Roadmap of Targeted Learning

A. ROADMAP FOR THE AVERAGE TREATMENT EFFECT (ATE)

Comparison of Standard Approach to Targeted Maximum Likelihood Estimation

1. DESCRIBE DATA

$n = 100$ subjects

For each subject, pre-treatment covariates (W), treatment (A), and outcome (Y) vectors were measured

W pre-treatment covariate

A indicating whether subject received treatment ($A = 1$) or not ($A = 0$)

Y continuous post-treatment outcome

Subjects were sampled independently from eachother and from the same population distribution P_0

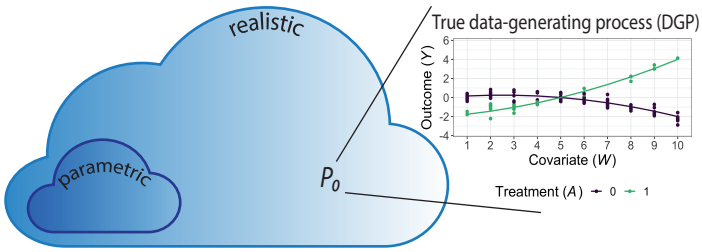
$$O_1, \ldots, O_n \overset{iid}{\sim} P_0$$

2. SPECIFY STATISTICAL MODEL

Standard Approach

Parametric statistical model

Does not contain P_0, the DGP (i.e., misspecified model)



Targeted Maximum Likelihood Estimation

Realistic semiparametric or nonparametric statistical model

Defined to ensure P_0 is contained in model

3. DEFINE ESTIMAND

Additional assumptions are required to interpret this estimand as causal

What is the average effect of treatment on outcome?

$$\Psi(P_0) = E_0\big(E_0[Y|A=1,W] - E_0[Y|A=0,W]\big)$$

Ψ is a function that takes as input P_0 and outputs the answer to the question of interest

The ***assumption of positivity*** is requied to estimate of this quantity from the data. That is, it must be possible to observe both levels of treatment for all strata of W.

4. CONSTRUCT ESTIMATOR

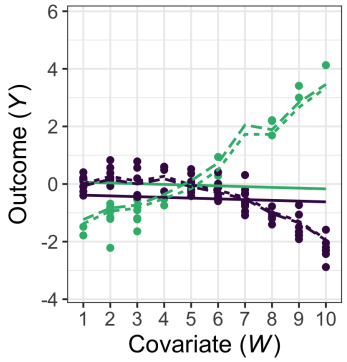
Standard Approach

Generalized Linear Model (GLM) to estimate

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{A} + \beta_2 \mathbf{W} + \boldsymbol{\epsilon}$$

Estimated coefficients are biased

Cannot detect heterogeneity in treatment effect



Targeted Maximum Likelihood Estimation

TMLE implements a two-step procedure

1. initial estimation of $E_0[Y|A,W]$ with super (machine) learning
2. targeting towards optimal bias-variance trade-off for $\Psi(P_0)$

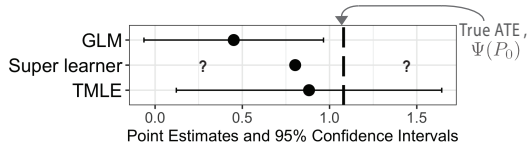
TMLE estimates are unbiased and doubly robust

To have legitimate estimate and inference, statistical model must be (1) correctly specified (i.e., contain P_0) and (2) selected a priori, before looking at the data.

5. FORM INFERENCE

Standard Approach

Inference (such as *p*-value and confidence interval) is misleading and erroneous



Targeted Maximum Likelihood Estimation

Targeting (step 2) improves estimate and makes inference possible

Trustworthy inference obtained with efficient influence function

B. ROADMAP FOR THE OPTIMAL TREATMENT EFFECT (steps identical to ATE omitted)

3. DEFINE ESTIMAND

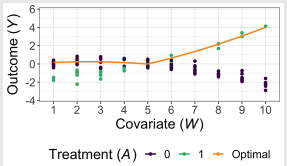
What is the effect of optimal individualized treatment on outcome?

$$\Psi(P_0) = E_0\big(E_0[Y|A = d_{\text{opt}}(W), W]\big) - E_0[Y]$$

$d_{opt}(W)$ is a decision rule that tailors treatment assignment to the subject (based on their characteristics W) to maximize their expected outcome.

$E_0[Y]$ is the mean outcome under the observed treatment assignment.

The optimal intervention assigns treatment $A = 1$ if $W \geq 5$ and $A = 0$ if $W < 5$

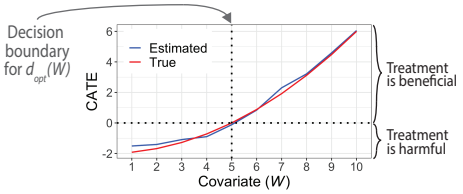


4. CONSTRUCT ESTIMATOR

Step 1 of TMLE requires estimation of the conditional additive treatment effect (CATE)

$$E_0(Y|A=1,W) - E_0(Y|A=0,W),$$

to learn the decision boundary for optimal rule $d_{opt}(W)$ to assign treatment A



5. FORM INFERENCE