

Contents

Part I Targeted Learning in Data Science: Introduction

1	Research Questions in Data Science	3
	Sherri Rose and Mark J. van der Laan	
1.1	Learning from (Big) Data	4
1.2	Traditional Approaches to Estimation Fail	5
1.3	Targeted Learning in Practice	6
1.4	The Statistical Estimation Problem	7
1.4.1	Data	8
1.4.2	Model and Parameter	8
1.4.3	Targeted Minimum Loss-Based Estimators	9
1.4.4	Other Common Estimation Problems	10
1.5	Roadmap for Targeted Learning	11
1.6	Notes and Further Reading	13
2	Defining the Model and Parameter	15
	Sherri Rose and Mark J. van der Laan	
2.1	Defining the Structural Causal Model	16
2.2	Causal Graphs	19
2.3	Defining the Causal Target Parameter	21
2.3.1	Interventions	21
2.3.2	Counterfactuals	22
2.3.3	Establishing Identifiability	22
2.3.4	Commit to a Statistical Model and Target Parameter ..	24
2.3.5	Interpretation of Target Parameter	25
2.4	Notes and Further Reading	25
3	Sequential Super Learning	27
	Sherri Rose and Mark J. van der Laan	
3.1	Background: Ensemble Learning	28
3.2	Defining the Estimation Problem	31
3.3	Sequential Super (Machine) Learning	32

3.4	Computation	34
3.5	Notes and Further Reading	34
4	LTMLE	35
	Sherril Rose and Mark J. van der Laan	
4.1	LTMLE in Action: When to Start HIV Treatment	35
4.2	Defining the Estimation Problem	37
4.3	What Does It Mean to Follow a Rule?	37
4.4	LTMLE for When to Start Treatment	40
	4.4.1 Determining the Efficient Influence Curve	40
	4.4.2 Determining the Loss Function and Fluctuation Submodel	41
	4.4.3 LTMLE Algorithm	42
4.5	Analysis of TMLE and Inference	44
	4.5.1 TMLE Solves Efficient Influence Curve Equation	44
	4.5.2 Second-Order Remainder for TMLE	44
	4.5.3 Asymptotic Efficiency	45
	4.5.4 Inference	46
4.6	Notes and Further Reading	47
 Part II Additional Core Topics		
5	One-Step TMLE	51
	Mark J. van der Laan, Wilson Cai, and Susan Gruber	
5.1	Local and Universal Least Favorable Submodels	51
5.2	A Universal Least Favorable Submodel for Targeted <i>Maximum</i> Likelihood Estimation	53
	5.2.1 Analytic Formula	54
	5.2.2 Universal Least Favorable Submodel in Terms of a Local Least Favorable Submodel	55
5.3	Example: One-Step TMLE for the ATT	57
5.4	Universal Least Favorable Model for Targeted <i>Minimum</i> Loss-Based Estimation	60
5.5	Universal Canonical One-dimensional Submodel for a Multidimensional Target Parameter	64
	5.5.1 Practical Construction	65
	5.5.2 Existence of MLE or Approximate MLE ϵ_n	66
	5.5.3 Universal Score-Specific One-Dimensional Submodel	67
5.6	Example: One-Step TMLE, Based on Universal Canonical One-Dimensional Submodel, of an Infinite-Dimensional Target Parameter	68
5.7	Universal Canonical One-Dimensional Submodel for Targeted Minimum Loss-Based Estimation of a Multidimensional Target Parameter	73

