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

Mark van der Laan[1] and Susan Gruber[2]

[1] Department of Biostatistics, University of California, Berkeley School of Public Health

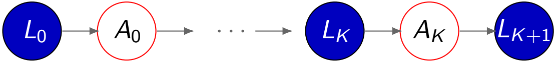
[2] Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute

Course Outline

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

TMLE for Longitudinal Data Analysis

- Goal: Assess Impact of Treament at Multiple Timepoints
- Longitudinal Data (K time points)



Covariate and Outcome nodes $(L_0, \dots, L_{K+1})$

Intervention nodes $(A_0, \dots, A_K)$ indicate treatment and censoring

Challenges to Analyzing Longitudinal Data

- Common default approach
  - View as time-to-event data
  - Impose a Cox Proportional Hazards Model
- However
  - Hazard ratio may not be the most relevant target parameter
  - Cox model is misspecified
  - Cox PH model ignores informative right censoring
  - Time-dependent Cox model does not appropriately handle time-varying covariates affected by prior treatment
- Longitudinal TMLEs appropriately address these challenges

Statistical Estimation Problem

- Data: *n* i.i.d. copies
  $O = \big(L_0, A_0, \ldots, L_K, A_K, Y = L_{K+1}\big) \sim P_0$
- Statistical Model: $\mathcal{M}$
  Collection of possible probability distributions of O
- Target Parameter: $E(Y_{\bar{a}})$
  Mean outcome under specified intervention $\bar{a} = (a_0, \ldots, a_K)$
- Mapping: $\Psi : \mathcal{M} \rightarrow \mathbb{R}$, such that $\Psi(P_{\bar{a}}) = E(Y_{\bar{a}})$
  $(P_{\bar{a}})$ is post-intervention distribution identified by G-computation formula when causal assumptions are met

Note: contrasts (e.g. ATE, RR, RD) are functions of intervention-specific means

Factorization of Likelihood

- Probability distribution P_0 of O factorizes according to time-ordering as

$$
\begin{aligned}
P_0(O) &= \prod_{k=0}^{K+1} P_0\left[L_k \mid Pa(L_k)\right] \prod_{k=0}^{K} P_0\left[A_k \mid Pa(A_k)\right] \\
&\equiv \prod_{k=0}^{K+1} Q_{0,L_k}(O) \prod_{k=0}^{K} g_{0,A_k}(O) \\
&\equiv Q_0 g_0
\end{aligned}
$$

where $Pa(L_k) \equiv (\bar{L}_{k-1}, \bar{Q}_{k-1})$ and $Pa(A_k) \equiv (\bar{L}_k, \bar{A}_{k-1})$ denote parents of L_k and A_k in the time-ordered sequence, respectively

- g_0-factor represents the intervention mechanism, e.g., treatment and right-censoring mechanisms.

G-Computation Formula for Post-Intervention Distribution

- $\bar{a}_K$ is a specific treatment regime of interest
- Consider an intervention that sets $\bar{A}_K = \bar{a}_K$ in the NPSEM
- The post-intervention distribution is given by Robins' G-computation formula

$$P^a(\bar{l}) = \prod_{k=0}^{K+1} Q^a_{L_k}(\bar{l}_k),$$

where $Q^a_{L_k}(\bar{l}_k) = Q_{L_k}\left(l_k \mid \bar{l}_{k-1}, \bar{A}_{k-1} = \bar{a}_{k-1}\right)$.

Statistical Target Parameter

- Let $L^a = \left(L_0, L_1^a, \ldots, Y^a = L_{K+1}^a\right)$ denote the random variable with probability distribution P^a
- Our statistical target parameter is the mean of $Y^a : \Psi(P) = E_{P^a} Y^a$, where $\Psi : \mathcal{M} \to \mathbb{R}$.
  - depends on P only through $Q = Q(P)$.
  - Equivalently denoted by the mapping $\Psi : \mathcal{Q} = \{Q(P) : P \in \mathcal{M}\} \to \mathbb{R}$ so that $\psi_0 = \Psi(Q_0)$.

