Targeted Learning: Application to Optimal Dynamic Treatments

Mark van der Laan
UC Berkeley

October 28, 2013, Data, Society and Inference, Stanford Business School

Outline

Outline

The statistical estimation problem

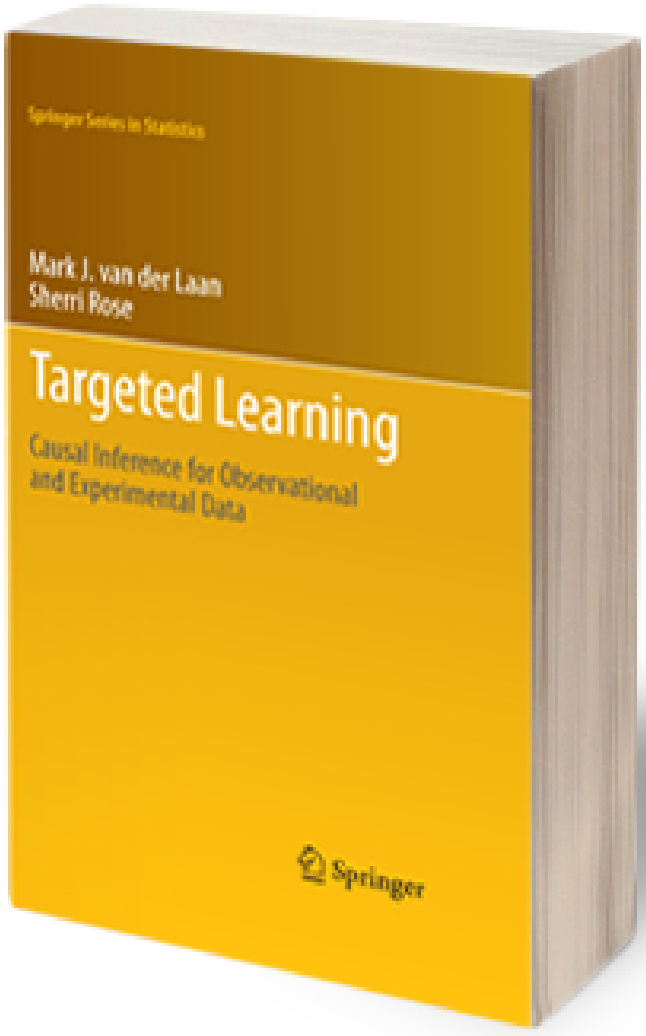
- **Observed data**: Realizations of random variables with a probability distribution.
- **Statistical model**: Set of possible distributions for the data-generating distribution, defined by actual knowledge about the data. e.g. in an RCT, we know the probability of each subject receiving treatment.
- **Statistical target parameter**: Function of the data-generating distribution that we wish to learn from the data.
- **Estimator**: An a priori-specified algorithm that takes the observed data and returns an estimate of the target parameter. Benchmarked by a dissimilarity-measure (e.g., MSE) w.r.t target parameter.

Example

- **Observed data**: n i.i.d. copies of $O = (W, A, Y) \sim P_0$.
- **Statistical model**: Nonparametric model.
- **Statistical target parameter**:
  $\Psi(P) = E_P\{E_P(Y \mid A = 1, W) - E_P(Y \mid A = 0, W)\}$. Only depends on P through $\bar{Q}(P) = E_P(Y \mid A, W)$ and $Q_W(P)$: So we can write $\Psi(Q)$, where $Q = (Q_W, \bar{Q})$.
- **Estimator**: E.g., plug-in estimator:
  $\psi_n = \Psi(Q_n) = \frac{1}{n}\sum_{i=1}^{n}\{\bar{Q}_n(1, W_i) - \bar{Q}_n(0, W_i)\}$.

Targeted learning

- Define valid (and thus LARGE) statistical semi parametric models and interesting target parameters
- Avoid reliance on human art and nonrealistic parametric models
- Plug-in estimator based on targeted fit of the (relevant part of) data-generating distribution to the parameter of interest
- Semiparametric efficient and robust
- Statistical inference
- Has been applied to: static or dynamic treatments, direct and indirect effects, parameters of MSMs, variable importance analysis in genomics, longitudinal/repeated measures data with time-dependent confounding, censoring/missingness, case-control studies, RCTs, dependent network data, etc.