6	Highly Adaptive Lasso (HAL)	77
	Mark J. van der Laan and David Benkeser	
6.1	Statistical Formulation of the Estimation Problem	79
6.2	Representation of a Cadlag Function as a Linear Combination of Basis Functions	80
6.3	A Minimum Loss-Based Estimator (MLE) Minimizing over all Functions with Variation Norm Smaller than λ	82
6.4	The HAL Estimator	83
6.5	Further Dimension Reduction Considerations	84
6.6	Applications	85
	6.6.1 Constructing the Highly Adaptive Lasso	86
	6.6.2 Prediction Simulation	88
	6.6.3 Prediction Data Analysis	91
	6.6.4 Simulation for Missing Data	91
	6.6.5 Conclusion	94
7	A Generally Efficient HAL-TMLE	95
	Mark J. van der Laan	
7.1	Treatment Specific Mean	96
	7.1.1 HAL-TMLE	96
	7.1.2 Asymptotic Efficiency	97
7.2	General HAL-TMLE and Asymptotic Efficiency	98
7.3	Discussion	101
8	HAL Estimator of the Efficient Influence Curve	103
	Mark J. van der Laan	
8.1	Formulation of HAL Least Squares Linear Regression Estimator of the Efficient Influence Curve	104
8.2	Rate of Convergence of the HAL Estimator of the Efficient Influence Curve	107
	8.2.1 Application to Estimate Projection of Initial Gradient onto Subtangent Spaces	108
	8.2.2 Using the Actual Data Set from the True Data Distribution	109
8.3	Truncated Mean Based on Current Status Data	109
8.4	Truncated Mean Based on Interval Censored Data	111
8.5	Causal Effect of Binary Treatment on Interval Censored Time to Event	113
8.6	Bivariate Survival Function Based on Bivariate Right-Censored Data	118
8.7	Causal Effect of Binary Treatment on Bivariate Survival Probability Based on Bivariate Right-Censored Data	121
8.8	Discussion	123

9	Data-Adaptive Target Parameters	12
	Alan E. Hubbard, Chris J. Kennedy, and Mark J. van der Laan	
9.1	Example: Defining Treatment or Exposure Levels	12
9.2	Methodology for Data-Adaptive Parameters	12
9.3	TMLE of v -Specific Data-Adaptive Parameter	12
9.4	Combining v -Specific TMLEs Across Estimation Samples	13
9.5	CV-TMLE	13
9.6	CV-TMLE for Data-Adaptive Parameters	13
9.7	CV-TMLE for Variable Importance Measure	13
9.8	Software for Data-Adaptive VIMs: varImpact	13
9.9	Data Analysis: Framingham Heart Study	13
	9.9.1 Super Learner Library	13
	9.9.2 Results	14
9.10	Discussion	14
10	C-TMLE for Continuous Tuning	14
	Mark J. van der Laan, Antoine Chambaz, and Cheng Ju	
10.1	Formal Motivation for Targeted Tuning of Nuisance Parameter Estimator in TMLE	14
	10.1.1 Contrasting Discrete and Continuous Tuning Parameters	14
	10.1.2 Key Theoretical Property and Rational for Proposed C-TMLE That Drives Its Asymptotic Superiority Relative to Standard TMLE	15
	10.1.3 Implicitly Defined Tuning Parameter	15
10.2	A General C-TMLE Algorithm	15
10.3	Verifying That C-TMLE Solves Critical Equation (10.4)	15
	10.3.1 Condition for C-TMLE Solving Critical Equation (10.4)	15
	10.3.2 A TMLE and C-TMLE that Solve Equation (10.3) Exactly	15
10.4	General Theorem for C-TMLE Asymptotic Linearity	15
10.5	Discussion	16

Part III Randomized Trials

11	Targeted Estimation of Cumulative Vaccine Sieve Effects	16
	David Benkeser, Marco Carone, and Peter Gilbert	
11.1	Observed Data	16
11.2	Causal Model and Parameters of Interest	16
11.3	Identification	16
11.4	Efficient Influence Function	16
11.5	Initial Estimates	17
11.6	Submodels and Loss Functions	17
11.7	TMLE Algorithm	17
11.8	Statistical Properties of TMLE	17
11.9	HVTN 505 HIV Vaccine Sieve Analysis	17
11.10	Discussion	17

