

Contents

List of Figures	page xi
List of Tables	xv
Preface	xvii
Acknowledgments	xxiii
Notes on Notation	xxv
PART I CONCEPTS, THEORY, AND IMPLEMENTATION	
1 Introduction to Maximum Likelihood	3
1.1 Introduction to Maximum Likelihood	3
1.2 Coin Flips and Maximum Likelihood	3
1.3 Samples and Sampling Distributions	6
1.4 Maximum Likelihood: An Overview	9
1.4.1 Maximum Likelihood: Specific	9
1.4.2 The Least Squares Approach	16
1.5 Maximum Likelihood: General	17
1.6 Conclusion	19
1.7 Further Reading	20
2 Theory and Properties of Maximum Likelihood Estimators	21
2.1 The Likelihood Function: A Deeper Dive	21
2.1.1 Likelihood and Bayesian Statistics	22
2.1.2 Regularity	23
2.1.3 Score	24
2.1.4 Fisher Information	25
2.2 Properties of Maximum Likelihood Estimators	28
2.2.1 Invariance	28
2.2.2 Large-Sample Properties	28

2.3	<i>Diagnostics for Maximum Likelihood</i>	32
2.3.1	<i>Likelihood Ratios</i>	32
2.3.2	<i>Score and Wald Tests</i>	35
2.3.3	<i>Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)</i>	37
2.4	<i>What If the Likelihood Is Wrong?</i>	38
2.4.1	<i>Robust Variance and Standard Errors</i>	39
2.5	<i>Conclusion</i>	40
2.6	<i>Further Reading</i>	41
3	<i>Maximum Likelihood for Binary Outcomes</i>	43
3.1	<i>Binary Responses</i>	43
3.2	<i>Binary Data</i>	44
3.2.1	<i>Odds, Odds Ratios, and Relative Risk</i>	45
3.3	<i>The Linear Probability Model</i>	47
3.4	<i>The Logit Model, a.k.a. Logistic Regression</i>	50
3.4.1	<i>The Logit Model</i>	51
3.4.2	<i>Estimation</i>	52
3.4.3	<i>Output and Interpretation</i>	54
3.4.4	<i>Estimation Uncertainty around the MLE</i>	58
3.4.5	<i>Graphical Presentation of Effects in Logistic Regression</i>	60
3.5	<i>An Example: US Senate Voting</i>	62
3.5.1	<i>Model Evaluation</i>	64
3.6	<i>Some Extensions</i>	71
3.6.1	<i>Latent Variable Formulation for Binary Data</i>	71
3.6.2	<i>Heteroskedastic Probit</i>	72
3.6.3	<i>Complimentary Logistic Model, a.k.a. cloglog</i>	73
3.6.4	<i>An Example: Civil Conflict</i>	74
3.7	<i>Summary of General Approach to Binary Variables</i>	76
3.8	<i>Further Reading</i>	77
4	<i>Implementing MLE</i>	79
4.1	<i>The Likelihood Surface</i>	79
4.1.1	<i>Examples</i>	80
4.1.2	<i>Profile Likelihood</i>	82
4.2	<i>Numerical Optimization</i>	82
4.2.1	<i>Newton-Raphson</i>	83
4.2.2	<i>Other Algorithms</i>	84
4.2.3	<i>Expectation Maximization (EM)</i>	85
4.3	<i>Estimation Challenges</i>	86
4.3.1	<i>Optimizer Fails to Converge</i>	86
4.3.2	<i>Starting Values</i>	86
4.3.3	<i>Algorithm Choice</i>	86
4.3.4	<i>Scaling of Variables</i>	87
4.3.5	<i>Step Sizes</i>	87

4.3.6	<i>Flat Likelihoods</i>	87
4.3.7	<i>Absence of Single Maximum</i>	88
4.4	<i>Different Software Yields “Different” Results</i>	88
4.5	<i>Modeling Challenges Appear as Estimation Problems</i>	89
4.5.1	<i>(Near-)Perfect Separation</i>	89
4.5.2	<i>Rare Events (and Small Samples)</i>	91
4.6	<i>Conclusion</i>	91
4.7	<i>Further Reading</i>	92

PART II MODEL EVALUATION AND INTERPRETATION

5	<i>Model Evaluation and Selection</i>	97
5.1	<i>Current Practice</i>	97
5.1.1	<i>BUTON</i>	98
5.1.2	<i>Graphical Displays</i>	99
5.1.3	<i>What’s Left Out?</i>	100
5.1.4	<i>Sign, Significance, and the Perils of p-Values</i>	102
5.2	<i>The Logic of Prediction, Out-of-Sample</i>	104
5.2.1	<i>The Process of Evaluating Out-of-Sample Predictions</i>	105
5.2.2	<i>Variations on a Theme</i>	110
5.2.3	<i>Out-of-Sample Prediction: Trade and the WTO, Act II</i>	112
5.2.4	<i>Benefits of Cross-Validation</i>	115
5.3	<i>Conclusion</i>	115
5.4	<i>Further Reading</i>	116
6	<i>Inference and Interpretation</i>	119
6.1	<i>The Mechanics of Inference</i>	119
6.2	<i>Interpreting Models and Producing Scenarios of Interest</i>	122
6.2.1	<i>Quantities of Interest</i>	122
6.2.2	<i>Scenarios</i>	123
6.2.3	<i>Interpolation, Extrapolation, and Convex Hulls</i>	124
6.2.4	<i>Simulating Estimation Uncertainty</i>	126
6.3	<i>Conclusion</i>	130
6.4	<i>Further Reading</i>	130

