

Contents

Foreword by Judea Pearl	vii
Foreword by Ira B. Tager	xi
Preface	xv
List of Contributors	xliii
Abbreviations and Notation	xlvi
Models, Inference, and Truth	li
Richard J.C.M. Starmans	
Part I Targeted Learning: The Basics	
1 The Open Problem	3
Sherri Rose, Mark J. van der Laan	
1.1 Learning from Data	4
1.2 Traditional Approach to Estimation	6
1.2.1 Experimental Studies	7
1.2.2 Observational Studies	8
1.2.3 Regression in (Misspecified) Parametric Statistical Models	9
1.2.4 The Complications of Human Art in Statistics	11
1.3 Data, Model, and Target Parameter	13
1.3.1 The Model	14
1.3.2 The Target Parameter	15
1.3.3 Summary of Concepts	16
1.4 The Need for Targeted Estimators	17
1.5 Road Map for Targeted Learning	17
1.6 Notes and Further Reading	20

2	Defining the Model and Parameter	21
	Sherri Rose, Mark J. van der Laan	
2.1	Defining the Structural Causal Model	22
2.2	Causal Graphs	27
2.2.1	Terminology	27
2.2.2	Assumptions	29
2.3	Defining the Causal Target Parameter	30
2.3.1	Interventions	30
2.3.2	Counterfactuals	32
2.3.3	Establishing Identifiability	33
2.3.4	Commit to a Statistical Model and Target Parameter	34
2.3.5	Interpretation of Target Parameter	36
2.4	Revisiting the Mortality Example	37
2.5	Road Map for Targeted Learning	37
2.6	Conceptual Framework	39
2.7	Notes and Further Reading	41
3	Super Learning	43
	Eric C. Polley, Sherri Rose, Mark J. van der Laan	
3.1	Background	44
3.2	Defining the Estimation Problem	46
3.3	Super (Machine) Learning	47
3.3.1	Discrete Super Learner	48
3.3.2	Super Learner	50
3.3.3	Finite Sample Performance and Applications	52
3.4	Road Maps	59
3.5	Conceptual Framework of Loss-Based Super Learning	62
3.6	Notes and Further Reading	65
4	Introduction to TMLE	67
	Sherri Rose, Mark J. van der Laan	
4.1	Motivation	69
4.2	TMLE in Action: Mortality Study Example	70
4.2.1	Estimator	71
4.2.2	Inference	75
4.3	Practical Implications	76
4.3.1	Randomized Controlled Trials	76
4.3.2	Observational Studies	78
4.4	Summary	79
4.5	Road Map for Targeted Learning	79
4.6	Notes and Further Reading	82

5	Understanding TMLE	83
	Sherri Rose, Mark J. van der Laan	
5.1	Conceptual Framework	83
5.2	Definition of TMLE in Context of the Mortality Example	87
5.2.1	Estimating \bar{Q}_0	88
5.2.2	Estimating g_0	88
5.2.3	Determining the Efficient Influence Curve $D^*(P)$	89
5.2.4	Determining the Fluctuation Working Model	89
5.2.5	Updating \bar{Q}_n^0	93
5.2.6	Estimating the Target Parameter	94
5.2.7	Calculating Standard Errors	94
5.3	Foundation and Philosophy of TMLE	97
5.4	Summary	100
6	Why TMLE?	101
	Sherri Rose, Mark J. van der Laan	
6.1	Landscape	102
6.1.1	MLE	102
6.1.2	Estimating Equation Methods	105
6.2	TMLE is Based on (Targeted) Loss-Based Learning	106
6.2.1	Competitors	107
6.2.2	TMLE	107
6.3	TMLE Is Well Defined	108
6.3.1	Competitors	108
6.3.2	TMLE	109
6.4	TMLE Is Unbiased	110
6.4.1	Competitors	111
6.4.2	TMLE	111
6.5	TMLE Is Efficient	113
6.5.1	Competitors	115
6.5.2	TMLE	115
6.6	TMLE Is a Substitution Estimator	116
6.6.1	Competitors	116
6.6.2	TMLE	117
6.7	Summary	117
6.8	Notes and Further Reading	117

Part II Additional Core Topics

7	Bounded Continuous Outcomes	121
	Susan Gruber, Mark J. van der Laan	
7.1	Introduction	121
7.2	TMLE for Causal Estimation on a Continuous Outcome	123
7.2.1	A Substitution Estimator Respecting the Statistical Model	123
7.2.2	Procedure	124
7.2.3	Robustness of TMLE in the Context of Sparsity	128