12	The Sample Average Treatment Effect	175
	Laura B. Balzer, Maya L. Petersen, and Mark J. van der Laan	
12.1	The Causal Model and Causal Parameters	177
12.2	Identifiability	180
12.3	Estimation and Inference	182
	12.3.1 TMLE for the Population Effect	183
	12.3.2 TMLE for the Sample Effect	185
12.4	Extensions to Pair-Matched Trials	187
12.5	Simulation	190
12.6	Discussion	193
13	Data-Adaptive Estimation in Cluster Randomized Trials	195
	Laura B. Balzer, Mark J. van der Laan, and Maya L. Petersen	
13.1	Motivating Example and Causal Parameters	198
13.2	Targeted Estimation in a Randomized Trial Without Matching ..	199
13.3	Targeted Estimation in a Randomized Trial with Matching	203
13.4	Collaborative Estimation of the Exposure Mechanism	206
13.5	Obtaining Inference	208
13.6	Small Sample Simulations	209
	13.6.1 Study 1	209
	13.6.2 Study 2	212
13.7	Discussion	214
Part IV Observational Data		
14	Stochastic Treatment Regimes	219
	Iván Díaz and Mark J. van der Laan	
14.1	Data, Notation, and Parameter of Interest	221
	14.1.1 Identification	223
	14.1.2 Positivity Assumption	224
14.2	Optimality Theory for Stochastic Regimes	224
14.3	Targeted Minimum Loss-Based Estimation	226
	14.3.1 Asymptotic Distribution of TMLE	228
14.4	Initial Estimators	229
	14.4.1 Super Learning for a Conditional Density	229
	14.4.2 Construction of the Library	230
14.5	Notes and Further Reading	232
15	LTMLE with Clustering	233
	Mireille E. Schnitzer, Mark J. van der Laan, Erica E. M. Moodie, and Robert W. Platt	
15.1	The PROBIT Study	234
	15.1.1 Observed Data	234
	15.1.2 Causal Assumptions	236
	15.1.3 Model and Parameter	238
15.2	Two Parametrizations of the g -Formula	239
	15.2.1 g -Computation for the PROBIT	240

	15.2.2	Sequential g -Computation	24
	15.2.3	Sequential g -Computation for the PROBIT	24
	15.2.4	g -Computation Assumptions	24
15.3		LTMLE for a Saturated Marginal Structural Model	24
	15.3.1	Construction of Weights	24
	15.3.2	Efficient Influence Function	24
	15.3.3	LTMLE	24
	15.3.4	LTMLE for the PROBIT	24
15.4		Variance Estimation and Clustering	24
	15.4.1	Distinction Between Clustering and Interference	24
	15.4.2	Estimation with the EIF	24
	15.4.3	Simulation Study	24
15.5		PROBIT Results	24
15.6		Discussion	25
16		Comparative Effectiveness of Adaptive Treatment Strategies	253
		Romain S. Neugebauer, Julie A. Schmittdiel, Patrick J. O'Connor, and Mark J. van der Laan	
	16.1	The Treatment Intensification Study	254
	16.2	Data	256
	16.3	Causal Model and Statistical Estimands	258
	16.4	Estimation	260
		16.4.1 TMLE	261
		16.4.2 Action Mechanism, g_0^g	265
		16.4.3 Outcome Regressions, Q_0^g	268
	16.5	Practical Performance	269
	16.6	Discussion	275
17		Mediation Analysis with Time-Varying Mediators and Exposures	277
		Wenjing Zheng and Mark J. van der Laan	
	17.1	The Mediation Formula, Natural Direct, and Natural Indirect Effects	279
		17.1.1 Counterfactual Outcome Under Conditional Mediator Distribution	280
		17.1.2 Causal Parameters and Identifiability	281
		17.1.3 Longitudinal Mediation Analysis with Marginal vs Conditional Random Interventions	284
	17.2	Efficient Influence Curve	286
	17.3	Estimators	290
		17.3.1 Nontargeted Substitution Estimator	290
		17.3.2 IPW Estimator	292
		17.3.3 TMLE	294
	17.4	Simulation	296
	17.5	Discussion	298

