Targeted Learning for Data Adaptive Causal Inference in Observational and Randomized Studies

Maya Petersen and Mark van der Laan

Department of Biostatistics, University of California, Berkeley School of Public Health

Outline of estimation part of course

- Part 1
  - Targeted Learning Overview
  - Estimation Roadmap
  - Super Learning
- Part 2
  - Targeted Minimum Loss-Based Estimation (TMLE)
- Part 3
  - TMLE for longitudinal data analysis
  - Concluding Remarks

Estimation is a Science, Not an Art

- Data: Realizations of random variables with probability distribution P_0
- Statistical Model ($\mathcal{M}$): Actual knowledge about the shape of P_0
- Statistical Target Parameter: A feature/function of P_0
- Estimator
  - Algorithm for mapping from data to a (d-dimensional) real number
  - Benchmarked by a dissimilarity measure (e.g., MSE) w.r.t. target parameter

Causal Inference

- Under non-testable assumptions P_0 can be described in terms of an underlying parameter varying over an underlying parameter space
  - e.g., intervention-specific counterfactuals
  - Parameter space described by a full data model such as a nonparametric structural equation model (NPSEM)
- Non-testable assumptions enrich the interpretation of the statistical target parameter
  - Model = Statistical Model + Untestable Assumptions
  - Allows identification of full data parameters and causal quantities
  - Defintion of statistical target parameter is clear. Causal interpretation when assumptions are met
- **Statistical estimation is concerned only with statistical model and statistical target parameter**

Example: TMLE for the ATE parameter

Marginal additive effect of a binary point treatment (ATE) parameter

- Data: n i.i.d. observations $O = (Y, A, W) \sim P_0$

  outcome Y, binary treatment indicator A, covariate vector W

- NPSEM

$$\begin{aligned} W &= f_W(U_W) \\ A &= f_A(W, U_A) \\ Y &= f_Y(W, A, U_Y) \end{aligned}$$

  where U_W, U_A, U_Y are uncorrelated exogenous random errors

- Target Parameter: $\psi_0^{ATE} = E_0(Y_1 - Y_0)$
  - Y_a is counterfactual outcome generated under NPSEM with A set to a

A Substitution Estimator

- Target parameter ψ_0 is a feature of P_0
  - ψ_0 can be expressed as a mapping, $\Psi(P_0)$
  - ψ_0 is sometimes a feature of only a portion of P_0 denoted by Q_0
  - Thus, $\Psi(Q_0) = \Psi(P_0)$
- A substitution estimator applies the target parameter mapping directly to an estimate of relevent component
  - Q_n is an estimator of Q_0 that conforms to statistical model $\mathcal{M}$
  - Substitution estimator can be represented as mapping $\psi_n = \Psi(Q_n)$.
- Use of a substitution estimator enhances robustness by respecting bounds on both $\mathcal{M}$ and ψ_0

TMLE for the ATE parameter

- Likelihood factorizes

$$\mathcal{L}(O) = \underbrace{P(Y \mid A, W)}_{Q_Y} \underbrace{P(A \mid W)}_{g} \underbrace{P(W)}_{Q_W}$$

- Define

$$Q_0 = (Q_{0_Y}, Q_{0_W})$$
$$g_0 = P_0(A \mid W)$$

- When causal assumptions are met ψ_0^{ATE} equals

$$\begin{aligned}\psi_0 &= E_{0_W}[E_0(Y \mid A = 1, W) - E_0(Y \mid A = 0, W)] \\ &= E_{0_W}\left[\bar{Q}_{0_Y}(1, W) - \bar{Q}_{0_Y}(0, W)\right]\end{aligned}$$

- Otherwise, ψ_0 remains a useful measure of variable importance

Motivation for TMLE

- Super Learner (SL) for estimating Q_0
- SL-based substitution estimator evaluates $\psi_n^{SL} = \Psi(Q_n)$
- TMLE fluctuates initial Q_n to obtain targeted Q_n^*
  - Targeting makes use of information in P_0 beyond Q_0 to improve estimation of ψ_0
  - Provides an opportunity to
    - reduce asymptotic bias if initial Q_n not consistent
    - reduce finite sample bias
    - reduce variance