Alternative target parameters

- Treatment-specific mean $E_{P^d} Y^d$ defined by the G-computation formula for a dynamic treatment d
- Dose-Response
  - Projection of a true dose-response curve $(E_{P^a} Y^a : a \in \mathcal{A})$ onto a working model $\{a \to m_\beta(a) : \beta\}$.
  - Projection of the true dose-response curve $(E_{P^d} Y^d : d \in \mathcal{D})$, $\mathcal{D}$ a collection of dynamic treatment rules, onto a working model $\{d \to m_\beta(d) : \beta\}$.
  - Summary measures of conditional dose-response curves $(E_{P^d}(Y^d | V) : d \in \mathcal{D})$, conditioning on baseline covariates of interest
- Related classes of target parameters defined by history adjusted marginal structural working models for history adjusted conditional treatment-specific means
- Effects of stochastic interventions, intention to treat interventions, etc.

A Sequential Regression G-Computation Formula

- By the iterative conditional expectation rule (tower rule), we have

$$E_{P^a} Y^a = E \dots E \left[E(Y^a \mid \bar{L}^a_K) | L^a_{K-1} \dots \mid L_0 \right].$$

  - Conditional expectation given $\bar{L}^a_K$ is equivalent to conditioning on $\bar{L}_K, \bar{A}_{K-1} = \bar{a}_{K-1}$.
- This yields the sequential regression G-computation formula
  - Compute $\bar{Q}^a_Y = E_{Q^a_Y} Y \equiv E\left(Y \mid \bar{L}_K, \bar{A}_K = \bar{a}_K\right)$
  - Given $\bar{Q}^a_Y$, next compute $\bar{Q}^a_{L_K} = E_{Q^a_{L_K}}\left(\bar{Q}^a_Y \mid \bar{L}_{K-1}, \bar{A}_{K-1} = \bar{a}_{K-1}\right)$
  - Iterate over all time points
    - given $\bar{Q}^a_{L_{k+1}}$, compute $\bar{Q}^a_{L_k} = E_{Q^a_{L_{k+1}}}\left(\bar{Q}^a_{L_{k+1}} \mid \bar{L}_{k-1}, \bar{A}_{k-1} = \bar{a}_{k-1}\right)$
    - Until final step, $\bar{Q}^a_{L_0} = E_{Q_{L_0}} \bar{Q}^a_{L_1}$.

TMLE for an Intervention-Specific Mean

- Iterated conditional expectations approach *(Bang and Robins, 2005)*

$$E(Y_{\bar{a}}) = \underbrace{E\Big(\underbrace{E\Big\{\dots \underbrace{E[\underbrace{E\{Y_{\bar{a}} \mid \bar{L}_K, \bar{A}_K = \bar{a}_K\}}_{\bar{Q}^a_{L(K+1)}} \mid \bar{L}_{K-1}, \bar{A}_{K-1} = \bar{a}_{K-1}]}_{\bar{Q}^a_{L(K)}} \dots \mid L_0\Big\}}_{\bar{Q}^a_{L(1)}}\Big)}_{\bar{Q}^a_{L(0)} = \Psi(Q)}$$

- TMLE target parameter mapping: target parameter is function of iteratively defined sequence of conditional means, $\Psi(\bar{Q}^a)$

$$\bar{Q}^a = \left(\bar{Q}^a_Y, \bar{Q}^a_{L(K)}, \dots, \bar{Q}^a_{L(0)}\right)$$

Efficient Influence Curve of Target Parameter

- Efficient influence curve representation as sum of iteratively defined scores of iteratively defined conditional means

$$D^* = \sum_{k=0}^{K+1} D_k^*$$

where

$$D_{K+1}^* = \frac{I(\bar{A}_K = \bar{a}_K)}{g_{0:K}} \left(Y - \bar{Q}_{K+1}^a\right),$$

and

$$\begin{aligned} D_k^* &= \frac{I(\bar{A}_{k-1} = \bar{a}_{k-1})}{g_{0:k-1}} \left(\bar{Q}_{L_{k+1}}^a - E_{Q_{L_k}^a} \bar{Q}_{L_{k+1}}^a\right), \\ &= \frac{I(\bar{A}_{k-1} = \bar{a}_{k-1})}{g_{0:k-1}} \left(\bar{Q}_{L_{k+1}}^a - \bar{Q}_{L_k}^a\right), k = K, \ldots, 1, \end{aligned}$$

and

$$D_0^* = \bar{Q}_{L_1}^a - E_{L_0} \bar{Q}_{L_1}^a = \bar{Q}_{L_1}^a - \Psi(\bar{Q}^a).$$

$g_{0:K} = \prod_{k=1}^{K} g_k$

L-TMLE Definition

- Initial estimate of $\bar{Q}^a_{L_k}$ (Assume $Y \in [0,1]$)

  super learning, parametric regression, etc.