Part V Online Learning

18 Online Super Learning	303
Mark J. van der Laan and David Benkeser	
18.1 Statistical Formulation of Estimation Problem	305
18.1.1 Statistical Model	305
18.1.2 Statistical Target Parameter and Loss Function	305
18.1.3 Regression Example	306
18.2 Cross-Validation for Ordered Sequence of Dependent Experiments	306
18.2.1 Online Cross-Validation Selector	306
18.2.2 Online Oracle Selector	307
18.2.3 The Online Super Learner for a Continuous Finite Dimensional Family of Candidate Estimators	309
18.3 An Oracle Inequality for Online Cross-Validation Selector	310
18.3.1 Quadratic Loss Functions	310
18.3.2 Nonquadratic Loss Functions	311
18.4 Special Online-Cross-Validation Selector for Independent Identically Distributed Observations	312
18.4.1 Online Cross-Validation Selector	312
18.4.2 Imitating V-Fold Cross-Validation	313
18.4.3 Online Oracle Selector	313
18.5 Discussion	315
19 Online Targeted Learning for Time Series	317
Mark J. van der Laan, Antoine Chambaz, and Sam Lendle	
19.1 Statistical Formulation of the Estimation Problem	318
19.1.1 Statistical Model: Stationarity and Markov Assumptions	319
19.1.2 Underlying Causal Model and Target Quantity	320
19.1.3 g-Computation Formula for Post-intervention Distribution	321
19.1.4 Statistical Estimand: Intervention-Specific Counterfactual Mean	321
19.1.5 Sequential Regression Representation of Counterfactual Mean	322
19.1.6 General Class of Target Parameters	322
19.1.7 Statistical Estimation Problem	323
19.2 Efficient Influence Curve of the Target Parameter	324
19.2.1 Monte-Carlo Approximation of the Efficient Influence Curve using the Nesting Assumption	326
19.2.2 A Special Representation of the Efficient Influence Curve for Binary Variables	328

19.3	First Order Expansions for the Target Parameter in Terms of Efficient Influence Curve	31
19.3.1	Expansion for Standard TMLE	31
19.3.2	Expansion for Online One-Step Estimator and Online TMLE	31
19.4	TMLE	31
19.4.1	Local Least Favorable Fluctuation Model	31
19.4.2	One-Step TMLE	31
19.4.3	Iterative TMLE	31
19.4.4	Analysis of the TMLE	31
19.5	Online One-Step Estimator	31
19.6	Online TMLE	31
19.7	Online Targeted Learning with Independent Identically Distributed Data	34
19.7.1	Online Targeted Learning of the Average Causal Effect	34
19.7.2	Online One-Step Estimator	34
19.7.3	Online TMLE	34
19.8	Discussion	34

Part VI Networks

20	Causal Inference in Longitudinal Network-Dependent Data	349
	Oleg Sofrygin and Mark J. van der Laan	
20.1	Modeling Approach	351
20.2	Data Structure	352
20.3	Example	352
20.4	Estimation Problem	353
20.4.1	Counterfactuals and Stochastic Interventions	354
20.4.2	Post-Intervention Distribution and Sequential Randomization Assumption	355
20.4.3	Target Parameter as the Average Causal Effect (ACE)	356
20.4.4	Dimension Reduction and Exchangeability Assumptions	357
20.4.5	Independence Assumptions on Exogenous Errors	357
20.4.6	Identifiability: g -Computation Formula for Stochastic Intervention	358
20.4.7	Likelihood and Statistical Model	359
20.4.8	Statistical Target Parameter	359
20.4.9	Statistical Estimation Problem	360
20.4.10	Summary	360
20.5	Efficient Influence Curve	361
20.6	Maximum Likelihood Estimation, Cross-Validation, and Super Learning	363
20.7	TMLE	365
20.7.1	Local Least Favorable Fluctuation Model	365
20.7.2	Estimation of the Efficient Influence Curve	366

20.8	Summary	368
20.9	Notes and Further Reading	369
21	Single Time Point Interventions in Network-Dependent Data	373
	Oleg Sofrygin, Elizabeth L. Ogburn, and Mark J. van der Laan	
21.1	Modeling Network Data	374
21.1.1	Statistical Model	374
21.1.2	Types of Interventions	376
21.1.3	Target Parameter: Sample-Average of Expected Outcomes	376
21.1.4	Sample Average Mean Direct Effect Under Interference	378
21.2	Estimation	378
21.2.1	The Estimator $\hat{Q}_{W,N}$ for $\bar{Q}_{W,0}$	381
21.2.2	The Initial (Nontargeted) Estimator \hat{Q}_N of \bar{Q}_0	381
21.2.3	Estimating Mixture Densities \bar{g}_0^* and \bar{g}_0	382
21.2.4	The TMLE Algorithm	382
21.3	Inference	383
21.3.1	Inference in a Restricted Model for Baseline Covariates	384
21.3.2	Ad-Hoc Upper Bound on Variance	386
21.3.3	Inference for Conditional Target Parameter	386
21.4	Simulating Network-Dependent Data in R	387
21.4.1	Defining the Data-Generating Distribution for Observed Network Data	387
21.4.2	Defining Intervention, Simulating Counterfactual Data and Evaluating the Target Causal Quantity	390
21.5	Causal Effects with Network-Dependent Data in R	392
21.6	Simulation Results	393
21.7	Notes and Further Reading	395

