Targeted Learning: Causal Inference for Observational and Experimental Data

Mark J. van der Laan[1], Maya L. Petersen[1], Sherri Rose[2]

[1]University of California, Berkeley School of Public Health
[2]Johns Hopkins Bloomberg School of Public Health

laan@berkeley.edu · mayaliv@berkeley.edu · srose@jhsph.edu
stat.berkeley.edu/~laan/
works.bepress.com/maya_petersen/
drsherrirose.com

targetedlearningbook.com

July 29, 2012

Course Goals

1. Understand the shortcomings of traditional parametric regression-based techniques for the estimation of causal effects.

2. Translate a scientific question and background knowledge into a formal causal model and target causal parameter.

3. Become familiar with the properties and basic implementation of three distinct estimators of the causal effect of a point treatment: Maximum Likelihood Substitution Estimator, Inverse Probability Weighted Estimator, and Targeted Maximum Likelihood Estimators.

4. Be introduced to the ideas behind machine learning approaches as tools for confronting the curse of dimensionality.

5. Understand the challenges raised by estimation of the causal effect of longitudinal exposures.

6. Obtain an overview of estimating the causal effect of a longitudinal exposure in the context of right censoring and time-dependent confounding.

LECTURE ONE

Motivation for Targeted Causal Inference Methods

Tradition Approach in Epidemiology and Clinical Medicine

1. Fit several parametric logistic regression models, and select a favorite one.

2. Report point estimate of coefficient in front of treatment, confidence intervals, and *p*-value, as if this parametric model was a priori-specified.

Complications of Human Art in Statistics

1. The parametric model is misspecified.
2. The target parameter is interpreted as if the parametric model is correct.
3. The parametric model is often data-adaptively (or worse!) deleted, and this part of the estimation procedure is not accounted for in the variance.

Complications of Human Art in Statistics

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

The New York Times
nytimes.com

September 16, 2007

Do We Really Know What Makes Us Healthy?

By GARY TAUBES

AMSTATNEWS
The Membership Magazine of the American Statistical Association

Statistics Ready for a Revolution

1 SEPTEMBER 2010   503 VIEWS   2 COMMENTS

Next Generation of Statisticians Must Build Tools for Massive Data Sets

Mark van der Laan, Jiann-Ping Hsu/Karl E. Peace Professor in Biostatistics and Statistics at UC Berkeley, and Sherri Rose, PhD candidate at UC Berkeley

Complications of Human Art in Statistics

Debate over HRT

Professional groups gave HRT their stamp of approval 15 years ago.

Studies indicated HRT protective against osteoporosis and heart disease.

In 1998, a study demonstrated increased risk of heart attack among women with heart disease taking HRT.

In 2002 a study showed increased risk for breast cancer, heart disease, and stroke, among other ailments, for women on HRT.

Why were there inconsistencies in the study results?

Complications of Human Art in Statistics

Debate over mammography

Mammography gained widespread acceptance as effective tool for breast cancer screening in the 1980s.

The Health Insurance Plan trial and Swedish Two-County trial demonstrated mammography saved lives.

In 2009, surprise over new recommendations from the U.S. Preventive Services Task Force.

Among women without a family history, mammography now recommended for women aged 50 to 74. Previous guidelines started at age 40.

Why was there a seemingly sudden paradigm shift?

Estimation is a Science, *Not an Art*

1. **Data**: realizations of random variables with a probability distribution.
2. **Statistical Model**: actual knowledge about the shape of the data-generating probability distribution.
3. **Statistical Target Parameter**: a feature/function of the data-generating probability distribution.
4. **Estimator**: an a priori-specified algorithm, benchmarked by a dissimilarity-measure (e.g., MSE) w.r.t. target parameter.

Modeling beyond the Statistical Model: Causal Inference

Under additional non-testable assumptions one may describe the probability distribution of the data in terms of an underlying parameter varying over an underlying parameter space. For example, the underlying parameter may be the probability distribution of a full-data random variable such as a collection of intervention specific counterfactuals, and the parameter space might be described by a full-data model (e.g., a nonparametric structural equation model).