PART III THE GENERALIZED LINEAR MODEL

7	<i>The Generalized Linear Model</i>	135
7.1	<i>The Generalized Linear Model</i>	135
7.2	<i>The Exponential Family</i>	136
7.3	<i>Past Examples as GLMs</i>	137
7.3.1	<i>GLMs in $\mathcal{R}$</i>	138
7.4	<i>“Quasi-” and “Pseudo-”likelihood</i>	138
7.5	<i>Conclusion</i>	140
7.6	<i>Further Reading</i>	140

8	Ordered Categorical Variable Models	141
8.1	<i>Motivation</i>	141
8.2	<i>The Ordered Logit Model</i>	142
8.2.1	<i>What about the Ordered Probit Model?</i>	145
8.2.2	<i>Results and Interpretation</i>	145
8.3	<i>Example: Categorical Assessments of Political Ideology</i>	146
8.4	<i>Parallel Regressions</i>	149
8.4.1	<i>Example: Genocide Severity</i>	151
8.4.2	<i>Parallel Regressions and Genocide Severity</i>	152
8.4.3	<i>Extensions</i>	154
8.5	<i>Ordered Categorical Variables as Regressors</i>	154
8.5.1	<i>An Example: Child Mortality and Democracy</i>	157
8.6	<i>Further Reading</i>	160
9	Models for Nominal Data	161
9.1	<i>Introduction</i>	161
9.2	<i>Multinomial Logit</i>	162
9.2.1	<i>A Latent Variable Formulation</i>	164
9.2.2	<i>IIA</i>	166
9.3	<i>An Example: Australian Voters in 2013</i>	168
9.3.1	<i>Evaluation</i>	168
9.3.2	<i>Interpretation</i>	173
9.4	<i>Conditional Multinomial Logit</i>	179
9.4.1	<i>A Note on Data Structure</i>	180
9.5	<i>Extensions</i>	184
9.5.1	<i>Heteroskedastic Multinomial Logit</i>	184
9.5.2	<i>Nested Logit</i>	185
9.5.3	<i>Multinomial Probit</i>	187
9.5.4	<i>Random Coefficients and Mixed Logit</i>	187
9.6	<i>Conclusion</i>	188
9.7	<i>Further Reading</i>	189
10	Strategies for Analyzing Count Data	190
10.1	<i>Introduction</i>	190
10.2	<i>The Poisson Distribution</i>	191
10.2.1	<i>An Example: Means and Mediation</i>	192
10.2.2	<i>Interpretation</i>	195
10.3	<i>Dispersion</i>	197
10.3.1	<i>Diagnosing Over-Dispersion</i>	198
10.4	<i>Modeling Over-Dispersion</i>	201
10.4.1	<i>Quasipoisson</i>	202
10.4.2	<i>The Negative Binomial Model</i>	202
10.4.3	<i>Mediations</i>	205
10.4.4	<i>Under-Dispersion</i>	207

10.5	<i>Hurdling 0s and 1s</i>	207
10.5.1	<i>How Many 0s Are Too Many?</i>	209
10.5.2	<i>Hurdle Models</i>	210
10.5.3	<i>Zero Inflation</i>	211
10.5.4	<i>Example: Droughts and Atrocities</i>	212
10.6	<i>Summary</i>	213
10.7	<i>Further Reading</i>	215
PART IV ADVANCED TOPICS		
11	<i>Strategies for Temporal Dependence: Duration Models</i>	219
11.1	<i>Introduction</i>	219
11.2	<i>Data Structures and Thinking about Time</i>	220
11.2.1	<i>Discrete and Continuous Time</i>	221
11.2.2	<i>Organizing Data as a Spell-Level</i>	221
11.2.3	<i>Continuous Time with Time-Varying Covariates: Counting Process</i>	222
11.3	<i>Discrete Time: The BTSCS Approach</i>	224
11.4	<i>Duration Models</i>	226
11.4.1	<i>Survivor and Hazard Functions</i>	226
11.4.2	<i>A Likelihood</i>	227
11.4.3	<i>Writing Down a Duration Model</i>	228
11.5	<i>The Cox Model</i>	230
11.5.1	<i>The Partial Likelihood</i>	231
11.5.2	<i>An Example: Cabinet Durations</i>	233
11.5.3	<i>Pros and Cons of the Cox Model</i>	235
11.6	<i>A Split-Population Model</i>	237
11.6.1	<i>Interpretation of Split-Duration Estimates</i>	241
11.6.2	<i>An Example</i>	243
11.7	<i>Conclusion</i>	245
11.8	<i>Further Reading</i>	247
12	<i>Strategies for Missing Data</i>	249
12.1	<i>Introduction</i>	249
12.1.1	<i>Concepts and Notation</i>	250
12.2	<i>Finding Missingness</i>	251
12.3	<i>A Taxonomy of Missing Data</i>	256
12.4	<i>Common Strategies</i>	257
12.4.1	<i>Complete Case Analysis</i>	257
12.4.2	<i>Available Case Analysis</i>	258
12.4.3	<i>Mean Imputation</i>	258
12.4.4	<i>Conditional Mean Imputation</i>	258
12.5	<i>Multiple Imputation</i>	259
12.5.1	<i>MLE-EM and Bayesian Multiple Imputation</i>	259
12.5.2	<i>MICE</i>	260

12.5.3	Copula Methods	261
12.5.4	Combining Results Across Imputations	263
12.6	Building an Imputation Model	264
12.7	Evaluating Imputation Models	265
12.7.1	Empirical Replications: Sorens and Ruger	265
12.8	Conclusion	268
12.9	Further Reading	269

PART V A LOOK AHEAD

13	Epilogue	273
13.1	The Broader Landscape	274
	Bibliography	277
	Index	293