7.3	Simulations	128
7.3.1	Estimators	129
7.3.2	Data-Generating Distributions	130
7.3.3	Results	130
7.4	Discussion	131
	Appendix	132
8	Direct Effects and Effect Among the Treated	133
	Alan E. Hubbard, Nicholas P. Jewell, Mark J. van der Laan	
8.1	Defining the Causal Direct Effect	133
8.2	TMLE	135
8.3	Simulation	139
8.4	Data Analysis	139
8.4.1	Unadjusted Mean Difference in Salary	140
8.4.2	Adjusted Mean Difference in Salary: $\Psi_{NDE}(P_0)$	141
8.5	Discussion	142
8.6	Notes and Further Reading	143
9	Marginal Structural Models	145
	Michael Rosenblum	
9.1	Impact of Missing Doses on Virologic Failure	145
9.2	Data	147
9.3	Statistical Model	147
9.4	Parameter of Interest	148
9.5	Effect Modification: Simplified Case	149
9.5.1	Obtaining Q_n^0 , an Initial Estimate of Q_0	149
9.5.2	Calculating the Optimal Fluctuation	150
9.5.3	Obtaining Q_n^* , a Targeted Estimate of Q_0	151
9.5.4	Estimation of Parameter	151
9.6	Effect Modification: Marginal Structural Models	152
9.6.1	Obtaining Q_n^0 , an Initial Estimate of Q_0	154
9.6.2	Calculating the Optimal Fluctuation	154
9.6.3	Obtaining Q_n^* , a Targeted Estimate of Q_0	155
9.6.4	Estimation of Parameter	156
9.7	Constructing Confidence Intervals	157
9.8	Results	157
9.9	Discussion	159
9.10	Notes and Further Reading	159
10	Positivity	161
	Maya L. Petersen, Kristin E. Porter, Susan Gruber, Yue Wang, Mark J. van der Laan	
10.1	Framework for Causal Effect Estimation	162
10.2	Estimator-Specific Behavior Under Positivity Violations	166
10.2.1	MLE	166
10.2.2	IPTW Estimator	167

10.2.3	Double Robust Estimators	168
10.3	Diagnosing Bias Due to Positivity Violations	169
10.3.1	The Parametric Bootstrap as a Diagnostic Tool	169
10.3.2	Simulations	172
10.3.3	HIV Data Application	176
10.4	Practical Approaches to Positivity Violations	178
10.5	Discussion	183
10.6	Notes and Further Reading	184

Part III TMLE and Parametric Regression in RCTs

11	Robust Analysis of RCTs Using Generalized Linear Models	187
	Michael Rosenblum	
11.1	Summary of Main Result	188
11.2	The Generalized Linear (Working) Models	190
11.3	TMLE Using Generalized Linear Model in Initial Estimator	192
11.3.1	Parameter as a Mapping from the Distribution of the Data	193
11.3.2	Obtaining Q_n^0 , an Initial Estimate of Q_0	194
11.3.3	Loss Function for TMLE Step	194
11.3.4	Calculating the Optimal Fluctuation/Submodel	194
11.3.5	Obtaining Q_n^* , a Targeted Estimate of Q_0	196
11.3.6	Estimation of Marginal Treatment Effect	196
11.4	Main Theorem	196
11.5	Special Robustness of Poisson Model with Only Main Terms	197
11.6	Standard Errors and Confidence Intervals	198
11.7	Discussion	198
11.8	Notes and Further Reading	199
12	Targeted ANCOVA Estimator in RCTs	201
	Daniel B. Rubin, Mark J. van der Laan	
12.1	Previously Proposed Estimators	204
12.2	Targeted ANCOVA	208
12.2.1	Asymptotic Optimality	209
12.2.2	Targeted ANCOVA Is a Substitution Estimator	210
12.2.3	Small and Moderate Sample Performance	210
12.3	Standard Error Estimation	211
12.4	Simulations	212
12.5	Discussion	214

Part IV Case-Control Studies

13	Independent Case-Control Studies	219
	Sherri Rose, Mark J. van der Laan	
13.1	Data, Model, and Target Parameter	219
13.2	Prevalence Probability	221
13.3	CCW-TMLE	221