Pathwise Differentiable Parameter

- Pathwise derivative for a path $\{P(\epsilon) : \epsilon\} \subset \mathcal{M}$ through P at $\epsilon = 0$ is defined by $\frac{d}{d\epsilon}\Psi(P(\epsilon))|_{\epsilon=0}$
- If for all paths through P this derivative can be represented as

$$P(D^*(P)S) \equiv \int D^*(P)(o)S(o)dP(o),$$

where S is the score of the path at $\epsilon = 0$,

and $D^*(P)$ is an element of tangent space at P,

then

  - the target parameter mapping is pathwise differentiable at P
  - its canonical gradient (i.e., efficient influence curve), is $D^*(P)$

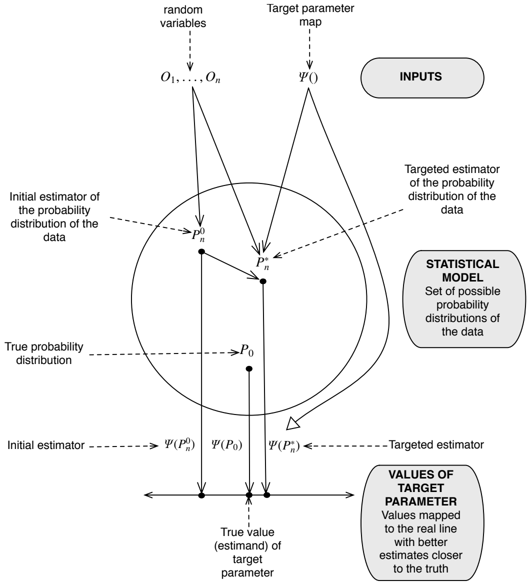
Efficient Influence Curve Equation

- An estimator is asymptotically efficient if and only if it is asymptotically linear with influence curve the efficient influence curve $D^*(P_0)$:

$$\psi_n - \psi_0 = \frac{1}{n} \sum D^*(P_0)(O_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- An efficient regular asymptotically linear estimator will need to solve the efficient influence curve equation $\sum_i D^*(P)(O_i) = 0$ (up to second order term)
- TMLE is a double robust semi-parametric efficient RAL substitution estimator that can be applied to estimate any pathwise differential parameter of P_0.

TMLE



random variables
Target parameter map
$O_1, \ldots, O_n$
$\Psi()$
INPUTS
Initial estimator of the probability distribution of the data
Targeted estimator of the probability distribution of the data
P_n^0
P_n^*
STATISTICAL MODEL
Set of possible probability distributions of the data
True probability distribution
P_0
Initial estimator
$\Psi(P_n^0)$
$\Psi(P_0)$
$\Psi(P_n^*)$
Targeted estimator
VALUES OF TARGET PARAMETER
Values mapped to the real line with better estimates closer to the truth
True value (estimand) of target parameter

Targeted Minimum Loss-Based Estimation

- TMLE Procedure
  1. Identify the "hardest" parametric submodel to fluctuate initial $\hat{P}$
     Small fluctuation $\rightarrow$ maximum change in target
  2. Identify optimal magnitude of fluctuation by MLE
  3. Apply optimal fluctuation to $\hat{P}$ to obtain 1st-step TMLE
  4. Repeat until incremental fluctuation is zero
     - 1-step convergence guaranteed in some important cases
  5. Final probability distribution solves efficient influence curve equation
     - basis for asymptotic linearity, normality, and efficiency.
     - confers double robustness, or, more general, makes bias a second order term.
- Asymptotically efficient when initial and treatment/censoring mechanism estimators are both consistent
- Allows incorporation of machine learning while preserving inference

TMLE Algorithm for our ATE example

- Step 1: Obtain initial estimate

$$\bar{Q}_n^0(A, W) = \hat{E}(Y \mid A, W)$$

- Step 2: Target initial estimate (logit scale)

$$\bar{Q}_n^*(A, W) = \bar{Q}_n^0(A, W) + \hat{\epsilon} h_{g_n}(A, W)$$

  - Estimate $g_0(A, W)$ (propensity score)
  - Construct parameter-specific fluctuation covariate, e..g,

$$h_{g_n}^{ATE} = \left[\frac{A}{g_n(1, W)} - \frac{1 - A}{g_n(0, W)} \right]$$

  - Maximum likelihood to fit ϵ

- Evaluate parameter: $\psi_n^{TMLE} = \Psi(\bar{Q}_n^*)$

Efficient Influence Curve for ATE

TMLE solves $P_n D^*(P_n^*) = 0$

- Efficient influence curve for ATE parameter

$$D^*(P) = \underbrace{\left[\frac{A}{g(1,W)} - \frac{1-A}{g(0,W)}\right]\left[Y - \bar{Q}(A,W)\right]}_{a} + \underbrace{\bar{Q}(1,W) - \bar{Q}(0,W) - \psi}_{b}$$

