Ensemble Learning Targeted Maximum Likelihood Estimation for Causal Inference

Miguel Angel Luque-Fernandez, BSc, MA, MPH, MSc, GradStat(UK), PhD

Associate Professor of Statistics
Department of Statistics and Operations Research
Faculty of Science
University of Granada, Granada, Spain

https://maluque.netlify.app/

IMAG: Mathematical Institute of Granada

April 10, 2023

Table of Contents

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Notation and definitions

Observed Data

- Treatment **A**.

  Often, A = 1 for treated and A = 0 for control.
- Confounders **W**.
- Outcome **Y**.

Potential Outcomes

- For patient i $\mathbf{Y_i(1)}$ and $\mathbf{Y_i(0)}$ set to A = a $Y^{(a)}$, namely A = 1 and A = 0.

Causal Effects

- Average Treatment Effect: **E[Y(1) - Y(0)]**.

Background: Potential Outcomes framework

Rubin and Heckman

- This framework was developed first by statisticians (Rubin, 1983) and econometricians (Heckman, 1978) as a new approach for the estimation of **causal effects** from observational data.

- We will keep separate the **causal framework** (a conceptual issue briefly introduce here) and the **"how to estimate causal effects"** (an statistical issue also introduced here)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Causal effects with OBSERVATIONAL data

ASSUMPTIONS for Identification

- Rosebaum & Rubin, 1983: **The Ignorable Treatment Assignment** (A.K.A Ignorability, Unconfoundeness or Conditional Mean Independence).
- **POSITIVITY**.
- **SUTVA**.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Causal effect with OBSERVATIONAL data

IGNORABILITY

$$(\mathbf{Y_i(1)}, \mathbf{Y_i(0)}) \perp \mathbf{A_i} \mid \mathbf{W_i}$$

POSITIVITY

POSITIVITY: $\mathbf{P(A = a \mid W) > 0}$ for all a, W

SUTVA

- We have assumed that there is only on version of the treatment (**consistency**) **Y(1)** if **A = 1** and **Y(0)** if **A = 0**.
- The assignment to the treatment to one unit doesn't affect the outcome of another unit (**no interference**) or **IID** random variables.
- The model used to estimate the assignment probability has to **be correctly specified**.

Causal effect

Potential Outcomes

We only observe:

$$Y_i(1) = Y_i(A = 1) \text{ and } Y_i(0) = Y_i(A = 0)$$

However we would like to know what would have happened if:

Treated $Y_i(1)$ would have been non-treated $Y_i(A = 0) = Y_i(0)$.

Controls $Y_i(0)$ would have been treated $\mathbf{Y_i(A = 1) = Y_i(1)}$.

Identifiability

- How we can identify the effect of the potential outcomes $\mathbf{Y}^a$ if they are not observed?
- How we can estimate the expected difference between the potential outcomes **E[Y(1) - Y(0)]**, namely the **ATE** or risk difference.

G-Formula, (Robins, 1986)

G-Formula for the **identification** of the ATE (Estimand) with observational data

$$
\begin{aligned}
E(Y^a) &= \sum_y E(Y^a \mid W = w) P(W = w) \\
&= \sum_y E(Y^a \mid A = a, W = w) P(W = w) \text{ by consistency} \\
&= \sum_y E(Y = y \mid A = a, W = w) P(W = w) \text{ by ignorability}
\end{aligned}
$$

The **ATE**=

$$
\sum_{\mathbf{w}} \left[\sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 1, W = w)} - \sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 0, W = w)} \right] \mathbf{P(W = w)}
$$

$$
P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)
$$

G-Formula, (Robins, 1986)

G-Formula for the identification of the ATE (Estimand) with observational data

The **ATE**=

$$\sum_{\mathbf{w}} \left[\sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 1, W = w)} - \sum_{\mathbf{y}} \mathbf{P(Y = y \mid A = 0, W = w)} \right] \mathbf{P(W = w)}$$

$$P(W = w) = \sum_{y,a} P(W = w, A = a, Y = y)$$

G-Formula

- The sums is generic notation. In reality, likely involves sums and integrals (we are just integrating out the W's).
- The **g-formula** is a **generalization of standardization** and allow to estimate unbiased treatment effect estimates.

HYGIENE &TROPICAL MEDICINE

ATE estimators

Nonparametric

- G-formula plug-in estimator (generalization of standardization).

Parametric

- Regression adjustment (RA).
- Inverse probability treatment weighting (IPTW).
- Inverse-probability treatment weighting with regression adjustment (IPTW-RA) (Kang and Schafer, 2007).

Semi-parametric Double robust (DR) methods

- Augmented inverse-probability treatment weighting (Estimation Equations) (AIPTW) (Robins, 1994).
- Targeted maximum likelihood estimation (TMLE) **(van der Laan, 2006)**.

&TROPICAL MEDICINE

ESTIMATORS: G-Computation

G-Computation: Regression adjustment (RA)

$$\text{ATE} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(E(Y_i \mid A_i = 1, \boldsymbol{W}_i) - E(Y_i \mid A_i = 0, \boldsymbol{W}_i)\right)}_{G-computation-Regression-Adjustment}$$

ESTIMATORS: IPTW

IPTW (Inverse probability treatment weighting)

Survey theory (Horvitz-Thompson)

$$\text{ATE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i}{P(A_i = 1 \mid \boldsymbol{W}_i)} - \frac{1 - A_i}{(1 - P(A_i = 1 \mid \boldsymbol{W}_i))}\right) Y_i.$$

IPTW standardized weights

$$\text{ATE} = \frac{\sum\left(\frac{AY}{P(A=1|\boldsymbol{W})}\right)}{\sum\left(\frac{A}{P(A=1|\boldsymbol{W})}\right)} - \frac{\sum\left(\frac{(1-A)Y}{1-P(A=1|\boldsymbol{W})}\right)}{\sum\left(\frac{(1-A)}{1-P(A=1|\boldsymbol{W})}\right)}.$$

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

ESTIMATORS: Double Robust type AIPTW

AIPTW

$$\text{ATE} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(E(Y_i \mid A_i = 1, \boldsymbol{W}_i) - E(Y_i \mid A_i = 0, \boldsymbol{W}_i)\right)}_{G-computation-Regression-Adjustment}$$

$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i[Y_i - E(Y_i \mid A_i = 1, \boldsymbol{W_i})]}{g(A_i = 1 \mid \boldsymbol{W}_i)} - \frac{(1 - A_i)[Y_i - E(Y_i \mid A_i = 1, \boldsymbol{W_i})]}{g(A_i = 0 \mid \boldsymbol{W}_i)}\right)}_{Zero-expectation},$$

Equivalence between IPTW and G-computation

Equivalence

By repeated use of the law of total expectation, the IPTW and the G-computation regression adjustment estimators for the ATE are equivalent as given by

$$\underbrace{E\left(\frac{I(a=1)}{P(A=1 \mid \boldsymbol{W})} Y\right)}_{IPTW} =$$

By definition of expectations...

$$= \sum_{w,a,y} \frac{I(a=1)}{P(A=1|W=w)} y\, P(Y=y, A=a, W=w)$$

By the law of total probability...

$$= \sum_{w,a,y} \frac{I(a=1)}{P(A=1 \mid W=w)} y\, P(Y=y \mid A=a, W=w)\, P(A=a \mid W=w)\, P(W=w)$$