Targeted Learning Book

Springer Series in Statistics

van der laan & Rose

targetedlearningbook.com

Outline

Two stage methodology

- Super learning (SL) van der Laan et al. (2007),Polley et al. (2012),Polley and van der Laan (2012)
  - Uses a library of candidate estimators (e.g. multiple parametric models, machine learning algorithms like neural networks, RandomForest, etc.)
  - Builds data-adaptive weighted combination of estimators using cross validation
- Targeted (maximum likelihood/minimum loss) estimation (TMLE) van der Laan and Rubin (2006)
  - Updates initial estimate, often a Super Learner, to remove bias for the parameter of interest
  - Update based on parametric submodel through initial estimate with score equal to efficient score/efficient influence curve.
  - Calculates final parameter from updated fit of the data-generating distribution

Super learning

- Generalization of stacking.
- No need to chose a priori a particular parametric model or machine learning algorithm for a particular problem
- Allows one to combine many data-adaptive estimators into one improved estimator.
- Grounded by oracle results for cross-validation as estimator selection (Van Der Laan and Dudoit (2003),van der Vaart et al. (2006)). Loss function needs to be bounded.
- Performs asymptotically as well as best (oracle) weighted combination, or achieves parametric rate of convergence.

Super learning: Simulation example

Figure: Relative Cross-Validated Mean Squared Error (compared to main terms least squares regression)

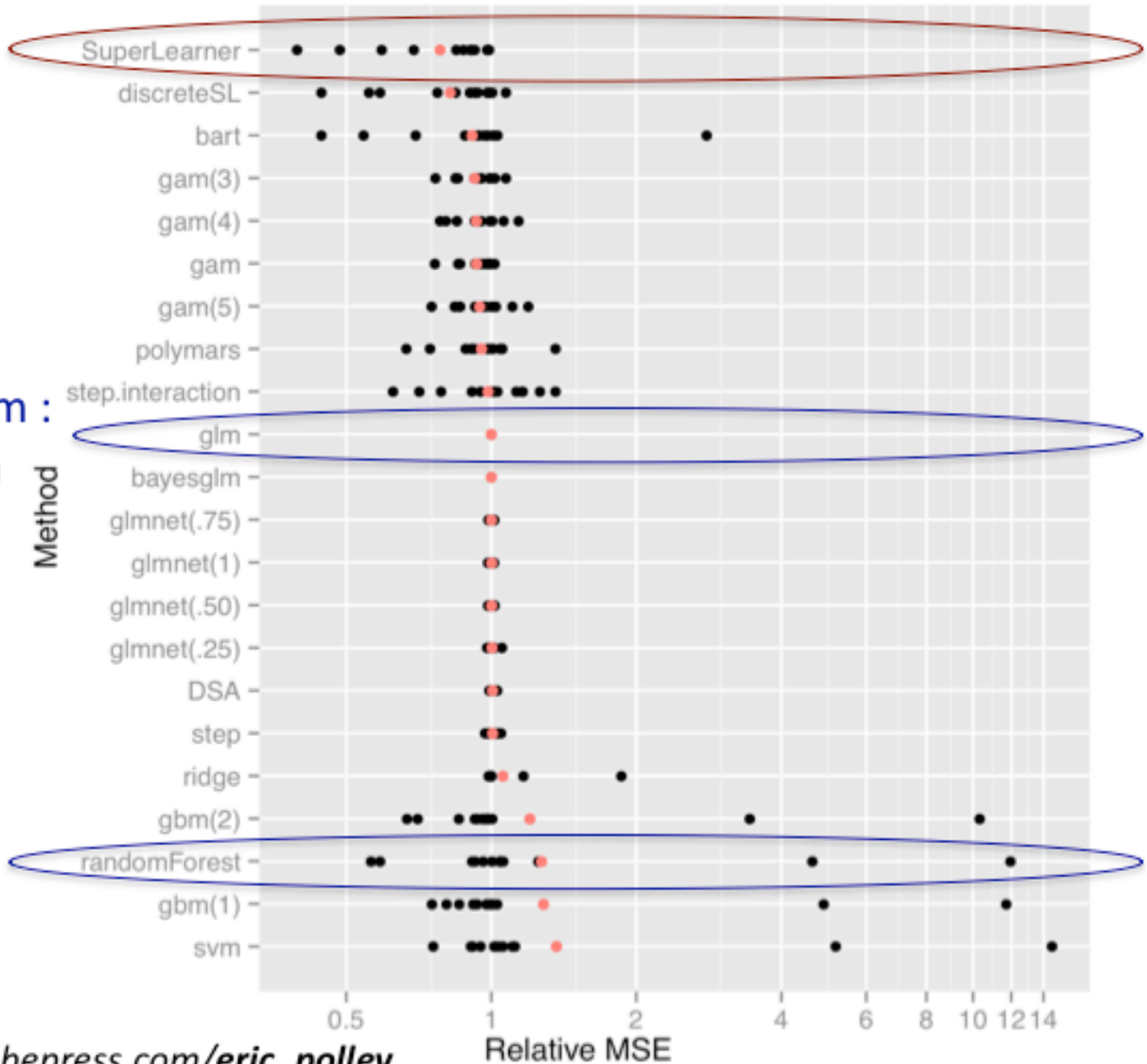
Method	Study 1	Study 2	Study 3	Study 4	Overall
Least Squares	1.00	1.00	1.00	1.00	1.00
LARS	0.91	0.95	1.00	0.91	0.95
D/S/A	0.22	0.95	1.04	0.43	0.71
Ridge	0.96	0.9	1.02	0.98	1.00
Random Forest	0.39	0.72	1.18	0.71	0.91
MARS	0.02	0.82	0.17	0.61	0.38
Super Learner	0.02	0.67	0.16	0.22	0.19