- Stage 2 targeting fits ϵ by maximum likelihood
  - MLE solves score equation $\sum_i h_{g_n}(A_i, W_i)[Y_i - \bar{Q}_n^*(A_i, W_i)] = 0$
  - We define parameter-specific h_g to ensure that the empirical mean of a equals 0.
- As a substitution estimator $\psi_n^{TMLE} = \frac{1}{n}\sum_i \bar{Q}_n^*(1, W_i) - \bar{Q}_n^*(0, W_i)$, thus empirical mean of b equals 0.

Inference

- Asymptotic Linearity

$$\sqrt{n}(\psi_n^{TMLE} - \psi_0) \xrightarrow{D} N(0, \sigma^2)$$

- 95% confidence intervals

$$\psi_n(Q_n^*) \pm 1.96\, \hat{\sigma}/\sqrt{n}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{D}^{*^2}(P_n^*)(O_i)$$

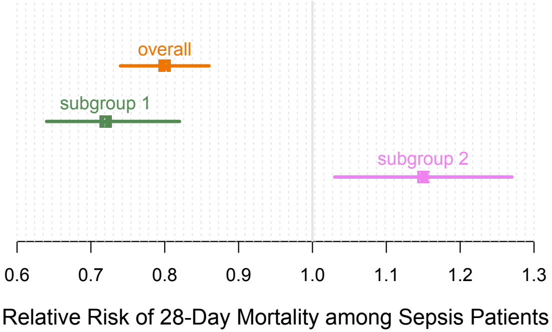
- Test statistic for null hypothesis H_0: $\psi_0 = 0$

$$T = \frac{\psi_n}{\sqrt{\hat{\sigma}^2/n}}$$

- p-values are calculated as $2\Phi(-abs(T))$
  Φ = CDF of standard normal distribution

Effect of Steroids on Mortality in Adults with Septic Shock

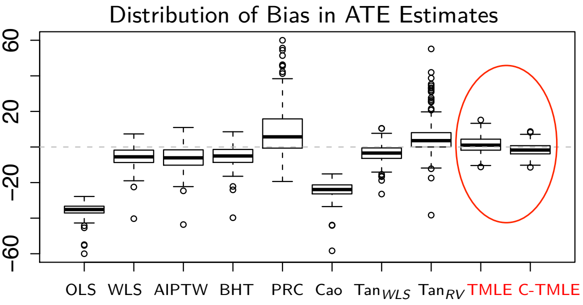
- Patient-level data from three major randomized controlled trials*
- Conflicting results among the three studies
- Using TMLE + SL to estimate RR of mortality (steroid vs. placebo)
  - Reduced variance
  - Power to estimate significant heterogeneous subgroup effects



*Pirracchio, et. al. under review

Simulation Study: ATE parameter estimation

- TMLE and C-TMLE compared with ordinary least squares and seven other double robust estimators
- Challenges: Misspecified outcome regression, correct propensity score model produces near positivity violations
- TMLE and C-TMLE were least biased and had smallest variance



Porter, Gruber, Sekhon, and van der Laan. *The International Journal of Biostatistics*, 2011.

Additional topics

- Point treatment parameters
  - Relative risk, odds ratio, risk difference
  - Mean outcome under missingness in the population
  - Effect of treatment among the treated (ATT)
  - Controlled direct effects
  - Marginal structural model parameters
- Case-control and other biased sampling designs
- Addressing common challenges in data analysis
  - Near positivity violations (poor overlap): Collaborative TMLE (C-TMLE) for data-adaptive nuisance parameter estimation
  - TMLE for bounded continuous outcomes
  - TMLE for rare outcomes

TMLE Demonstration

- *tmle* R package (CRAN)
- Using the package
  - Examples taken from: S. Gruber and MJ van der Laan. *tmle*: An R Package for Targeted Maximum Likelihood Estimation. Journal of Statistical Software 2012; 51(13)
- Practical Considerations
  - Setting bounds on Q
  - Truncation level for g
  - Examining results
    - Summary
    - Obtaining untargeted parameter estimates