Equivalence between IPTW and G-computation

Equivalence

By repeated use of the law of total expectation, the IPTW and the G-computation regression adjustment estimators for the ATE are equivalent as given by

Cancellation by evaluating at A=1...

$$= \sum_{w,y} y\, P(Y = y \mid A = 1, W = w)\, P(w = w)$$

By definition of expectations...

$$= \sum_{w} E(Y \mid A = 1, W = w)\, P(W = w)$$

Finally, again by definition of expectations...

$$= \underbrace{E[E(Y \mid A = 1, W)]}_{G-computation}$$

ATE estimators: drawbacks

Nonparametric

- Course of dimensionality (sparsity: zero empty cell)

Parametric

- Parametric models are misspecified (all models are wrong but some are useful, Box, 1976), and break down for high-dimensional data.
- (RA) Issue: extrapolation and biased if misspecification, no information about treatment mechanism.
- (IPTW) Issue: sensitive to course of dimensionality, inefficient in case of extreme weights and biased if misspecification. Non information about the outcome.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Double-robust (DR) estimators

Pros: Semi-parametric Double-Robust Methods

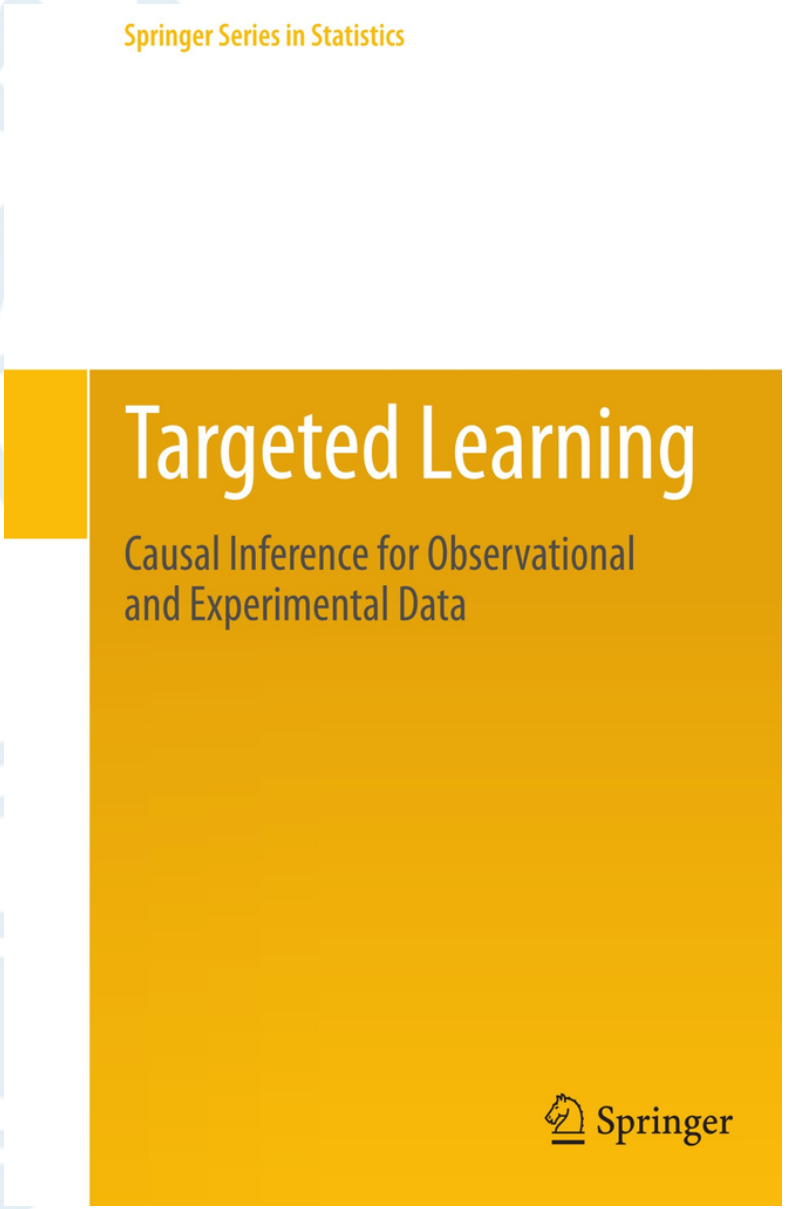
- DR methods give **two chances at consistency** if any of two nuisance parameters is consistently estimated.
- DR methods are **less sensitive** to course of dimensionality.

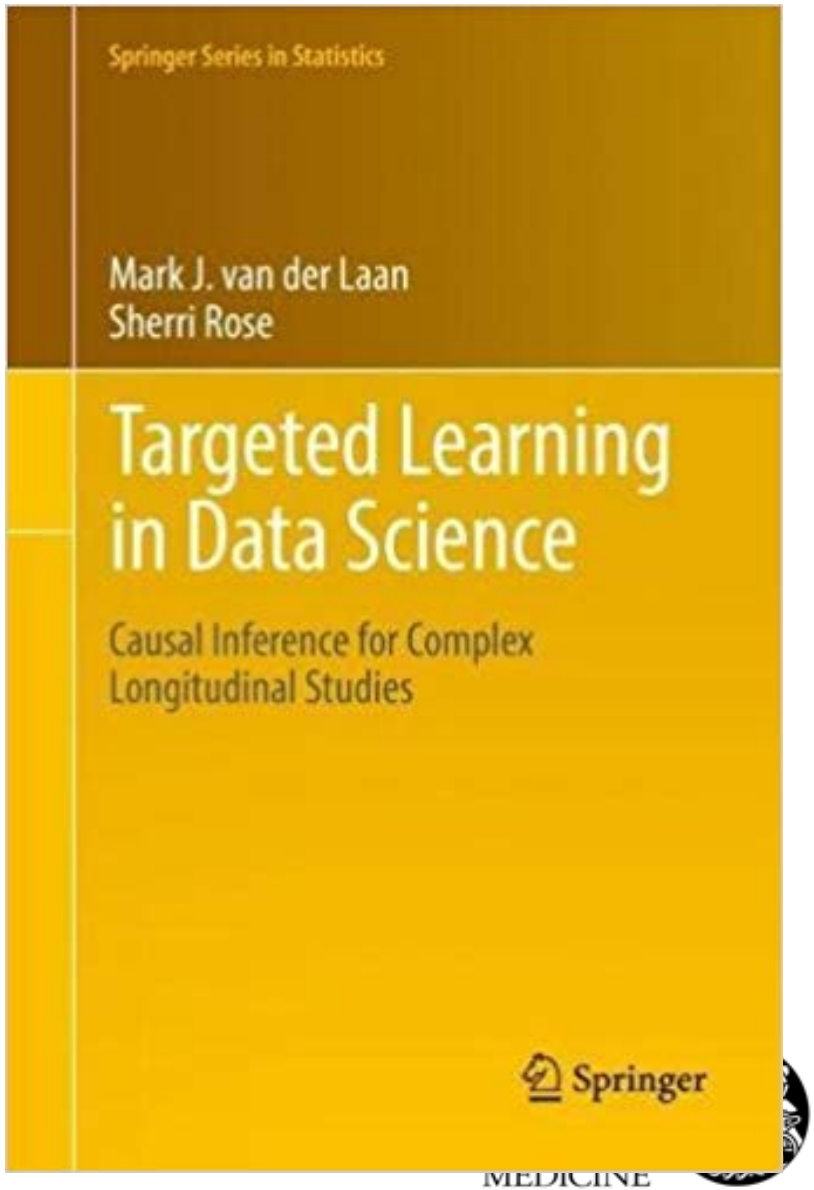
Cons: Semi-parametric Double-Robust Methods

- DR methods are unstable and inefficient if the propensity score (PS) is small (**violation of positivity assumption**) (vand der Laan, 2007).
- AIPTW and IPTW-RA do not respect the **limits of the boundary space of Y**.
- **Poor performance** if **dual misspecification** (Benkeser, 2016).