- Submodel and Loss Function (e.g., negative log likelihood)

$$logit\,\bar{Q}^{a,*}_{L_k}(\epsilon_k, g) = logit\,\bar{Q}^a_{L_k} + \epsilon_k \frac{1}{g_{0:k-1}}, k = K+1, \ldots, 0$$

$$\mathcal{L}_{k,\bar{Q}^a_{L_{k+1}},g}(\bar{Q}^a_{L_k}) =$$

$$-\frac{I(\bar{A}_{k-1} = \bar{a}_{k-1})}{g_{0:k-1}} \left\{ \bar{Q}^a_{L_{k+1}} \log \bar{Q}^a_{L_k} + (1 - \bar{Q}^a_{L_{k+1}}) \log\{1 - \bar{Q}^a_{L_k}\} \right\}$$

- Mapping

  $\Psi(\bar{Q}^{a,*}_n) = \bar{Q}^{a,*}_{L_0,n} = \frac{1}{n} \sum_{i=1}^n \bar{Q}^{a,*}_{1,n}(L_{0_i})$

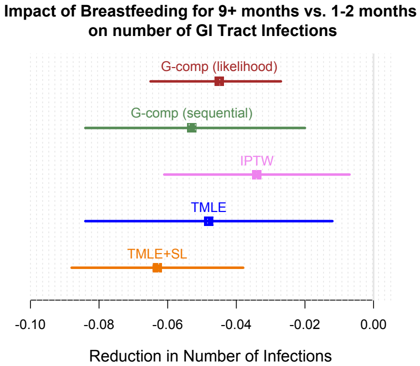
L-TMLE

- Known bounds (e.g. rare outcome) on conditional means can be respected using logistic link.
- TMLE is double robust and asymptotically efficient if both g_0 and the conditional means $\bar{Q}^a_{L_k}$ are consistently estimated.
- Statistical inference can be based on efficient influence curve: conservative as long as g_0 is estimated well.

Example: PROBIT Study Re-Analysis

Investigate the impact of increasing duration of breastfeeding on number of GI tract infections in 1st year of life*

- Breastfeeding at time t impacts infection at time $t+1$, which impacts decision to continue breastfeeding at $t+2$
- TMLE + SL estimates largest effect, with variance close to that of efficient parametric G-computation estimator



* Schnitzer, et al, 2014

Extension to Marginal Structural Models (simplified)

- Impose a MSM to smooth over areas where there is little support in the data
- Consider a working logistic MSM, $logit m_\beta(d,t) = \beta_1 + \beta_2 t + \beta_3 f(d,t)$
- Define target parameter as

$$\psi_0 = \underset{\beta}{\text{argmin}} \; -E_0 \sum_{t \in \tau} \sum_{d \in D} \{Y_d(t) log \; m_\beta(d,t) + (1 - Y_d(t)) log(1 - m_\beta(d,t))\}.$$

- See Petersen, et al (2013), *ltmle* package on CRAN

Stratified TMLE for Longitudinal MSM Parameter

- Estimand $\psi_0 = \beta$ solves the equation

$$0 = E_0 \sum_{t \in \tau} \sum_{d \in D} \frac{\frac{d}{d\beta} m_\beta(d,t)}{m_\beta(1-m_\beta)} \left(E_0(Y^d(t) \mid L_0) - m_\beta(d,t) \right).$$

- Estimate $\bar{Q}_{L_0}^{d,t*}$, for each time point, t, and rule $d \in D$ using targeted iterated conditional expectations approach
- Finally, stack $\bar{Q}_{L_0}^{d,t*}$, and regress onto appropriate covariates in the model $(1, t, f(d,t))$