13.3.1	Case-Control-Weighted Estimators for Q_0 and g_0	222
13.3.2	Parametric Submodel for Full-Data TMLE	223
13.3.3	Obtaining a Targeted Estimate of Q_0	223
13.3.4	Estimator of the Target Parameter	223
13.3.5	Calculating Standard Errors	224
13.4	Simulations	224
13.5	Discussion	227
13.6	Notes and Further Reading	227
14	Why Match? Matched Case-Control Studies	229
	Sherri Rose, Mark J. van der Laan	
14.1	Data, Model, and Target Parameter	230
14.2	CCW-TMLE for Individual Matching	230
14.3	Simulations	232
14.4	Discussion	236
14.5	Notes and Further Reading	236
15	Nested Case-Control Risk Score Prediction	239
	Sherri Rose, Bruce Fireman, Mark J. van der Laan	
15.1	Data, Model, and Parameter	240
15.2	Loss Function	240
15.3	Data Analysis	241
15.4	Discussion	243
15.5	Notes and Further Reading	244
Part V RCTs with Survival Outcomes		
16	Super Learning for Right-Censored Data	249
	Eric C. Polley, Mark J. van der Laan	
16.1	Data Structure	250
16.2	Parameters of Interest	252
16.3	Cross-Validation for Censored Data Structures	252
16.4	Super Learner for Hazard Estimation in Lung Cancer	255
16.5	Notes and Further Reading	258
17	RCTs with Time-to-Event Outcomes	259
	Kelly L. Moore, Mark J. van der Laan	
17.1	Data, Likelihood, and Model	260
17.2	Causal Quantity, Identifiability, and Statistical Parameter	261
17.3	Efficient Influence Curve	262
17.4	TMLE of Additive Effect on Survival at a Fixed End Point	263
17.5	Statistical Properties	265
17.6	Variance Estimator	266
17.7	Simulations	266
17.8	Discussion	268
17.9	Notes and Further Reading	268

18 RCTs with Time-to-Event Outcomes and Effect Modification Parameters	271
Ori M. Stitelman, Victor De Gruttola, C. William Wester, Mark J. van der Laan	
18.1 Data Structure	272
18.2 Cox Proportional Hazards and Failure Time Models	274
18.3 Model and Parameters of Interest	277
18.4 Effect Modification Parameters	281
18.5 The TMLE	283
18.6 Data Application: Tshepo Study	288
18.6.1 Causal Effect of NNRTI	289
18.6.2 Causal Effect Modification by Baseline CD4	291
18.6.3 Causal Effect Modification Due to Gender	292
18.7 Discussion	293
Appendix	294

Part VI C-TMLE

19 C-TMLE of an Additive Point Treatment Effect	301
Susan Gruber, Mark J. van der Laan	
19.1 Linear Fluctuation and Squared Error Loss	303
19.1.1 Simulations: Estimator Comparison	307
19.1.2 Simulations: Comparison of C-TMLE and TMLE	313
19.1.3 Data Analysis	315
19.2 Logistic Fluctuation for Bounded Continuous Outcomes	317
19.2.1 Simulations: Logistic vs. Linear Fluctuation	318
19.2.2 Simulations: Estimator Comparison	319
19.3 Discussion	321
20 C-TMLE for Time-to-Event Outcomes	323
Ori M. Stitelman, Mark J. van der Laan	
20.1 Estimating Parameters Based on Coarsened Data	324
20.2 C-TMLEs	325
20.3 Data, Model, and Parameters of Interest	327
20.4 Estimators of the Treatment-Specific Survival Function	327
20.5 C-TMLE of the Treatment-Specific Survival Function	328
20.6 Simulations	334
20.6.1 Results: Point Estimates	335
20.6.2 Results: Inference	339
20.7 Discussion	342
21 Propensity-Score-Based Estimators and C-TMLE	343
Jasjeet S. Sekhon, Susan Gruber, Kristin E. Porter, Mark J. van der Laan	
21.1 Neyman–Rubin Causal Model and Potential Outcomes	347
21.2 Kang and Schafer Censored-Data Simulations	349
21.2.1 Estimators	351

21.2.2	Results	353
21.2.3	Super Learning and the Kang and Schafer Simulations	357
21.3	Freedman and Berk Simulations	358
21.3.1	Estimators	359
21.3.2	Simulations	360
21.3.3	Results	361
21.4	Discussion	364

Part VII Genomics

22	Targeted Methods for Biomarker Discovery	367
	Catherine Tuglus, Mark J. van der Laan	
22.1	Semiparametric-Model-Based Variable Importance	368
22.2	The TMLE	370
22.3	Variable Importance Methods	372
22.4	Simulations	373
22.5	Leukemia Data Application	377
22.6	Discussion	382
23	Finding Quantitative Trait Loci Genes	383
	Hui Wang, Sherri Rose, Mark J. van der Laan	
23.1	Semiparametric Regression Model and TMLE	384
23.2	The C-TMLE	386
23.3	Simulation	389
23.4	Wound-Healing Application	390
23.5	Listeria Application	391
23.6	Discussion	394