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Targeted learning





Targeted Maximum Likelihood Estimation (TMLE)

Pros: TMLE

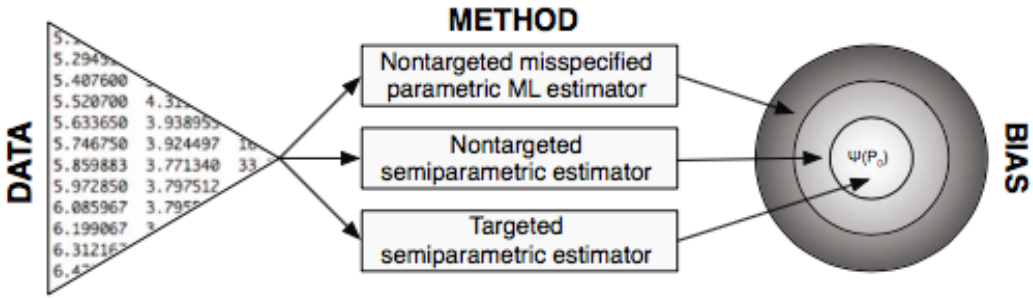
- (TMLE) is a general algorithm for the construction of **double-robust**, **semiparametric** MLE, efficient **substitution** estimator (Van der Laan, 2011)
- Better performance than competitors has been largely documented (Porter, et. al.,2011).
- (TMLE) **Respect bounds on Y**, **less sensitive** to **misspecification** and to **near-positivity** violations (Benkeser, 2016).
- (TMLE) **Reduces bias** through **ensemble learning** if misspecification, even dual misspecification.
- For the ATE, **Inference** is based on the **Efficient Influence Curve**. Hence, the **CLT** applies, making inference easier.

Cons: TMLE

- The procedure is only available in R: **tmle** package (Gruber, 2011).

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Why Targeted learning?



Source: Mark van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer Series in Statistics, 2011.

Statistics: The two cultures

Home >

Open Access

Seleccionar idioma

Translator Disclaimer

August 2001

Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)

Leo Breiman

Statist. Sci. 16(3): 199-231 (August 2001). DOI: 10.1214/ss/1009213726

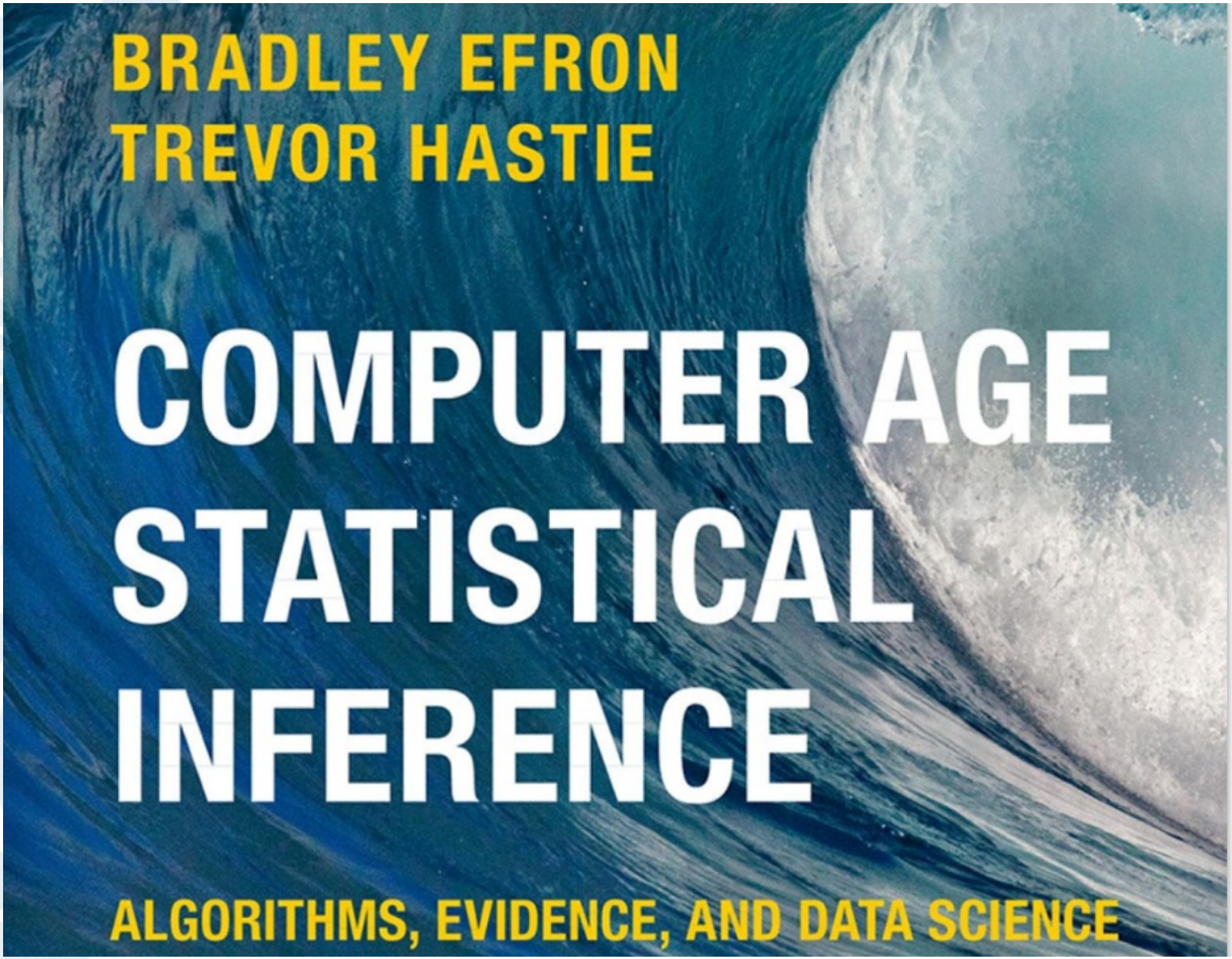
ABOUT | FIRST PAGE | CITED BY | REFERENCES