Part VII Optimal Dynamic Rules

22	Optimal Dynamic Treatment Rules	399
	Alexander R. Luedtke and Mark J. van der Laan	
22.1	Optimal Dynamic Treatment Estimation Problem	400
22.2	Efficient Influence Curve of the Mean Outcome Under V-Optimal Rule	403
22.3	Statistical Inference for the Average of Sample-Split Specific Mean Counterfactual Outcomes Under Data Adaptively Determined Dynamic Treatments	405
22.3.1	General Description of CV-TMLE	406
22.3.2	Statistical Inference for the Data-Adaptive Parameter $\tilde{\psi}_{0n}$	407
22.3.3	Statistical Inference for the True Optimal Rule ψ_0	408
22.4	Discussion	410
22.5	Proofs	411

22.6	CV-TMLE for the Mean Outcome Under Data-Adaptive V-Optimal Rule	414
22.7	Notes and Further Reading	416
23	Optimal Individualized Treatments Under Limited Resources	419
	Alexander R. Luedtke and Mark J. van der Laan	
23.1	Optimal Resource-Constrained Rule and Value	419
23.2	Estimating the Optimal Resource-Constrained Value	422
23.3	Canonical Gradient of the Optimal Resource-Constrained Value	423
23.4	Inference for $\Psi(P_0)$	425
23.5	Discussion of Theorem 23.4 Conditions	426
23.6	Discussion	428
23.7	Proofs	430
24	Targeting a Simple Statistical Bandit Problem	437
	Antoine Chambaz, Wenjing Zheng, and Mark J. van der Laan	
24.1	Sampling Strategy and TMLE	440
	24.1.1 Sampling Strategy	441
	24.1.2 TMLE	442
24.2	Convergence of Sampling Strategy and Asymptotic Normality of TMLE	443
24.3	Confidence Intervals	445
24.4	Simulation	446
24.5	Conclusion (on a Twist)	450
Part VIII Special Topics		
25	CV-TMLE for Nonpathwise Differentiable Target Parameters	455
	Mark J. van der Laan, Aurélien Bibaut, and Alexander R. Luedtke	
25.1	Definition of the Statistical Estimation Problem	456
25.2	Approximating Our Target Parameter by a Family of Pathwise Differentiable Target Parameters	458
25.3	CV-TMLE of h -Specific Target Parameter Approximation	461
	25.3.1 CV-TMLE of $\Psi_h(P_0)$	461
	25.3.2 Asymptotic Normality of CV-TMLE	462
	25.3.3 Asymptotic Normality of CV-TMLE as an Estimator of ψ_0	465
25.4	A Data-Adaptive Selector of the Smoothing Bandwidth	466
25.5	Generalization of Result for Data-Adaptive Bandwidth Selector	470
	25.5.1 Selecting among Different Classes of Pathwise Differentiable Approximations of Target Parameter ...	473
25.6	Example: Estimation of a Univariate Density at a Point	474
25.7	Example: Causal Dose Response Curve Estimation at a Point ...	477
25.8	Notes and Further Reading	480