The model is the statistical model enriched with this parameterization (i.e. non-testable assumptions).

One may then define the target quantity as the function of the underlying parameter. We may call this the full-data parameter, or causal quantity if it concerns a parameter in a causal model.

One may then establish identifiability of this target quantity from the probability distribution of the data (i.e., $P_0 = P_{\theta_0}$ may identify our target quantity $\Psi^F(\theta_0)$), possibly even under further restrictions on the underlying parameter. This presents us with a statistical target parameter.

The statistical target parameter can always be interpreted pure statistically within the statistical model, and under the stated non-testable assumptions, it can also be interpreted as the underlying target quantity.

The statistical estimation problem only cares about the statistical model, and statistical target parameter, while the non-testable assumptions enrich the interpretation of the statistical target parameter.

The Need for Targeted Learning in Semiparametric Models

1. MLE/machine learning are not targeted for effect parameters.
2. For that, we need a subsequent targeted bias-reduction step.

Targeted MLE

Targeted Learning

- Avoid reliance on human art and nonrealistic (parametric) models
- Define interesting parameters
- Target the fit of data-generating distribution to the parameter of interest
- Statistical inference

TMLE/SL Toolbox

Targeted effects

- Effect of static or dynamic treatments (e.g. on survival time)
- Direct and Indirect effects
- Parameters of Marginal Structural Models
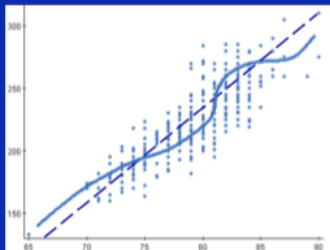- Variable importance analysis in genomics

Types of data

- Point treatment
- Longitudinal/Repeated Measures
- Censoring/Missingness/Time-dependent confounding
- Case-Control
- Randomized clinical trials and observational data

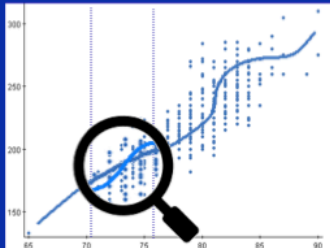# Two-stage Methodology: SL/TMLE

1. Super Learning

- Uses a library of estimators
- Builds data-adaptive weighted combination of estimators
- Weights are optimized based on loss-function specific cross-validation to guarantee best overall fit

2. Targeted Maximum Likelihood Estimation

- Zooms in on one aspect of the estimator—the target feature
- Removes bias for the target.

Targeted Maximum Likelihood Learning

Super Learner (van der Laan, Polley, and Hubbard; 2007)

Allows researchers to use multiple algorithms to outperform a single algorithm in realistic non-parametric and semi-parametric statistical models that are based on actual knowledge. It generalizes stacking, and provides its theoretical underpinning.

TMLE (van der Laan and Rubin; 2006)

Produces a well-defined, unbiased, efficient substitution estimator of target parameter of the data-generating distribution.

It updates an initial estimator by iteratively fitting an amount of fluctuation along a least favorable parametric submodel through the initial estimator.

TMLE for Causal Effects Based on Censored Data

TMLE: Double Robust

- Removes asymptotic residual bias of initial estimator for the target parameter, if it uses a consistent estimator of the treatment and censoring mechanism.

- If initial estimator was consistent for the target parameter, the additional fitting of the data in the targeting step may remove finite sample bias, and preserves consistency property of the initial estimator.
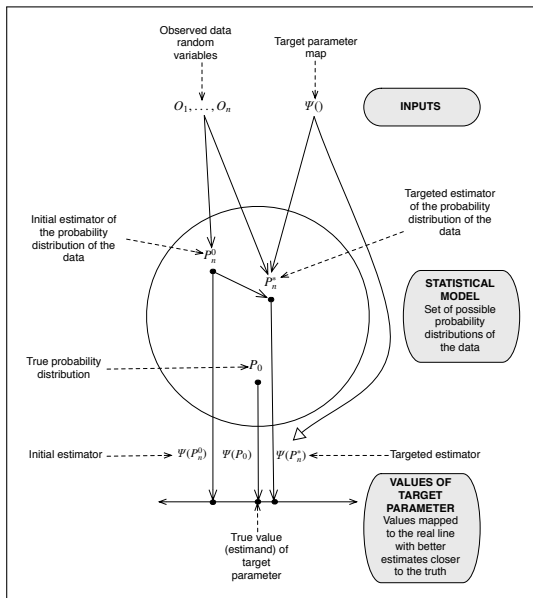
TMLE: Efficiency