Abstract

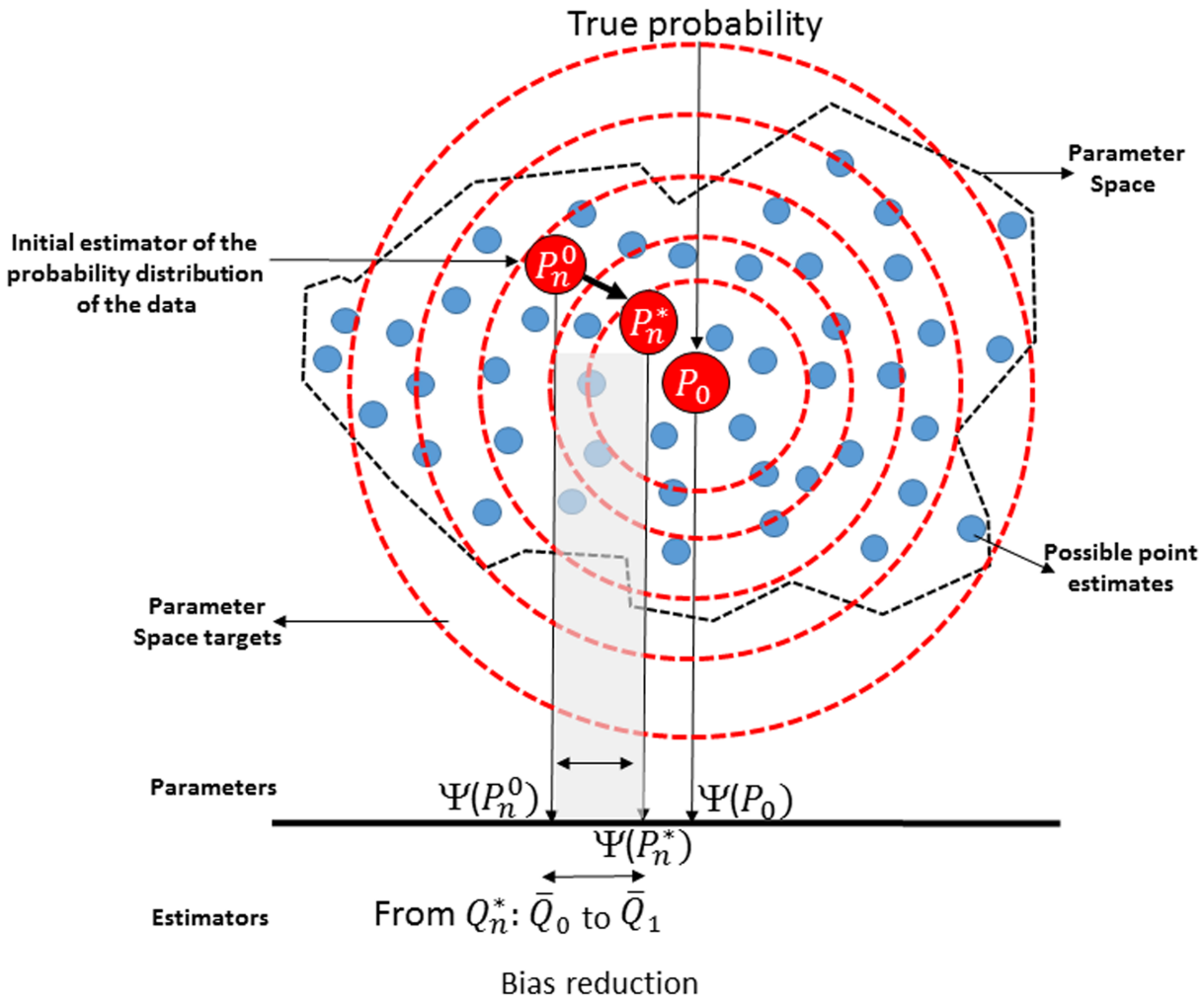
There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Statistics: The two cultures



TMLE ROAD MAP



TMLE ROAD MAP

MC simulations: Luque-Fernandez et al, 2017 (in press, American Journal of Epidemiology)

	ATE		BIAS (%)		RMSE		95%CI coverage (%)	
	N=1,000	N=10,000	N=1,000	N=10,000	N=1,000	N=10,000	N=1,000	N=10,000
First scenario* (correctly specified models)								
True ATE	**-0.1813**							
Naïve	-0.2234	-0.2218	23.2	22.3	0.0575	0.0423	77	89
AIPTW	-0.1843	-0.1848	1.6	1.9	0.0534	0.0180	93	94
IPTW-RA	-0.1831	-0.1838	1.0	1.4	0.0500	0.0174	91	95
TMLE	-0.1832	-0.1821	1.0	**0.4**	0.0482	0.0158	**95**	**95**
Second scenario ** (misspecified models)								
True ATE	**-0.1172**							
Naïve	-0.0127	-0.0121	89.2	89.7	0.1470	0.1100	0	0
BFit AIPTW	-0.1155	-0.0920	1.5	**11.7**	0.0928	0.0773	65	65
BFit IPTW-RA	-0.1268	-0.1192	8.2	**1.7**	0.0442	0.0305	52	73
TMLE	-0.1181	-0.1177	**0.8**	**0.4**	0.0281	0.0107	**93**	**95**

*First scenario : correctly specified models and near-positivity violation

**Second scenario: misspecification, near-positivity violation and adaptive model selection

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

TMLE STEPS

Substitution estimation: $\hat{E}(Y \mid A, W)$

- First compute the outcome regression $\mathbf{E}(\mathbf{Y} \mid \mathbf{A}, \mathbf{W})$ using the **Super-Learner** to then derive the Potential Outcomes and compute $\psi^{(0)} = \mathbf{E}(Y(1) \mid A = 1, W) - \mathbf{E}(Y(0) \mid A = 0, W)$.

- Estimate the exposure mechanism P(A=1|,W) using the **Super-Learner** to predict the values of the propensity score.

- Compute **HAW** = $\left(\frac{\mathbb{I}(A_i=1)}{P(A_i=1|W_i)} - \frac{\mathbb{I}(A_i=0)}{P(A_i=0|W_i)} \right)$ for each individual, named the **clever covariate H**.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Fluctuation step: Epsilon

Fluctuation step ($\hat{\epsilon}_0$, $\hat{\epsilon}_1$)

- Update $\Psi^{(0)}$ through a fluctuation step incorporating the information from the exposure mechanism:

  $$\mathbf{H(1)W} = \frac{\mathbb{I}(A_i=1)}{\hat{P}(A_i=1|W_i)} \text{ and,} \mathbf{H(0)W} = -\frac{\mathbb{I}(A_i=0)}{\hat{P}(A_i=0|W_i)}.$$

- This step aims to **reduce bias** minimising the mean squared error (MSE) for (Ψ) and considering the **bounds of the limits of Y**.
- The fluctuation parameters ($\hat{\epsilon}_0$, $\hat{\epsilon}_1$) are estimated using maximum likelihood procedures (in Stata):

  . glm Y HAW, fam(binomial) nocons offset(E(Y| A, W))

  . mat e = e(b),

  . gen double $\epsilon = e[1,1]$,

HYGIENE &TROPICAL MEDICINE

Targeted estimate of the ATE ($\widehat{\Psi}$)

$\Psi^{(0)}$ update using ϵ (epsilon)

$$\mathbf{E}^*(Y \mid A = 1, W) = \text{expit}\left[\textbf{logit}\left[E(Y \mid A = 1, W)\right] + \hat{\epsilon_1} H_1(1, W)\right]$$