26 Higher-Order Targeted Loss-Based Estimation	483
Marco Carone, Iván Díaz, and Mark J. van der Laan	
26.1 Overview of Higher-Order TMLE	485
26.1.1 TMLE	485
26.1.2 Extensions of TMLE	487
26.1.3 Second-Order Asymptotic Expansions	488
26.1.4 Construction of a 2-TMLE	489
26.1.5 Insufficiently Differentiable Target Parameters	491
26.2 Inference Using Higher-Order TMLE	492
26.2.1 Asymptotic Linearity and Efficiency	492
26.2.2 Constructing Confidence Intervals	493
26.2.3 Implementing a Higher-Order TMLE	494
26.3 Inference Using Approximate Second-Order Gradients	495
26.3.1 Asymptotic Linearity and Efficiency	496
26.3.2 Implementation and Selection of Tuning Parameter ...	496
26.4 Illustration: Estimation of a g -Computation Parameter	498
26.4.1 Case I: Finite Support	499
26.4.2 Case II: Infinite Support	501
26.4.3 Numerical Results	504
26.5 Concluding Remarks	507
26.6 Notes and Further Reading	509
27 Sensitivity Analysis	511
Iván Díaz, Alexander R. Luedtke, and Mark J. van der Laan	
27.1 The Problem	512
27.2 Sensitivity Analysis	514
27.3 Bounds on the Causal Bias Are Unknown	519
27.4 Notes and Further Reading	521
28 Targeted Bootstrap	523
Jeremy Coyle and Mark J. van der Laan	
28.1 Problem Statement	524
28.2 TMLE	525
28.2.1 TMLE for Treatment Specific Mean	525
28.2.2 TMLE of the Variance of the Influence Curve	526
28.2.3 Joint TMLE of Both the Target Parameter and Its Asymptotic Variance	528
28.3 Super Learner	528
28.4 Bootstrap	530
28.4.1 Nonparametric Bootstrap	530
28.4.2 Model-Based Bootstrap	531
28.4.3 Targeted Bootstrap	531
28.4.4 Bootstrap Confidence Intervals	532
28.5 Simulation	534
28.6 Conclusion	539

29 Targeted Learning Using Adaptive Survey Sampling	541
Antoine Chambaz, Emilien Joly, and Xavier Mary	
29.1 Template for Targeted Inference by Survey Sampling	542
29.1.1 Retrieving the Observations by Survey Sampling	542
29.1.2 CLT on the TMLE and Resulting Confidence Intervals	543
29.2 Survey Sampling Designs and Assumption A1	545
29.2.1 Sampford's Survey Sampling Design	545
29.2.2 Determinantal Survey Sampling Design	546
29.3 Optimizing the Survey Sampling Designs	549
29.4 Example: Variable Importance of a Continuous Exposure	549
29.4.1 Preliminaries	550
29.4.2 Construction of the TMLE	551
29.5 Simulation	553
29.6 Elements of Proof	555
29.6.1 Proof of Proposition 29.1	555
29.6.2 Proof of Eqs. (29.8) and (29.9)	556
29.6.3 Proof of Proposition 29.3	557
29.6.4 Proof of Proposition 29.4	558
30 The Predicament of Truth: On Statistics, Causality, Physics, and the Philosophy of Science	561
Richard J. C. M. Starmans	
30.1 Statistics and the Fragility of Truth	561
30.2 Truth in Epistemology and Methodology	563
30.3 Eroded Models, Von Neumann and the End of Theory	566
30.4 Physics, Statistics and the Philosophy of Science	569
30.5 The Triptych of True Knowledge	571
30.6 Some Roots and Aspects of Causality	574
30.7 Elimination, Dualism, the Probabilistic Revolution, and Unification	578
30.8 Conclusion	582
A Appendix: Foundations	585
A.1 Data-Adaptive Target Parameters	585
A.1.1 Statistical Inference Based on the CV-TMLE	585
A.2 Mediation Analysis	588
A.2.1 Proof of Lemma 17.1: Identifiability Result	589
A.2.2 Proof of Theorem 17.1	590
A.3 Online Super Learning	593
A.3.1 Online Cross-Validated Risk Minus Online Cross-Validated True Risk Is a Discrete Martingale ...	593
A.3.2 Martingale Exponential Inequality for Tail Probability	593
A.3.3 Proof of Theorem 18.1	595
A.3.4 Brief Review of Literature on Online Estimation	601

Contents	xxxiii
A.4 Online Targeted Learning	602
A.4.1 First Order Expansion of Target Parameter Based on Marginal Expectation of Efficient Influence Curve	603
A.4.2 First Order Expansion of Target Parameter Based on Conditional Expectations of Efficient Influence Curve Components	606
A.4.3 Discussion of $R_{22,g^*,N}$ Remainder: Finite Memory Assumption	608
A.4.4 First Order Expansion for Online Estimation Based on Marginal Expectation of Efficient Influence Curve	610
A.4.5 First Order Expansion for Online Estimation Based on Conditional Expectation of Efficient Influence Curve	612
References	613