- If the initial estimator and the estimator of treatment and censoring mechanism are both consistent, then it is also asymptotically efficient according to semi-parametric statistical model efficiency theory.

TMLE

TMLE: In Practice

Allows the incorporation of machine learning methods for the estimation of relevant part of data generating distribution and the treatment/censoring mechanism, so that we do not make enforce restrictions that are not dictated by the actual knowledge/statistical model.

Thus, every effort is made to achieve minimal bias and the asymptotic semi-parametric efficiency bound for the variance.

TMLE Algorithm

Targeted MLE

1. Identify the parametric model for fluctuating initial $\hat{P}$
   – Small "fluctuation" $\rightarrow$ maximum change in **target**.
2. Given strategy, identify optimum amount of fluctuation by MLE.
3. Apply optimal fluctuation to $\hat{P}$ $\rightarrow$ **1st-step targeted maximum likelihood estimator**.
4. Repeat until the incremental "fluctuation" is zero
   – Some important cases: 1 step to convergence.
5. Final probability distribution solves efficient influence curve equation

$\rightarrow$ **T-MLE is double robust & locally efficient**

Targeted Minimum-Loss-Based Estimation (TMLE)

$\Psi(Q_0)$ target parameter

$Q_0 = \arg\min_Q P_0 L(Q) \equiv \int L(Q)(o) dP_0(o)$

$\hat{Q}(P_n)$: Initial estimator, Loss-based SL

$\{\hat{Q}_g(\epsilon) : \epsilon\}$ fluct. model for fitting ψ_0

$\hat{g} = \hat{g}(P_n)$ loss based SL of treatment/cens mech

$\frac{d}{d\epsilon} L(\hat{Q}_{\hat{g}}(\epsilon))\big|_{\epsilon=0} = D^*(\hat{Q}, \hat{g})$

$\epsilon_n = \arg\min_\epsilon P_n L(\hat{Q}_{\hat{g}}(\epsilon))$

Iterate till convergence: $\hat{Q}^*$

Solves efficient influence curve equation:

$P_n D^*(\hat{Q}^*, \hat{g}) = 0$
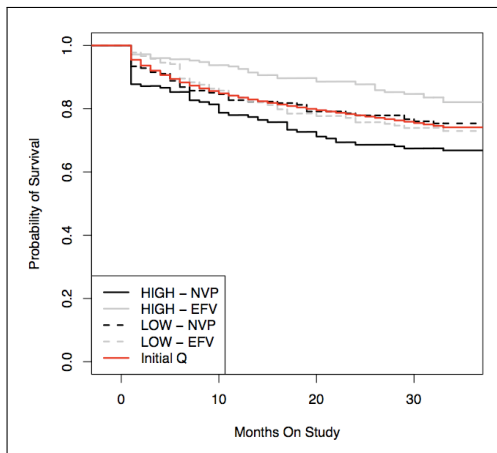
TMLE: $\Psi(\hat{Q}^*)$

Example: Tshepo HIV Study

- Clinical trial evaluating the long-term efficacy of Efavirenz (EFV)- vs. Nevirapine (NVP)-based cART among adults in Botswana.
- 650 individuals randomized to either EFV or NVP.
- Outcomes: Time-to-death, virologic failure, and time to loss of virological response.
- Interest in additive treatment effect modification by low/high baseline CD4 level and gender.