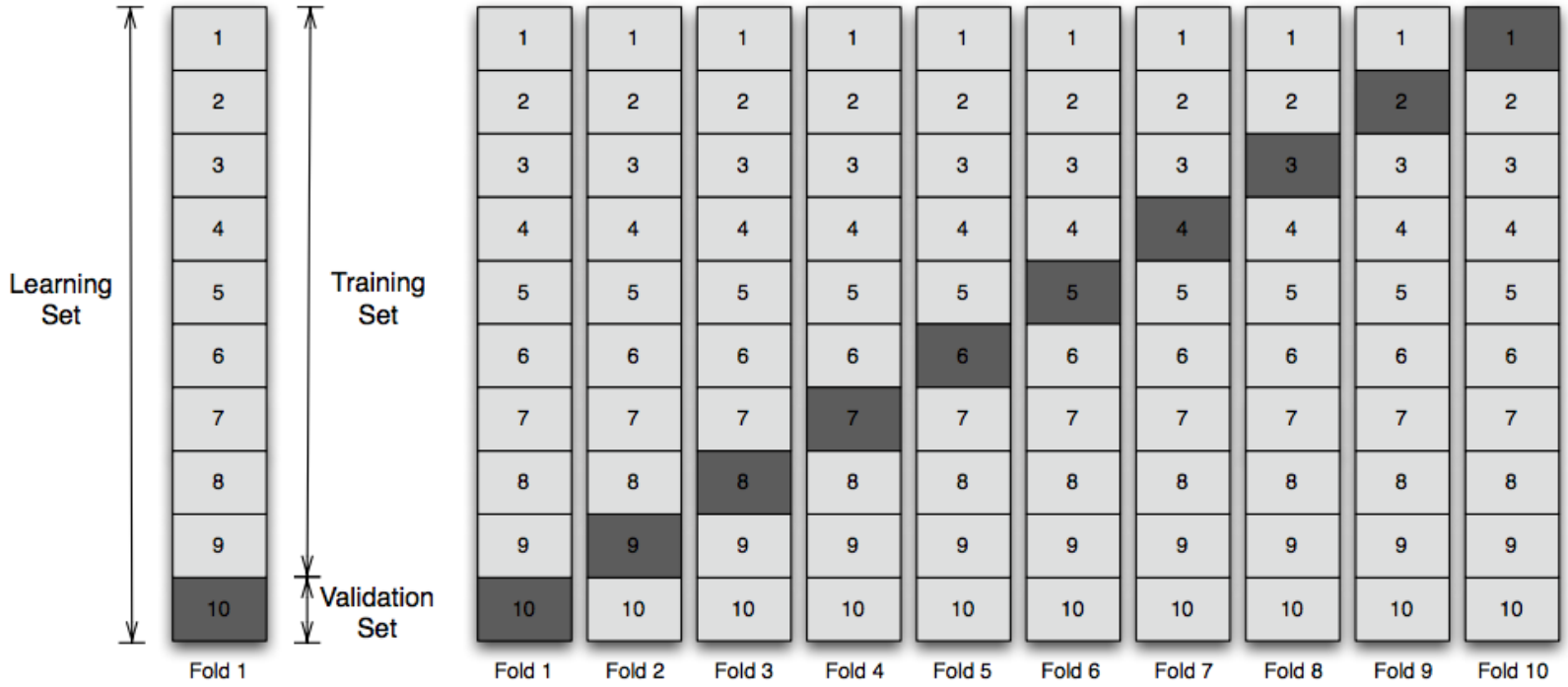
$$\mathbf{E}^*(Y \mid A = 0, W) = \text{expit}\left[\textbf{logit}\left[E(Y \mid A = 0, W)\right] + \hat{\epsilon_0} H_0(0, W)\right]$$

Targeted estimate of the ATE from $\Psi^{(0)}$ to $\Psi^{(1)}$: ($\widehat{\Psi}$)

$$\Psi^{(1)} : \hat{\Psi} = \left[\mathbf{E}^*(Y(1) \mid A = 1, W) - \mathbf{E}^*(Y(0) \mid A = 0, W)\right]$$

expit(x) = 1/(1+exp(-x)); logit(x) = log(x/1-x)

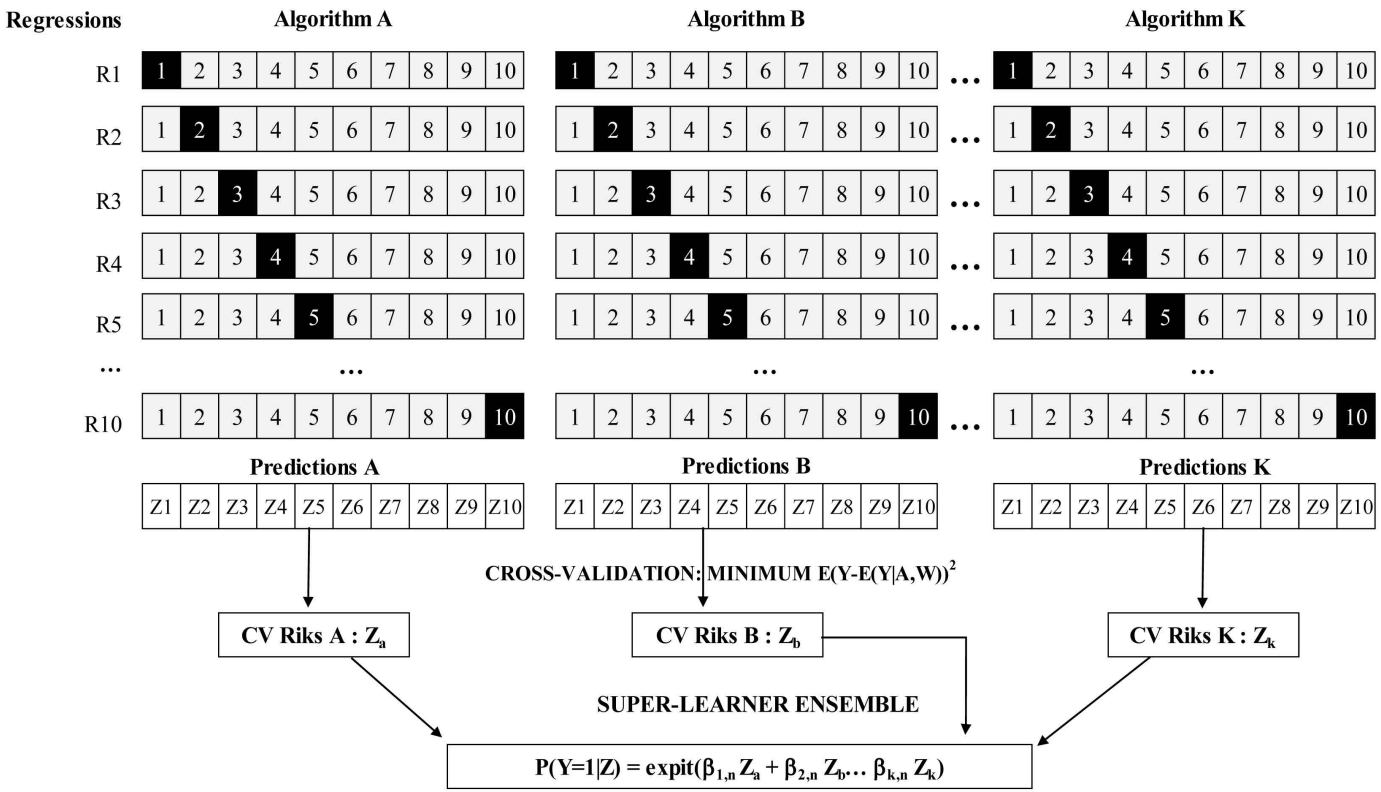
Targeted learning



Source: Mark van der Laan and Sherri Rose. Targeted learning: causal inference for observational and experimental data. Springer Series in Statistics, 2011. **Reference** NNLS: Breiman L. Stacked regressions. Mach Learn 1996;24:49–64.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Super-Learner: Ensemble learning



To apply the **EIC** we need data-adaptive estimation for both, the model of the outcome, and the model of the treatment.
Asymptotically, the final weighted combination of algorithms (Super Learner) performs as well as or better than the best-fitting algorithm (van der Laan, 2007).