Part VIII Longitudinal Data Structures

24	Case Study: Longitudinal HIV Cohort Data	397
	Maya L. Petersen, Mark J. van der Laan	
24.1	Data	398
24.2	Causal Model	399
24.3	Target Causal Parameters	400
24.3.1	Standard Marginal Structural Models	401
24.3.2	History-Adjusted Marginal Structural Models	403
24.3.3	Dynamic Marginal Structural Models	405
24.4	Statistical Model and Identifiability Results	407
24.4.1	Likelihood	408
24.4.2	Target Parameters $\Psi(P_0)$	409
24.5	Estimation	411
24.5.1	MLE	412
24.5.2	Inverse-Probability-Weighted Estimation	413
24.5.3	TMLE	413
24.5.4	IPCW Reduced-Data TMLE	415

24.6	Discussion	416
25	Probability of Success of an In Vitro Fertilization Program	419
	Antoine Chambaz	
25.1	The DAIFI Study	419
25.2	Data, Model, and Parameter	421
25.3	The TMLE	424
	25.3.1 TMLE Procedure	426
	25.3.2 Merits of TMLE Procedure	429
	25.3.3 Implementing TMLE	431
25.4	Simulations	431
25.5	Data Application	433
25.6	Discussion	434
26	Individualized Antiretroviral Initiation Rules	435
	Romain Neugebauer, Michael J. Silverberg, Mark J. van der Laan	
26.1	Longitudinal Data Structure	436
26.2	Likelihood of the Observed Data Structure	440
26.3	Target Parameter	442
26.4	IPCW-R-TMLE	445
	26.4.1 R-TMLE Implementation and Results	446
	26.4.2 IPAW-R-TMLE Implementation and Results	454
26.5	Discussion	456
 Part IX Advanced Topics		
27	Cross-Validated Targeted Minimum-Loss-Based Estimation	459
	Wenjing Zheng, Mark J. van der Laan	
27.1	The CV-TMLE	460
27.2	The CV-TMLE for the Additive Causal Effect	463
27.3	Asymptotics of the CV-TMLE	465
27.4	Discussion of Conditions of the Theorem	473
28	Targeted Bayesian Learning	475
	Iván Díaz Muñoz, Alan E. Hubbard, Mark J. van der Laan	
28.1	Prior, Likelihood, and Posterior Distributions	477
28.2	Convergence of Targeted Posterior Distribution	481
28.3	Frequentist Properties of Targeted Posterior Distribution	482
28.4	Simulations	483
28.5	Discussion	487
	Appendix 1	488
	Appendix 2	491
	Appendix 3	493

29	TMLE in Adaptive Group Sequential Covariate-Adjusted RCTs	495
	Antoine Chambaz, Mark J. van der Laan	
29.1	Statistical Framework	496
29.2	Data-Generating Mechanism	498
29.3	Optimal Design	499
29.4	TMLE Procedure	503
29.4.1	Initial ML-Based Substitution Estimator	503
29.4.2	Convergence of the Adaptive Design	505
29.4.3	The TMLE	506
29.5	Asymptotics	507
29.6	Simulations	512

Part X Appendices

A	Foundations of TMLE	521
A.1	Asymptotic Linearity: The Functional Delta Method	521
A.2	Influence Curve of an Asymptotically Linear Estimator	524
A.3	Computation of the Influence Curve: An Example	526
A.4	Cramer–Rao Lower Bound	529
A.5	Invariance of Statistical Properties	533
A.6	Targeted Minimum-Loss-Based Estimation	537
A.7	Efficient Influence Curve for Longitudinal Data Structures	538
A.8	Factorization in Terms of Binary Conditional Distributions	542
A.9	Efficient Influence Curve Collaborative Double Robustness	545
A.10	Example: TMLE with the Outcome Subject to Missingness	547
A.11	Example: TMLE of Causal Effect in a Two-Stage RCT	548
A.12	Example: TMLE with Right-Censored Survival Time	551
A.13	Example: TMLE of a Causal Effect Among the Treated	552
A.14	Example: TMLE Based on an Instrumental Variable	556
A.15	Example: TMLE of the Conditional Relative Risk	558
A.16	IPCW Reduced-Data TMLE	561
A.17	Collaborative Double Robust TMLE	567
A.18	Asymptotic Linearity of (C-)TMLE	570
A.19	Efficiency Maximization and TMLE	572
A.20	Oracle Inequality of Cross-Validation Selector	577
A.21	Loss-Based Dissimilarity	578
A.22	Examples: Loss-Based Dissimilarity	579
A.23	Example: EEM and TMLE	581
B	Introduction to R Code Implementation	585
	References	589
	Index	613
	Author Index	617

Targeted Learning

Causal Inference for Observational and Experimental
Data

van der Laan, M.J.; Rose, S.

2011, LXXII, 628 p., Hardcover

ISBN: 978-1-4419-9781-4