See Chapter 18 of Targeted Learning Book as well as Targeted Maximum Likelihood Estimation of Effect Modification Parameters in Survival Analysis, Int J Biostat (2011), OM Stitelman, CW Wester, V De Gruttola, MJ van der Laan.
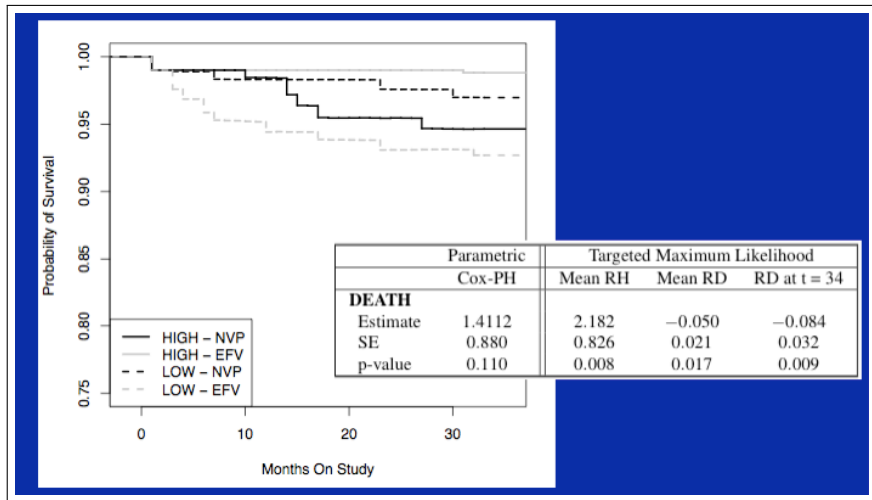
Example: Tshepo HIV Study

- Initial hazard is purposely misspecified; it does not include treatment and CD4 level.
- TMLE-hazard still identifies that high-CD4 has a strong treatment effect on survival, while low-CD4 has a small treatment effect.

Example: Tshepo HIV Study

Causal effect modification by CD4 level and death.

	Parametric	Targeted Maximum Likelihood		
	Cox-PH	Mean RH	Mean RD	RD at t = 34
DEATH				
Estimate	1.4112	2.182	-0.050	-0.084
SE	0.880	0.826	0.021	0.032
p-value	0.110	0.008	0.017	0.009

Legend:
- HIGH – NVP
- HIGH – EFV
- LOW – NVP
- LOW – EFV

Probability of Survival

Months On Study

FDA RCT Study

- International, multicenter, double-blind, parallel, placebo controlled RCT, which aims to evaluate safety based on mortality because of drug-to-drug interaction. We randomized the patients to receive either Drug1 or a placebo. All patients received Drug2 concomitantly as a background therapy.

- The primary objective was to determine whether the mortality rates between patients receiving Drug1 and placebo remained within a 1% margin or less.

- Postmarket data analysis had suggested a harmful interaction.

Robust extraction of covariate information to improve estimation efficiency in randomized trials, Stat Med (2011), KL Moore, R Neugebauer, T Valappil, MJ van der Laan.

Casual Effect of Warfarin (Kaiser Permanente DOR)

- Studying the causal effect of Warfarin on time-to-stroke, subject to right-censoring informed by time-dependent covariates.

Sentinel Project

- FDA initiative.
- Building a database with drug use and side effects/safety events.
- Requires automated unbiased estimation of effects.

Abacovir Observational Study

- Population: HIV-infected patients on antiretroviral treatment.
- Abacavir is a specific antiretroviral drug used to treat HIV.
- Analysis of observational data from several cohorts suggested abacavir use is associated with increased risk of myocardial infarction.
- Other analyses found no evidence of such an association.

Summary

Traditional approaches for prediction/density estimation and effect estimation are biased

- **Super Learning** allows researchers to combine multiple estimators of an infinite dimensional parameter into an improved estimator.
- **Targeted MLE** provides bias reduction for efficient plug-in effect estimation of the target parameter

Landscape of Estimators of the Parameter of Interest

- **G-Computation/MLE**: Graph-Computation/Maximum Likelihood Estimation

- **Estimating-Equation-Based Methods**: Inverse-Probability-of-Treatment Weighted (IPTW) and A-IPTW

- **TMLE**: Targeted Maximum Likelihood Estimation

Course Companion Text

www.targetedlearningbook.com