Non Negative Less Square and negative log-likelihood algorithms to get predictions from the ensemble learning (Breiman, Van der Laan)

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

TMLE inference: INFLUENCE FUNCTION

M-ESTIMATORS: Semi-parametric and Empirical processes theory

An estimator is **asymptotically linear** with **influence function** φ **(IC)** if the estimator can be **approximate by an empirical average** in the sense that

$$(\hat{\theta} - \theta_0) = \frac{1}{n}\sum_{i=1}^{n}(IC) + Op(1/\sqrt{n})$$

(Bickel, 1997).

TMLE inference: Bickel (1993); Tsiatis (2007); Van der Laan (2011); Kennedy (2016)

- The **IC** estimation is a more general approach than M-estimation.
- The **Efficient IC** has mean zero $E(IC_{\hat{\psi}}(y_i, \psi_0)) = 0$ and **finite variance**.
- By the Weak Law of the Large Numbers, the **Op** converges to zero in a rate $1/\sqrt{n}$ as n $\rightarrow\infty$ (Bickel, 1993).
- The **Efficient IC** requires **asymptotically linear** estimators.

MEDICINE

Influence Function for the ATE

TMLE statistical inference for the ATE

$$\mathbf{IF_{ATE}} = \phi'(\theta)(\hat{\theta} - \theta) =$$
$$1 \times \left[\psi - \left(\frac{(A_i = 1)}{P(A_i = 1 \mid W_i)} - \frac{(A_i = 0)}{P(A_i = 0 \mid W_i)} \right) [Y_i - E_1(Y \mid A_i, W_i)] + \right.$$
$$\left. [E_1(Y(1) \mid A_i = 1, W_i) - E_1(Y(0) \mid A_i = 0, W_i)] \right]$$

Type Wald Confidence Intervals

$$\textbf{Standard Error} : \sigma(\psi_0) = \frac{SD(IF_n)}{\sqrt{n}}$$

Functional Delta Method: Tutorial

Statistics > Methodology

[Submitted on 30 Jun 2022]

The Delta-Method and Influence Function in Medical Statistics: a Reproducible Tutorial

Rodrigo Zepeda-Tello, Michael Schomaker, Camille Maringe, Matthew J. Smith, Aurelien Belot, Bernard Rachet, Mireille E. Schnitzer, Miguel Angel Luque-Fernandez

Approximate statistical inference via determination of the asymptotic distribution of a statistic is routinely used for inference in applied medical statistics (e.g. to estimate the standard error of the marginal or conditional risk ratio). One method for variance estimation is the classical Delta-method but there is a knowledge gap as this method is not routinely included in training for applied medical statistics and its uses are not widely understood. Given that a smooth function of an asymptotically normal estimator is also asymptotically normally distributed, the Delta-method allows approximating the large-sample variance of a function of an estimator with known large-sample properties. In a more general setting, it is a technique for approximating the variance of a functional (i.e., an estimand) that takes a function as an input and applies another function to it (e.g. the expectation function). Specifically, we may approximate the variance of the function using the functional Delta-method based on the influence function (IF). The IF explores how a functional $\phi(\theta)$ changes in response to small perturbations in the sample distribution of the estimator and allows computing the empirical standard error of the distribution of the functional. The ongoing development of new methods and techniques may pose a challenge for applied statisticians who are interested in mastering the application of these methods. In this tutorial, we review the use of the classical and functional Delta-method and their links to the IF from a practical perspective. We illustrate the methods using a

Recall that the derivative, $\partial_v \phi(\theta)$, represents the slope of the line tangent to the function. Intuitively, if $\hat{\theta}$ is close to θ, the tangent line at $\hat{\theta}$ should provide an adequate approximation of $\phi(\theta)$. This is stated in the Taylor first order approximation of $\phi(\hat{\theta})$ around $\phi(\theta)$ as follows:

$$\phi(\hat{\theta}_n) \approx \phi(\theta) + \partial_v \phi(\theta) \tag{1}$$

with $v = \hat{\theta} - \theta$ and the sign $\approx$ is interpreted as *approximately equal*. This can be rewritten as the more classical approach:

$$\phi(\hat{\theta}) - \phi(\theta) \approx \partial_v \phi(\theta) \quad \text{with } v = \hat{\theta} - \theta. \tag{2}$$

Readers might be familiar with the theorem in the classical notation of univariate calculus which states the approximation:

$$\phi(\hat{\theta}) \approx \phi(\theta) + \phi'(\theta) \underbrace{(\hat{\theta} - \theta)}_{v}. \tag{3}$$

In this case, the Hadamard derivative coincides with the classical one multiplied by $v = \hat{\theta} - \theta$:

$$\partial_v \phi(\theta) = \phi'(\theta)(\hat{\theta} - \theta).$$

The justification for this connection is given by Fréchet's derivative which represents the slope of the tangent plane. Intuitively, if the Hadamard (one-sided directional) derivatives $\partial_v \phi(\theta)$ exist for all directions v we can talk about the tangent plane to ϕ at θ. The tangent plane is "made up" of all the individual (infinite) tangent lines. The slope of the tangent plane is the Fréchet derivative $\nabla \phi$. For univariate functions in $\phi : \mathbb{R} \to \mathbb{R}$ the Fréchet derivative is ϕ'; for functions of a multivariate θ returning one value, $\phi : \mathbb{R}^n \to \mathbb{R}$, this derivative is called the gradient and corresponds to the derivative of the function by each entry:

$$\nabla \phi = \Big(\frac{\partial \phi}{\partial \theta_1}, \frac{\partial \phi}{\partial \theta_2}, \ldots, \frac{\partial \phi}{\partial \theta_n} \Big).$$

For multivariate functions, $\phi : \mathbb{R}^n \to \mathbb{R}^m$, the Fréchet derivative is an $m \times n$ matrix called the Jacobian (matrix):

$$\nabla \phi = \begin{pmatrix} \frac{\partial \phi_1}{\partial \theta_1} & \frac{\partial \phi_1}{\partial \theta_2} & \cdots & \frac{\partial \phi_1}{\partial \theta_n} \\ \frac{\partial \phi_2}{\partial \theta_1} & \frac{\partial \phi_2}{\partial \theta_2} & \cdots & \frac{\partial \phi_2}{\partial \theta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_m}{\partial \theta_1} & \frac{\partial \phi_m}{\partial \theta_2} & \cdots & \frac{\partial \phi_m}{\partial \theta_n} \end{pmatrix}. \tag{4}$$

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Stata **ELTMLE**

Ensemble Learning Targeted Maximum Likelihood Estimation

- **eltmle** is a Stata program implementing R-TMLE for the ATE for a binary or continuous outcome and binary treatment.
- **eltmle** includes the use of a **super-learner**(Polley E., et al. 2011).
- I used the default Super-Learner algorithms implemented in the base installation of the tmle-R package v.1.2.0-5 (Susan G. and Van der Laan M., 2007).
- i) stepwise selection, ii) GLM, iii) a GLM interaction.
- Additionally, **eltmle** users will have the option to include Bayes GLM and GAM.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Stata ELTMLE

Syntax eltmle Stata command

eltmle Y A W [, **slapiw slaipwbgam tmle tmlebgam**]

Y: Outcome: numeric binary or continuous variable.

