Common Abbreviations	286
<i>References</i>	288
<i>Index</i>	293