Super learning: Data example

Super Learner-
Best weighted combination of algorithms for a given prediction problem

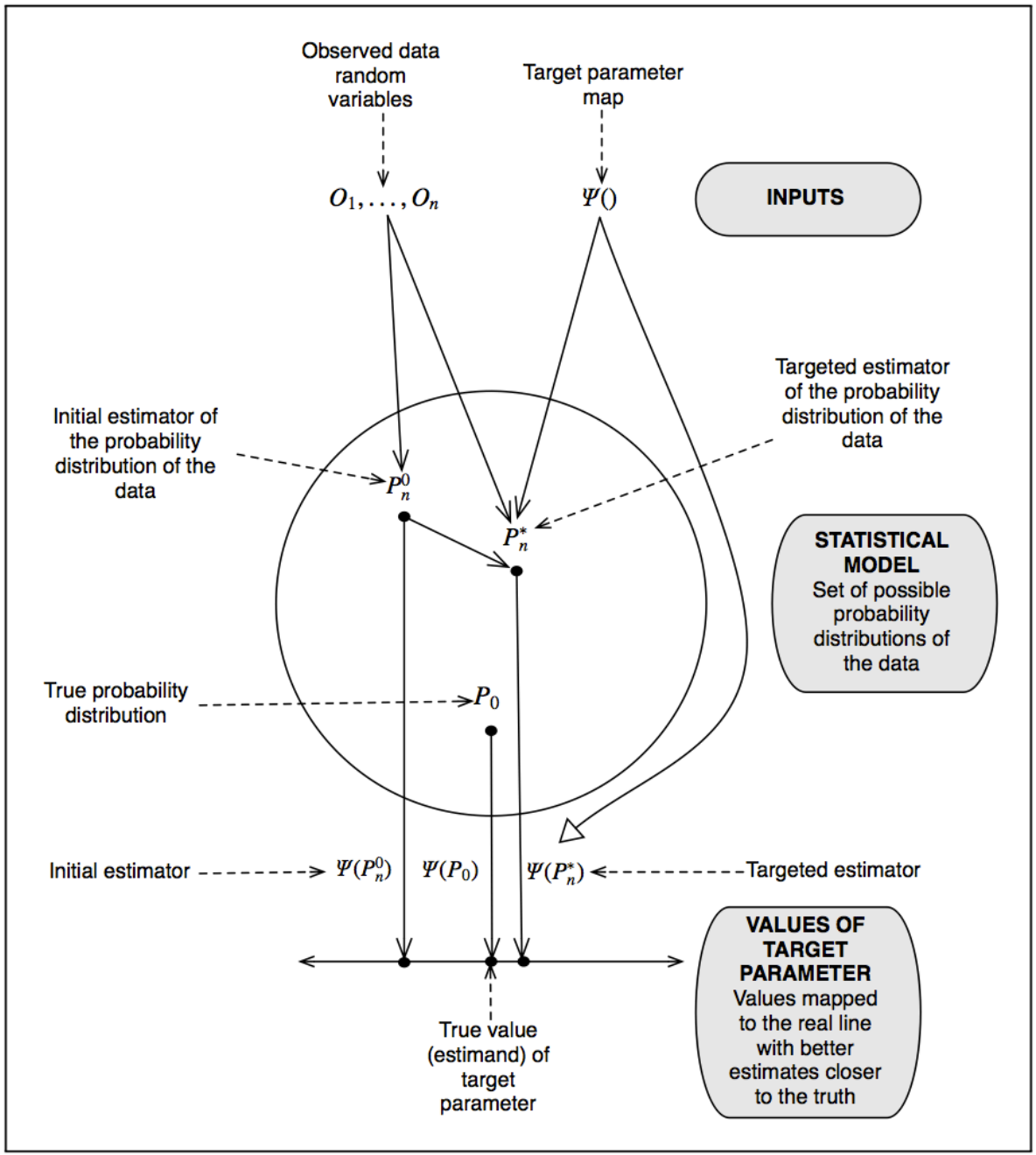
Example algorithm : Linear Main Term Regression

Example algorithm: Random Forest



Technical Report: *works.bepress.com/**eric_polley***

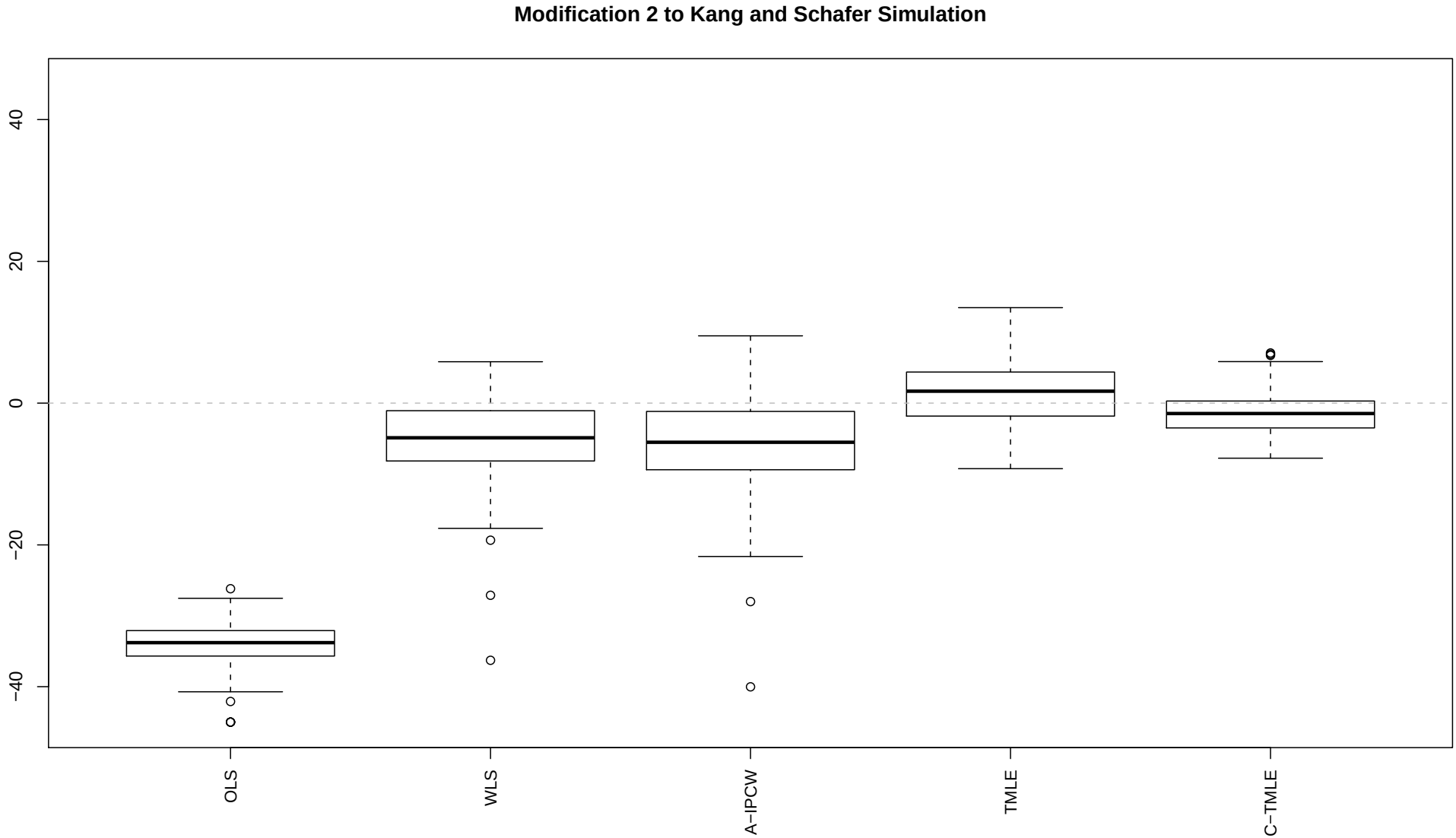
TMLE algorithm



TMLE algorithm: Formal Template

$\Psi(Q_0)$ 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: $\Psi(\hat{Q}^*)$

Simulations



Outline

General Longitudinal Data Structure

We observe n i.i.d. copies of a longitudinal data structure

$$O = (L(0), A(0), \ldots, L(K), A(K), Y = L(K+1)),$$

where $A(t)$ denotes a discrete valued intervention node, $L(t)$ is an intermediate covariate realized after $A(t-1)$ and before $A(t)$, $t = 0, \ldots, K$, and Y is a final outcome of interest.

For example, $A(t) = (A_1(t), A_2(t))$ could be a vector of two binary indicators of censoring and treatment, respectively.

Likelihood and Statistical Model

The probability distribution P_0 of O can be factorized according to the time-ordering as

$$\begin{aligned} P_0(O) &= \prod_{t=0}^{K+1} P_0(L(t) \mid Pa(L(t))) \prod_{t=0}^{K} P_0(A(t) \mid Pa(A(t))) \\ &\equiv \prod_{t=0}^{K+1} Q_{0,L(t)}(O) \prod_{t=0}^{K} g_{0,A(t)}(O) \\ &\equiv Q_0 g_0, \end{aligned}$$

where $Pa(L(t)) \equiv (\bar{L}(t-1), \bar{A}(t-1))$ and $Pa(A(t)) \equiv (\bar{L}(t), \bar{A}(t-1))$ denote the parents of $L(t)$ and $A(t)$ in the time-ordered sequence, respectively. The g_0-factor represents the intervention mechanism: e..g, treatment and right-censoring mechanism.

Statistical Model: We make no assumptions on Q_0, but could make assumptions on g_0.

Statistical Target Parameter: G-computation Formula for Post-Intervention Distribution

- Let

$$P^d(l) = \prod_{t=0}^{K+1} Q^d_{L(t)}(\bar{l}(t)), \tag{1}$$

  where $Q^d_{L(t)}(\bar{l}(t)) = Q_{L(t)}(l(t) \mid \bar{l}(t-1), \bar{A}(t-1) = \bar{d}(t-1))$.
- Let $L^d = (L(0), L^d(1), \ldots, Y^d = L^d(K+1))$ denote the random variable with probability distribution P^d.
- This is the so called G-computation formula for the post-intervention distribution corresponding with the dynamic intervention d.

Outline

A Sequential Regression G-computation Formula (Bang, Robins, 2005)

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

$$E_{P^d} Y^d = E \ldots E(E(Y^d \mid \bar{L}^d(K)) \mid L^d(K-1)) \ldots \mid L(0)).$$

- In addition, the conditional expectation, given $\bar{L}^d(K)$ is equivalent with conditioning on $\bar{L}(K), \bar{A}(K-1) = \bar{d}(K-1)$.

In this manner, one can represent $E_{P^d} Y^d$ as an iterative conditional expectation, first take conditional expectation, given $\bar{L}^d(K)$ (equivalent with $\bar{L}(K), \bar{A}(K-1)$), then take the conditional expectation, given $\bar{L}^d(K-1)$ (equivalent with $\bar{L}(K-1), \bar{A}(K-2)$), and so on, until the conditional expectation given $L(0)$, and finally take the mean over $L(0)$.

Outline

Marginal structural working models

We can define a target parameter as projection of the true dose-response curve $(E_{P^d} Y^d : d \in \mathcal{D})$ onto a working model $\{d \to m_\beta(d) : \beta\}$. For example, if $Y(t) \in [0, 1]$, we can define

$$\Psi(P_0)) = \arg\max \beta$$
$$E_0 \sum_t \sum_{d \in \mathcal{D}} h(d, t, V) \Big\{ E_0(Y^d(t)) \log m_\beta(d, t) + (1 - E_0(Y^d(t)) \log(1 - m_\beta(d, t)) \Big\}$$

Outline

ltmle package (Petersen et al. (2013), van der Laan, Gruber (2012), Schitzer et al. (2013))

R package (April 2013): `ltmle`

- Causal effect estimation with multiple intervention nodes
  - Intervention-specific mean under longitudinal static and dynamic interventions
  - Static and dynamic marginal structural models
- General longitudinal data structures
  - Repeated measures outcomes
  - Right censoring
- Estimators
  - IPTW
  - Non-targeted MLE
  - TMLE (two algorithms for MSM)
- Options include nuisance parameter estimation via glm regression formulas or calling `SuperLearner()`

Outline

Optimal dynamic treatment

Let $\mathcal{D}$ be the class of all dynamic treatments for which the rule for assigning $A(k)$ only uses $(\bar{V}(k) \subset \bar{L}(k), \bar{A}(k-1))$, $k = 0, \ldots, K$. Define the optimal rule

$$d_0 = \arg\max_{d \in \mathcal{D}} E_{P_0} Y_d.$$

This optimal rule is given by $d_{0,k} = I(\bar{Q}_{0,k}^{\underline{d}_{0,k+1}} > 0)$ in terms of the iteratively defined blip functions (starting with $k = K$):

$$\bar{Q}_{0,k}^{\underline{d}_{0,k+1}} = E_0(Y_{\bar{a}(k-1),A(k)=1,\underline{d}_{0,k+1}} - Y_{\bar{a}(k-1),A(k)=0,\underline{d}_{0,k+1}} \mid V_{\bar{a}}(k)),$$

where $\underline{d}_{k+1} = (d_j : j = k+1, \ldots, K)$. These blip-functions are modeled with parametric models in Murphy (2003), Robins (2003, 2004) to yield their structural nested mean models for optimal dynamic treatments.

Sequential Super Learning of blip functions

For simplicity, consider case $L(0), A(0), L(1), A(1), Y$ (i.e., $K = 1$).

- We need to construct a data adaptive estimator of $\bar{Q}_{02}(a(0), v(1)) = E_{P_0}(Y_{a(0)1} - Y_{a(0)0} \mid V_{a(0)}(1) = v(1))$ and, given the resulting estimator $d_{n,A(1)}$ of $d_{0,A(1)}$, we subsequently need to construct a data adaptive estimator of $\bar{Q}_{01,d}(v(0)) = E_{P_0}(Y_{1d_{A(1)}} - Y_{0d_{A(1)}} \mid V(0) = v(0))$ for a given $d_{A(1)} = d_{n,A(1)}$.
- For that purpose we propose to use sequential loss-based super-learning defined by the application of two subsequent super-learners.

Example of loss function: IPCW-loss functions

For the super-learner of $\bar{Q}_{20}$ we can use the following loss function:

$$L_{2,g_0}(\bar{Q}_2)(O) = \sum_{a(0)} \frac{I(A(0)=a(0))}{g_{0,A(0)}(O)} \left(D(g_0)(O) - \bar{Q}_2(A(0), V(1)\right)^2,$$

where

$$D(g_0)(O) = I(A_2(1) = 1)\frac{2A_1(1) - 1}{g_{0,A(1)}(O)} Y.$$

Given, the fitted rule of $d_{0,A(1)}$, for the super-learner of $\bar{Q}^d_{10}$, we can use the loss function

$$L_{1,d,g_0}(\bar{Q}^d_1)(O) = \frac{I(A(1)=d_{A(1)}(V(1))}{g_{0,A(1)}(O)}(D(g_0)(O) - \bar{Q}^d_1(V(0)))^2,$$

where

$$D(g_0)(O) = I(A_2(0) = 1)\frac{2A_1(0) - 1}{g_{0,A(0)}(O)} Y.$$

Super-learning based on performance of candidate rule

- Given a collection of candidate estimators $\hat{d}_\alpha(P_n)$ of d_0, we can also select α with a maximizer of a cross-validated estimator of the conditional risk $\alpha \rightarrow E_{B_n} E_{P_0} Y_{\hat{d}_\alpha(P^0_{n,B_n})}$, where $B_n \in \{0,1\}^n$ is sample split-vector, P^0_{n,B_n} is empirical distribution of training sample $\{i : B_n(i) = 0\}$. P^1_{n,B_n} denotes the empirical distribution of validation sample.
- For example, we could use the cross-validated empirical mean $E_{B_n} P^1_{n,B_n} L_{g_0}(\hat{d}_\alpha(P^0_{n,B_n})$ of the IPCW-loss

$$L_{g_0}(d) = \frac{I(\bar{A} = d(V))}{g_0(O)} Y,$$

  or the DR-IPCW loss L_{g_0,Q_0}.
- Alternatively, we can estimate this conditional risk $E_{B_n} E_{P_0} Y_{\hat{d}_\alpha(P^0_{n,B_n})}$ with a cross-validated-TMLE.

Statistical Inference for mean outcome under optimal rule

- The pathwise derivative of this target parameter $E_0 Y_{d_0}$ equals the pathwise derivative of the mean counterfactual outcome $E_0 Y_d$ under a given dynamic treatment rule d set at the optimal rule d_0, treating the latter as known!
- Thus, we apply the targeted minimum loss-based estimator for the mean outcome under a given rule, but set the given rule equal to our data adaptive estimator of the optimal rule.
- This TMLE of $E_0 Y_{d_0}$ is asymptotically linear, allowing us to construct confidence intervals for the mean outcome under the optimal dynamic treatment or its contrast $E_0(Y_{d_0} - Y_0)$ w.r.t. a standard treatment.
- In a SMART the statistical inference would *only* rely upon a second order difference between the estimator of the optimal dynamic treatment and the optimal dynamic treatment itself to be asymptotically negligible.

Statistical Inference for Mean outcome under fitted optimal rule

- The same targeted minimum loss based estimator can be viewed as an estimator of the data adaptive target parameter $E_0 Y_d|_{d=d_n}$ defined as the mean outcome under the *estimate* of the optimal dynamic treatment.
- In particular, we develop a cross-validated TMLE that provides asymptotic inference for $E_{B_n} E_0 Y_{\hat{d}(P^0_{n,B_n})}$ under minimal conditions.

Concluding remarks

- Thinking in terms of parametric models avoids natural results and progress.
- (Different versions of) Super-Learning of Optimal dynamic treatment based on sequentially randomized trials or observational studies is an exciting future area.
- It can be combined with statistical inference for the mean outcome.
- Statistical package development.

References I

H. Bang and J. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:962–972, 2005.

M. Petersen, J. Schwab, S. Gruber, N. Blaser, M. Schomaker, and M. van der Laan. Targeted minimum loss based estimation of marginal structural working models. *Journal of Causal Inference*, submitted, 2013.

E. Polley and M. van der Laan. *SuperLearner: Super Learner Prediction*, 2012. URL `http://CRAN.R-project.org/package=SuperLearner`. R package version 2.0-6.

E. Polley, S. Rose, and M. van der Laan. Super learning. *Chapter 3 in Targeted Learning, van der Laan,Rose (2012)*, 2012.

References II

M. Schnitzer, E. Moodie, M. van der Laan, R. Platt, and M. Klein. Marginal structural modeling of a survival outcome with targeted maximum likelihood estimation. *Biometrics*, to appear, 2013.

M. Van Der Laan and S. Dudoit. Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples. *UC Berkeley Division of Biostatistics Working Paper Series*, page 130, 2003.

M. J. van der Laan and S. Gruber. Targeted Minimum Loss Based Estimation of Causal Effects of Multiple Time Point Interventions. *The International Journal of Biostatistics*, 8(1), Jan. 2012. ISSN 1557-4679. doi: $10.1515/1557\text{-}4679.1370$.

References III

M. J. van der Laan and D. Rubin. Targeted Maximum Likelihood Learning. *The International Journal of Biostatistics*, 2(1), Jan. 2006. ISSN 1557-4679. doi: $10.2202/1557\text{-}4679.1043$.

M. J. van der Laan, E. C. Polley, and A. E. Hubbard. Super learner. *Statistical applications in genetics and molecular biology*, 6(1), Jan. 2007. ISSN 1544-6115. doi: $10.2202/1544\text{-}6115.1309$.

A. van der Vaart, S. Dudoit, and M. van der Laan. Oracle inequalities for multi-fold cross-validation. *Stat Decis*, 24(3): 351–371, 2006.

Targeted Learning of Causal Effect on Network of Units

Structural Equation Model for Interconnected Units

$$
\begin{aligned}
W_i &= L_i(0) = f_{W_i}(U_{W_i}) \\
A_i &= A_i(0) = f_A(C_i^A, U_{A_i}) \\
Y_i &= L_i(1) = f_Y(C_i^Y, U_{Y_i}) \\
i &= 1, \ldots, N,
\end{aligned}
$$

where $C_i^A = c_i^A(W) \in \mathbb{R}^{d_1}$ is determined by $W = (W_1, \ldots, W_N)$, and $C_i^Y = c_i^Y(W, A) \in \mathbb{R}^{d_2}$ is determined by W, A with $A = (A_1, \ldots, A_N)$. Important case is that F_i is set of "friends" of unit i, and

$$
\begin{aligned}
c_i^A(W) &= (W_j : j \in F_i) \\
c_i^Y(A, W) &= ((W_i, A_i), (W_j, A_j : j \in F_i)).
\end{aligned}
$$

It is assumed that U_{W_i} are independent, and, conditional on W, (U_{A_i}, U_{Y_i}) are independent (RA), and iid across i.

Causal Quantity Defined by Stochastic Intervention

- Let g^* be a conditional distribution of A, given W.
- Our goal is to estimate the mean of the counterfactual outcome of $Y^c = 1/N \sum_{i=1}^N Y_i$ under the stochastic intervention g^*. Let $Y_{g^*} = (Y_{g^*,i} : i = 1, \ldots, N)$ be the counterfactual indexed by a stochastic intervention g^* on A, and $Y^c_{g^*} = 1/N \sum_{i=1}^N Y_{g^*,i}$.
- The causal quantity of interest is defined as

$$\Psi^F(P_{U,W,A,Y}) = EY^c_{g^*},$$

which is a parameter of the distribution of (U, W, A, Y).

Observed Data and Likelihood

We observe $O = (O_1, \ldots, O_N)$, where $O_i = (W_i, A_i, Y_i)$. Due to the above structural assumptions, the probability distribution of O is given by:

$$P(O) = \prod_{i=1}^{N} P_{W_i}(W_i) P_{A|C^A}(A_i \mid C_i^A) P_{Y|C^Y}(Y_i \mid C_i^Y), \qquad (2)$$

where $P_{A|c^A}(\cdot \mid c^A)$ is a common (in i) density for A for each c^A, and $P_{Y|C^Y}(\cdot \mid c^Y)$ is a common density for Y for each c^Y.

Identifiability: G-Computation Formula

Since, by assumption, $A = (A_1, \ldots, A_N)$ is independent of $U_Y = (U_{Y_i} : i = 1, \ldots, N)$, given $W = (W_1, \ldots, W_N)$, the post-intervention probability distribution P_{g^*} of $(W, Y_{g^*}) = (W_i, Y_{i,g^*} : i = 1, \ldots, N)$ is identified by the following G-computation formula applied to the probability distribution P of O:

$$\begin{aligned} P_{g^*}(W, A^*, Y) &= \prod_{i=1}^{N} P_{W_i}(W_i) P_{Y|C^Y}(Y_i \mid C_i^{Y,*}) g_i^*(A_i^* \mid C_i^{A,*}) \\ &\equiv P^{g^*}(W, A^*, Y), \end{aligned}$$

where $C_i^{Y,*} = c_i^Y(A^*, W)$.

Statistical Estimation Problem

- Let $\mathcal{M}$ be the statistical model for the data distribution P in which P_{W_i}, $i = 1, \ldots, N$, and the common $P_{A|C^A}$, $P_{Y|C^Y}$ are unspecified.
- Let the statistical target parameter mapping $\Psi : \mathcal{M} \to \mathbb{R}$ be defined as $\Psi(P) = E_{P^{g^*}} Y^{g^*,c}$.
- Under the stated causal model and identifiability assumptions under which $P = P_{P_{U,W,A,Y}}$, we have

$$\Psi(P) = \Psi^F(P_{U,W,A,Y}),$$

  so that $\Psi(P)$ can be interpreted as the desired causal quantity.
- Our goal is to construct an estimator of $\psi_0 = \Psi(P_0)$ based on $O = (O_1, \ldots, O_N) \sim P_0 \in \mathcal{M}$.

Let Q_{W_i} be the marginal distribution of W_i and let $\bar{Q}_Y$ be the common conditional mean of Y_i, given C_i^Y. The target parameter $\Psi(P)$ only depends on P through Q_{W_i}, $i = 1, \ldots, N$, and $\bar{Q}_Y$:

$$\Psi(P) = \Psi(Q) = \frac{1}{N} \sum_{i=1}^{N} \int_{a^*, w} \bar{Q}_Y(C_i^Y(a^*, w)) g^*(a^* \mid w) Q_W(w).$$

TMLE

- Recall the target parameter is given by

$$\begin{aligned}\psi_0 &= E_0 Y^{c,g^*} = \Psi(\bar{Q}_{Y,0}, Q_{W,0}) \\ &= \tfrac{1}{N}\sum_{j=1}^{N} \int_{a,w} \bar{Q}_{Y,0}(c_j^Y(a,w)) g^*(a \mid w) Q_{W,0}(w).\end{aligned}$$

- Let $\bar{Q}_N$ be an estimator of $\bar{Q}_{Y,0}(c) = E_0(Y_i \mid C_i^Y = c)$. Suppose $Y_i \in \{0,1\}$ or continuous in $(0,1)$. This estimator $\bar{Q}_N$ could be based on the log-likelihood loss function

$$-L(\bar{Q}_Y)(O) = \sum_{i=1}^{N} \log \bar{Q}_Y(c_i^Y)^{Y_i}(1 - \bar{Q}_Y(c_i^Y))^{1-Y_i}.$$

- Let $\bar{Q}_{W,N}$ be a nonparametric maximum likelihood estimator of Q_W, thus respecting the model for the joint distribution of $W_1, \ldots, W_N$.
- A plug-in estimator could now be defined as $\Psi(\bar{Q}_N, Q_{W,N})$.

- Let g_N be an estimator of g_0. Given the model assumption $g(A \mid W) = \prod_i \bar{g}(A_i \mid c_i^A(W))$ for a common conditional density $\bar{g}$, this estimator can be based on the log-likelihood loss:

$$L(g)(O) = -\sum_{i=1}^{N} \log \bar{g}(A_i \mid c_i^A).$$

- Given g_N, $Q_{W,N}$, $\bar{Q}_N$, let $\bar{Q}_N(\epsilon)$ be a target-parameter specific submodel through $\bar{Q}_N$ defined by

$$\text{Logit}\bar{Q}_N(\epsilon) = \text{Logit}\bar{Q}_N + \epsilon \frac{\bar{h}(g^*, Q_{W,N})}{\bar{h}(g_N, Q_{W,N})}.$$

- Let
$$\epsilon^N = \arg\min_{\epsilon} L(\bar{Q}_N(\epsilon))(O)$$
be the maximum likelihood estimator, which simply involves running univariate logistic regression on a pooled data set with binary outcomes Y_i and covariate $\frac{\bar{h}(g^*, Q_{W,N})}{\bar{h}(g_N, Q_{W,N})}(c_i^Y)$, using as off-set $\text{Logit}\bar{Q}_N$.
- Here $\bar{h} = \sum_i h_i$, $h_i(c) = P(C_i^Y(A, W) = c)$, and $\bar{h}^* = \sum_i h_i^*$ with $h_i^* = h_i(g^*, Q_W)$.
- This defines now an update $\bar{Q}_N^* = \bar{Q}_N(\epsilon^N)$.
- The TMLE of ψ_0 is defined as the corresponding plug-in estimator
$$\psi_N^* = \Psi(\bar{Q}_N^*, Q_{W,N}).$$

Statistical Inference

This TMLE has been shown to be a double robust asymptotically normally distributed estimator, under the assumption that the number of friends does not grow to fast to infinity with N (**?**).

Comparing Estimators in SEARCH trial simulation