A: Treatment or exposure: numeric binary variable.

W: Covariates: vector of numeric and categorical variables.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Stata Implementation: overall structure

```
45
46  capture program drop eltmle
47  program define eltmle
48       syntax [varlist] [if] [pw] [, slaipw slaipwbgam tmle tmlebgam]
49       version 13.2
50       marksample touse
51       local var `varlist' if `touse'
52       tokenize `var'
53       local yvar = "`1'"
54       global flag = cond(`yvar'<=1,1,0)
55       qui sum `yvar'
56       global b = `r(max)'
57       global a = `r(min)'
58       qui replace `yvar' = (`yvar' - `r(min)') / (`r(max)' - `r(min)') if `yvar'>1
59       local dir `c(pwd)'
60       cd "`dir'"
61       qui export delimited `var' using "data.csv", nolabel replace
62       if "`slaipw'" == "" & "`slaipwbgam'" == "" & "`tmlebgam'" == "" {
63          tmle `varlist'
64       }
65       else if "`tmlebgam'" == "tmlebgam" {
66          tmlebgam `varlist'
67       }
68       else if "`slaipw'" == "slaipw" {
69          slaipw `varlist'
70       }
71       else if "`slaipwbgam'" == "slaipwbgam" {
72          slaipwbgam `varlist'
73       }
74  end
```

Stata Implementation: R code for calling the SL

```
program tmle
// Write R Code dependencies: foreign Surperlearner
set more off
qui: file close _all
qui: file open rcode using SLS.R, write replace
qui: file write rcode ///
        `"set.seed(123)"' _newline ///
        `"list.of.packages <- c("foreign","SuperLearner")"' _newline ///
        `"new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]"' _newline ///
        `"if(length(new.packages)) install.packages(new.packages, repos='http://cran.us.r-project.org')"' _newline ///
        `"library(SuperLearner)"' _newline ///
        `"library(foreign)"' _newline ///
        `"data <- read.csv("data.csv", sep=",")"' _newline ///
        `"attach(data)"' _newline ///
        `"SL.library <- c("SL.glm","SL.step","SL.glm.interaction")"' _newline ///
        `"n <- nrow(data)"' _newline ///
        `"nvar <- dim(data)[[2]]"' _newline ///
        `"Y <- data[,1]"' _newline ///
        `"A <- data[,2]"' _newline ///
        `"X <- data[,2:nvar]"' _newline ///
        `"W <- data[,3:nvar]"' _newline ///
        `"X1 <- X0 <- X"' _newline ///
        `"X1[,1] <- 1"' _newline ///
        `"X0[,1] <- 0"' _newline ///
        `"newdata <- rbind(X,X1,X0)"' _newline ///
        `"Q <- try(SuperLearner(Y = data[,1] ,X = X, SL.library=SL.library, family=binomial(), newX=newdata, method="me
        `"Q <- as.data.frame(Q[[4]])"' _newline ///
        `"QAW <- Q[1:n,]"' _newline ///
        `"Q1W <- Q[((n+1):(2*n)),]"' _newline ///
        `"Q0W <- Q[((2*n+1):(3*n)),]"' _newline ///
        `"g <- suppressWarnings(SuperLearner(Y = data[,2], X = W, SL.library = SL.library, family = binomial(), method
        `"ps <- g[[4]]"' _newline ///
        `"ps[ps<0.025] <- 0.025"' _newline ///
        `"ps[ps>0.975] <- 0.975"' _newline ///
        `"data <- cbind(data,QAW,Q1W,Q0W,ps,Y,A)"' _newline ///
        `"write.dta(data, "data2.dta")"'
qui: file close rcode
```

Stata Implementation: Batch file executing R

```
112  qui: file close rcode
113
114  // Write bacth file to find R.exe path and R version
115  set more off
116  qui: file close _all
117  qui: file open bat using setup.bat, write replace
118  qui: file write bat ///
119  `"@echo off"' _newline ///
120  `"SET PATHROOT=C:\Program Files\R\"' _newline ///
121  `"echo Locating path of R..."' _newline ///
122  `"echo."' _newline ///
123  `"if not exist "%PATHROOT%" goto:NO_R"' _newline ///
124  `"for /f "delims=" %%r in (' dir /b "%PATHROOT%R*" ') do ("' _newline ///
125          `"echo Found %%r"' _newline ///
126          `"echo shell "%PATHROOT%%%r\bin\x64\R.exe" CMD BATCH SLS.R > runr.do"' _newline ///
127          `"echo All set!"' _newline ///
128          `"goto:DONE"' _newline ///
129  `")"' _newline ///
130  `":NO_R"' _newline ///
131  `"echo R is not installed in your system."' _newline ///
132  `"echo."' _newline ///
133  `"echo Download it from https://cran.r-project.org/bin/windows/base/"' _newline ///
134  `"echo Install it and re-run this script"' _newline ///
135  `":DONE"' _newline ///
136  `"echo."' _newline ///
137  `"pause"'
138  qui: file close bat
139
140  //Run batch
141  shell setup.bat
142  //Run R
143  do runr.do
144
145  // Read Revised Data Back to Stata
146  clear
147  quietly: use "data2.dta", clear
148
149  // Q to logit scale
150  gen logQAW = log(QAW / (1 - QAW))
151  gen logQ1W = log(Q1W / (1 - Q1W))
152  gen logQ0W = log(Q0W / (1 - Q0W))
153
154  // Clever covariate HAW
```

Output for continuous outcome

```
.use http://www.stata-press.com/data/r14/cattaneo2.dta
.eltmle bweight mbsmoke mage medu prenatal mmarried, tmle

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        POM1 |      4,642    2832.384    74.56757   2580.186   2957.627
        POM0 |      4,642    3063.015    89.53935   2868.071   3167.264
          WT |      4,642   -.0409955    2.830591  -6.644464   21.43709
          PS |      4,642    .1861267     .110755   .0372202   .8494988
ACE:
Additive Effect: -230.63; Estimated Variance: 600.93; p-value: 0.0000;
 95%CI:(-278.68, -182.58)

Risk Differences:-0.0447; SE: 0.0047; p-value: 0.0000;
     95%CI:(-0.05, -0.04)
```

Simulations comparing Stata ELTMLE vs R-TMLE

```
. mean psi aipw slaipw tmle
Mean estimation
Number of obs    =  1,000
------------------------
             |   Mean
-------------+----------
        True |   .173
        aipw |   .170
      slaipw |   .170
  Stata-tmle |   .170
------------------------
      R-TMLE |   .170
------------------------
```

ONLINE open free tutorial

Link to the tutorial

https://migariane.github.io/TMLE.nb.html

Stata Implementation: source code

https://github.com/migariane/meltmle for MAC users
https://github.com/migariane/weltmle for Windows users

Stata installation and step by step commented syntax

github install migariane/meltmle (For MAC users)
github install migariane/weltmle (For Windows users)
which eltmle
viewsource eltmle.ado

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

One sample simulation: TMLE reduces bias

https://github.com/migariane/SUGML

Wiley Online Library | Universidad De Granada

Statistics in Medicine

TUTORIAL IN BIOSTATISTICS | Open Access

Targeted maximum likelihood estimation for a binary treatment: A tutorial

Miguel Angel Luque-Fernandez, Michael Schomaker, Bernard Rachet, Mireille E. Schnitzer