Notes on Targeting

- Multi-dimensional target parameter requires multi-dimensional fluctuation at each step, $\epsilon_k = (\epsilon_{1_k}, \epsilon_{2_k}, \epsilon_{3_k})$

$$\bar{Q}_{L_k}^{d*} = \bar{Q}_{L_k}^{d} + \epsilon_k \frac{h_1(d,t)}{g_{0:k-1}},$$

with $h_1(d,t) = \frac{\frac{d}{d\beta} m_\beta(d,t)}{m_\beta(1-m_\beta)}$

- Fit ϵ using observations where $\bar{A}_{k-1} = \bar{a}_{k-1}$

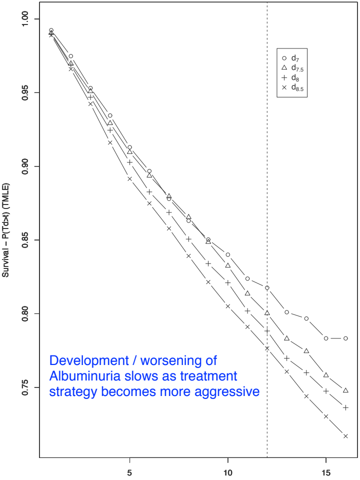
Pooled TMLE for Longitudinal MSM Parameter

- *Pooling across rules* refers to using stacking the datasets for all rules to estimate a single (multi-dimensional) ϵ_k common across all rules.
- The resulting dataset has $n \times |D|$ rows
- Pooling helps if there is sufficient support for some rules but not others
- This simultaneous targeting across all rules still solves the efficient influence curve equation $P_n D^* = 0$.
- An alternative pooled TMLE pools also over the time points t at final outcome $Y(t)$ at the targeting step. Initial estimates of $\bar{Q}^{t,d}_{L_k}$ are obtained for all k from 0 to t, across d and t, and then targeted simultaneously with a common ϵ. Updates are iterated until convergence. The dataset has $n \times |D| \times K + 1$ observations.

Example: Progression of Albuminuria in Type-2 Diabetics

- Lowering glucose levels known to prevent or slow development of Albuminuria
- Best glucose-lowering strategy is not known
- Four candidate strategies, d_θ,
  intensify treatment when patient's A1c level reaches $\theta = 7\%, 7.5\%, 8\%, \text{ or } 8.5\%$
- HMO Research Network EHR data
  7 sites, $n = 51,179$
- Longitudinal TMLE was used to evaluate a set of increasingly agressive dynamic strategies for lowering glucose levels

Counterfactual Survival Curves
(L-TMLE + SL)



Quarter since study entry

Neugebauer, Schmittdiel, van der Laan, 2014

Concluding Remarks

- TMLE provides a template for construction of efficient substitution estimators
- Three basic requirements
  - Loss function
  - Submodel for fluctuation so that its loss-based score spans the efficient influence curve
  - Procedure for iteratively minimizing the empirical risk along the fluctuation model through a current estimator

Concluding Remarks

- All TMLEs are double robust and efficient, but may have different finite sample performance
- Sequential regression
  - particularly effective representation of post-intervention distribution, and thereby causal effects
  - Estimate only smallest portion of Q needed for evaluating the parameter

Core Concepts in Targeted Learning

- Translate a scientific question and background knowledge into a formal causal model, target causal quantity, statistical model and statistical target parameter
- Target statistical parameter has a causal interpretation when assumptions are met, variable importance otherwise
- SL + TMLE for estimation
  - Optimal bias/variance trade-off for target parameter
  - Loss-based estimation using cross-validation
  - Flexible fitting of relevant components of the likelihood
  - Double robust to mitigate misspecification bias
  - Substitution estimator that respects domain knowledge

Selected Resources

- http://www.targetedlearningbook.com
- M.J. van der Laan and S. Rose. *Targeted Learning: Prediction and Causal Inference for Observational and Experimental Data*. Springer, New York, 2011.
- M.J. van der Laan and R. Starmans. Entering the era of data science: Targeted learning and the integration of statistics and computational data analysis. Adv Stat. 2014;2014:1Ð19.
- S. Gruber and MJ van der Laan. tmle: An R Package for Targeted Maximum Likelihood Estimation. *Journal of Statistical Software* 2012; 51(13).
- M.J. van der Laan, E. Polley, and A. Hubbard. Super learner. *Statistical Applications in Genetics and Molecular Biology*, 6(25), 2007. ISSN 1.
- E.C. Polley and M.J. van der Laan. Super Learner in Prediction. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 266 (2010).

Software

- E.C. Polley, SuperLearner: Super Learner in Prediction, v2.0-19, `http://cran.r-project.org/web/packages/SuperLearner`, 2016.
- S. Gruber. tmle. R package version 1.2.0-4, `http://CRAN.R-project.org/package=tmle`, 2014.