First published: 23 April 2018 | https://doi.org/10.1002/sim.7628 | Citations: 45

Abstract

When estimating the average effect of a binary treatment (or exposure) on an outcome, methods that incorporate propensity scores, the G-formula, or targeted maximum likelihood estimation (TMLE) are preferred over naïve regression approaches, which are biased under misspecification of a parametric outcome model. In contrast propensity score methods require the correct specification of an exposure model. Double-robust methods only

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Tutorials Causal Inference

TMLE R-markdown 2019

https://migariane.github.io/TMLE.nb.html

TMLE Statistics in Medicine 2018

https://www.ncbi.nlm.nih.gov/pubmed/29687470

SIM-2018. CODE in R and STATA

https://github.com/migariane/SIM-TMLE-tutorial

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Statistics in Medicine

TUTORIAL IN BIOSTATISTICS | Open Access

Introduction to computational causal inference using reproducible Stata, R, and Python code: A tutorial

Matthew J. Smith, Mohammad A. Mansournia, Camille Maringe, Paul N. Zivich, Stephen R. Cole, Clémence Leyrat, Aurélien Belot, Bernard Rachet, Miguel A. Luque-Fernandez ... See fewer authors

First published: 28 October 2021 | https://doi.org/10.1002/sim.9234

Funding information: Cancer Research UK, Instituto de Salud Carlos III

Abstract

The main purpose of many medical studies is to estimate the effects of a treatment or exposure on an outcome. However, it is not always possible to randomize the study participants to a particular treatment, therefore observational study designs may be used. There are major challenges with

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

WILEY

Congratulations!

Your work is a top downloaded article.

We're pleased to offer you a certificate of achievement in recognition of your paper.

Download Certificate

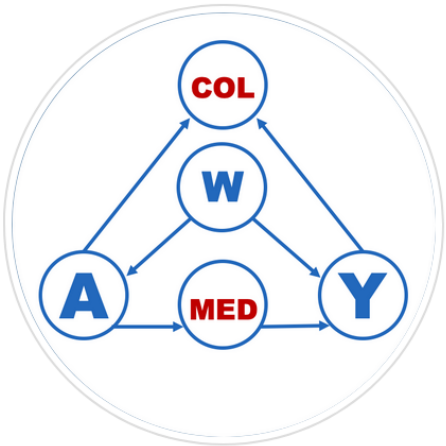
Dear Miguel,

We are excited to share that your work, published in *Statistics in Medicine*, is among our most downloaded papers!*

- Introduction to computational causal inference using reproducible Stata, R, and Python code: A tutorial

Tell everyone about your achievement

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE



Curso de Introducción a la Inferencia Causal en Epidemiología

3 de septiembre, 2019

Facultad de Medicina y Ciencias de la Salud de Oviedo

Curso precongreso SEE

Hugo SP-Minimal Theme | Based on Minimal

Curso de Introducción a la Inferencia Causal en Epidemiología

Descripción del curso

El Curso de Introducción a la Inferencia Causal en Epidemiología está dirigido a estadísticos y epidemiólogos de la administración sanitaria interesados en la Inferencia Causal, con conocimientos en Estadística Aplicada a las Ciencias de la Salud y Epidemiología.

Objetivos

El objetivo general del curso es ofrecer una visión general de los nuevos métodos en epidemiología basados en el marco conceptual de las "*Potential Outcomes*" y la "*Inferencia Causal*" utilizada para evaluar la efectividad de intervenciones o tratamientos basada en datos observacionales.

Los objetivos específicos son:

1. Obtener una visión general del desarrollo de la disciplina y la necesidad de la "*Inferencia Causal*".
2. Entender y aplicar el marco conceptual de las "*Potential Outcomes*" en escenarios simples.

CI crash short course

LABs 1-2: G-Comp. and IPTW

https://ccci.netlify.app/ G-Comp. and IPTW using R (RStudio cloud)

RStudio Cloud link

https://rstudio.cloud/spaces/19488/project/434105

Next steps for ELTMLE

Next steps

- Stata Journal manuscript.
- SIM manuscritp running simulations comparing elmtle and cveltmle.
- Improving the user interface for **eltmle**.
- Include positivity and near positivity violations diagnostic tools (tbalance table and figure).
- Implementation of Ensemble Learning calling Python 3.
- Cross-validate TMLE. Recently, we have implemented the **cross-validated AUC**: https://github.com/migariane/cvAUROC. Also available at **ssc**: ssc install cvAUROC

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

cvauroc

Cross-validation in Stata

The Stata Journal: Promoting communications on statistics and Stata

Impact Factor: **4.450**
5-Year Impact Factor: **5.880**

JOURNAL HOM

Available access | Research article | First published online September 20, 2019

cvauroc: Command to compute cross-validated area under the curve for ROC anal modeling for binary outcomes

Miguel Angel Luque-Fernandez, Daniel Redondo-Sánchez, and Camille Maringe View all authors and affiliations

Volume 19, Issue 3 | https://doi.org/10.1177/1536867X19874237

Contents | PDF / ePub

Abstract

Receiver operating characteristic (ROC) analysis is used for comparing predictive models in both model selection and model evaluation. ROC analysis is often applied in clinical medicine and social science to assess the tradeoff between model sensitivity and specificity. After fitting a binary logistic or probit regression model with a set of independent variables, the predictive performance of this set of variables can be assessed by the area under the curve (AUC) from an ROC curve. An important aspect of predictive

References

References

1. Bickel, Peter J.; Klaassen, Chris A.J.; Ritov, Yaacov; Wellner Jon A. (1997). Efficient and adaptive estimation for semiparametric models. New York: Springer.
2. Hample, F.R., (1974). The influence curve and its role in robust estimation. J Amer Statist Asso. 69, 375-391.
3. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. J Amer Statist Assoc. 1994;89:846–866.
4. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005;61:962–972.
5. Tsiatis AA. Semiparametric Theory and Missing Data. Springer; New York: 2006
6. Kang JD, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical Science. 2007;22(4):523–539
7. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974;66:688–701

MEDICINE

References

References

1. Luque-Fernandez, Miguel Angel. (2017). Targeted Maximum Likelihood Estimation for a Binary Outcome: Tutorial and Guided Implementation.
2. StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.
3. Gruber S, Laan M van der. (2011). Tmle: An R package for targeted maximum likelihood estimation. UC Berkeley Division of Biostatistics Working Paper Series.
4. Laan M van der, Rose S. (2011). Targeted learning: Causal inference for observational and experimental data. Springer Series in Statistics.626p.
5. Van der Laan MJ, Polley EC, Hubbard AE. (2007). Super learner. Statistical applications in genetics and molecular biology 6.
6. Bickel, Peter J.; Klaassen, Chris A.J.; Ritov, Ya'acov; Wellner Jon A. (1997). Efficient and adaptive estimation for semiparametric models. New York: Springer.
7. E. H. Kennedy. Semiparametric theory and empirical processes in causal inference. In: Statistical Causal Inferences and Their Applications in Public Health Research, in press.

MEDICINE

Acknowledgements

Berkeley Division de Bioestadistica

The motivation to implement the TMLE framework in Stata come from MALF's work in a visiting academic position in the Division of Biostatistics at the Berkeley School of Public Health in 2018. We acknowledge Professor Mark J. van der Laan for his support and guidance.

UK MRC GRANT

This work was supported by the UK Medical Research Council [grant number MR/W021021/1]

Thank YOU!!!

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE