

Targeted Minimum Loss Estimation with the Highly-Adaptive Lasso

Mark van der Laan
Division of Biostatistics, UC Berkeley

June 18, 2019, Leiden University
Workshop in honor of Aad van der Vaart, 60-th

Joint work with Cheng Ju, David Benkeser and Wilson Cai

Thanks to Charles Kooperberg: Aad as Berkeley Soccer Player



Aad as Berkeley Soccer Player



Aad showing off his tanned body



Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Running Example: Nonparametric estimation of treatment specific mean

- We observe n iid observations of $O = (W, A, Y) \sim P_0$, nonparametric model $\mathcal{M}$.
- W vector of baseline covariates, A binary treatment, Y binary outcome.
- Statistical target parameter $\Psi : \mathcal{M} \rightarrow \mathbb{R}$,
 $\Psi(P) = E_P E_P(Y \mid A = 1, W)$.
- Canonical gradient of pathwise derivative of Ψ at P :

$$D^*(P) = \frac{A}{P(A|W)}(Y - E(Y \mid A, W)) + E(Y|A = 1, W) - \Psi(P).$$

- Exact second order remainder
 $R_2(P, P_0) = \Psi(P) - \Psi(P_0) - (P - P_0)D^*(P)$ equals

$$R_2(P, P_0) = P_0(\bar{Q} - \bar{Q}_0)(\bar{G} - \bar{G}_0)/\bar{G},$$

where $\bar{G}(W) = P(A = 1 \mid W)$, $\bar{Q} = E(Y|A = 1, W)$.

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Targeted Minimum Loss Based Estimation (TMLE)

- Super learning provides an initial estimator $\mathbf{P}_n$ of stochastic system P_0 .
- Determine mathematically the fluctuation strategy (least favorable submodel) $\mathbf{P}_{n,\epsilon}$ of the super-learner fit $\mathbf{P}_n$ with tuning parameter ϵ **so that a small change in density corresponds with a maximal small change** in estimated answer $\Psi(\mathbf{P}_{n,\epsilon})$ to query $\Psi(P_0)$: i.e., score spans canonical gradient/**efficient influence curve** $D^*(\mathbf{P}_n)$.
- Determine the optimal amount ϵ_n of fluctuation based on the data (e.g., maximum likelihood estimation).
- The resulting update $\mathbf{P}_n^* = \mathbf{P}_{n,\epsilon_n}$ is the TMLE of P_0 , and it implies the TMLE $\Psi(\mathbf{P}_n^*)$ of target estimand.
- Thanks to TMLE-update, TMLE solves optimal score equation $P_n D^*(\mathbf{P}_n^*) \approx 0$ (or exactly if we use a universal LFM), and is asymptotically normally distributed around true answer to query with minimal asymptotic variance.

Three general methods for efficient estimation in literature

Undersmoothing/overfitting MLE does not always work, depends on sieve, and how to undersmooth is challenge.

Three general methods result in asymptotically efficient estimators, given good initial estimator $\mathbf{P}_n$:

- One-step estimator: $\psi_n^1 = \Psi(\mathbf{P}_n) + P_n D^*(\mathbf{P}_n)$.
- Estimating equation estimator: Assume estimating function representation $D^*(P) = D^*(\psi, \eta(P))$; let ψ_n solution of $P_n D^*(\psi, \eta(\mathbf{P}_n)) = 0$.
- TMLE: $\mathbf{P}_{n,\epsilon}$ least favorable submodel through initial $\mathbf{P}_n$; ϵ_n MLE; $P_n^* = \mathbf{P}_{n,\epsilon_n}$; TMLE is $\Psi(P_n^*)$.
- TMLE is general method that updates initial $\mathbf{P}_n$ into improved fit $\mathbf{P}_n^*$ that solves user supplied set of equations $P_n D(\mathbf{P}_n^*) \approx 0$, allowing for various additional properties beyond asymptotic efficiency.

Each one of the methods has a sample splitting analogue removing Donsker class condition.

Analysis of TMLE

Due to pathwise differentiability of Ψ , we have

$$\Psi(P_n^*) - \Psi(P_0) = -P_0 D^*(P_n^*) + R_2(P_n^*, P_0),$$

for a second order term $R_2()$. Combined with score equation $P_n D^*(P_n^*) = 0$ (or $o_P(n^{-1/2})$), this yields

$$\Psi(P_n^*) - \Psi(P_0) = (P_n - P_0) D^*(P_n^*) + R_2(P_n^*, P_0).$$

If $R_2(P_n^*, P_0) = o_P(n^{-1/2})$ (and Donsker class condition), then

$$\Psi(P_n^*) - \Psi(P_0) = \frac{1}{n} \sum_{i=1}^n D^*(P_0)(O_i) + o_P(n^{-1/2}).$$

That is, ψ_n^* is asymptotically efficient plug-in estimator of ψ_0 .

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Representation of multivariate cadlag function (Gill et al, 94)

- For a bivariate real valued cadlag function $Q : [0, \tau] \rightarrow \mathbb{R}$, we have:

$$\begin{aligned} Q(x_1, x_2) &= Q(0, 0) + \int_{(0, x_1]} Q(du, 0) \\ &\quad + \int_{(0, x_2]} Q(0, dv) + \int_{(0, x_1] \times (0, x_2]} Q(du, dv). \end{aligned}$$

- In general, for a cadlag function $Q : [0, \tau] \subset \mathbb{R}^d \rightarrow \mathbb{R}$ with finite sectional variation norm (and thus its sections generate a finite signed measure), we have

$$Q(x) = Q(0) + \sum_{s \subset \{1, \dots, d\}} \int I(x_s \geq u_s) Q_s(du_s),$$

where $Q_s(u) = Q(u_s, 0_{s^c})$ is the section of Q that sets the coordinates in s equal to zero.

- Here $x_s = (x(j) : j \in s)$ and the sum is over all subsets of $\{1, \dots, d\}$. 

Sectional variation norm (Gill et al., 94)

- We define the sectional variation norm of Q as sum of total variations of s -specific measures generated by Q_s :

$$\|Q\|_v^* = |Q(0)| + \sum_{s \in \{1, \dots, d\}} \int |Q_s(du_s)|.$$

Highly Adaptive Lasso (HAL-MLE)

- $Q_{n,M} = \arg \min_{Q, \|Q\|_V^* < M} P_n L(Q)$ MLE over cadlag with bounded sectional variation norm.
- Q can be arbitrarily well approximated by linear combination of tensor products of indicators $I(W_j > c)$, $j = 1, \dots, d$, basis functions for dense set of cut-offs c .
- The L_1 -norm of coefficient vector represents the sectional variation norm of the function.
- Therefore, HAL-MLE can be fitted with Lasso software (e.g., glmnet), M chosen with cross-validation.
- By being an MLE, it follows that HAL-MLE Q_n converges in loss-based dissimilarity (like square of $L^2(P_0)$ -norm) at a rate faster than $n^{-1/2-\alpha(d)}$ for an $\alpha(d) > 0$ behaving as $1/d$
- Assumption: uniform bound on loss and true function is Cadlag with Finite Sectional Variation.

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

HAL-TMLE is *asymptotically* efficient

- Due to the HAL-super learner converging to true nuisance parameter at faster rate than $n^{-1/4}$, the HAL-TMLE (using HAL or HAL-SL for all nuisance parameters) is efficient in great generality (vdL, 15).
- The only necessary model assumptions are:
 - the true nuisance parameters have finite sectional variation norm;
 - their loss functions are uniformly bounded (so that oracle inequality for cross-validation applies);
 - the strong positivity assumption holds (i.e. uniformly bounded canonical gradient $D^*(P_0)$).

Example: *Asymptotic* efficiency of HAL-TMLE for treatment specific mean

Consider the HAL-TMLE of $EY_1 = EE(Y \mid A = 1, W)$ based on $(W, A, Y) \sim P_0$ in a nonparametric model.

It is asymptotically efficient if

- $\delta < P_0(A = 1 \mid W)$ for some $\delta > 0$;
- $W \rightarrow E_0(Y \mid A = 1, W)$ and $W \rightarrow P_0(A = 1 \mid W)$ are cadlag and have finite sectional variation norm.

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Simulation design

We repeatedly sampled random data generating mechanisms and simulated samples of size $n \in \{100, 500, 1000, 2000\}$ for a total of 25,000 different data generating mechanisms of (W, A, Y) .

We computed TMLEs of the ATE based on different estimators of $E_0(Y \mid A, W)$ and $P_0(A = 1 \mid W)$.

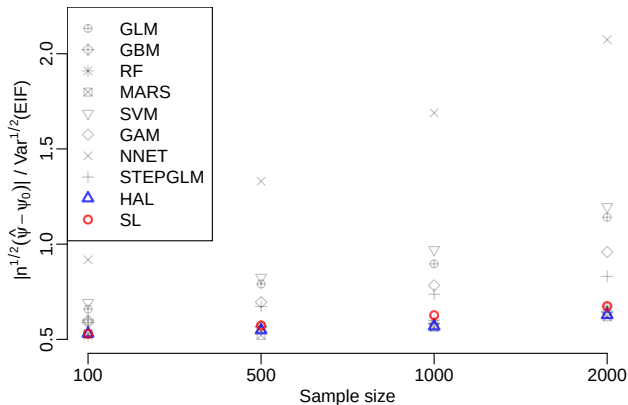
- GLM, Bayes GLM, stepwise GLM (AIC), stepwise GLM (p-value), stepwise GLM with two-way interactions, intercept-only GLM, GAM, GBM*, random forest*, linear SVM*, neural nets*, regression trees*, HAL
- Super Learner (based on these algorithms)
- * = tuning parameters selected via cross-validation

Estimators compared on their absolute error (relative to best achievable SE) and coverage probability of 95% oracle confidence intervals.

Results – absolute error by sample size

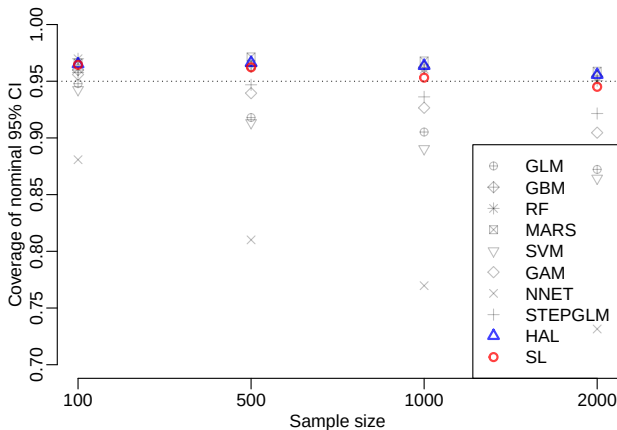
HAL-TMLE exhibited excellent accuracy relative to competitors.

Mean absolute error



Results – coverage by sample size

HAL-TMLE achieves approximate Normality in reasonable sample sizes.



Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient**
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Undersmoothing HAL-MLE is asymptotically efficient

- By being an MLE, the HAL-MLE $Q_{n,M}$ of Q_0 solves a lot of score equations.
- By choosing the L_1 -norm M in HAL-MLE large enough, so that even basis functions with empirical support $o(1/n^{1/2})$ are included, it follows that the linear span of score equations can approximate efficient influence curve for smooth functionals $\Psi(Q_0)$ up till $o_P(n^{-1/2})$. HAL-MLE solves the efficient influence curve equation.
- For example, for ATE, $P_n D^*(Q_{n,M}, G_0) = o_P(n^{-1/2})$ for M large enough.
- The undersmoothing condition is not parameter specific, so that such an undersmoothed HAL-MLE is efficient for all smooth functionals.
- In practice, we see that targeted undersmoothing works best.

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE**
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Targeted HAL-MLE

- We can add the ϵ -extension (e.g., clever covariate) of the least favorable submodel as an additional clever covariate in the linear span of indicator basis functions.
- We then define a Targeted HAL-MLE as the HAL-MLE that does not penalize the coefficient in front of clever covariate.
- This Targeted HAL-MLE solves the efficient influence curve equation, so that it is a TMLE (corresponding with initial estimator the HAL-MLE).
- So this targeted HAL-MLE is asymptotically efficient.
- One could combine undersmoothing the HAL-MLE, with this targeting.

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Collaborative estimation of treatment/censoring mechanism in TMLE

- The target estimand $\Psi(Q_0)$ (e.g, ATE) depends on true data distribution P_0 through a functional parameter Q_0 .
- The least favorable submodel $\{Q_{n,\epsilon} : \epsilon\}$ through initial estimator Q_n depends on an estimator G_n of another (often orthogonal) nuisance parameter G_0 (e.g. treatment/censoring mechanism).
- Let's estimate G_n with an HAL-MLE, but let's adapt its fit to its purpose, as in Collaborative-TMLE, where the increase in log-likelihood during TMLE step is iteratively used a local criterion.
- We have proposed to both tune the L_1 -norm and truncation level with $C - TMLE$, resulting in highly significant finite sample gains.

C-TMLE to tune HAL-MLE of G_0 : Simulation

- 1 Sample A conditionally on W from the Bernoulli law with parameter

$$G_0(W) \equiv \text{expit} \left(\delta + \sum_{k=1}^p \beta_k W_k \right),$$

where $(\beta_1, \dots, \beta_p) = (1, 1, 3/(p-2), \dots, 3/(p-2))$. We fix the ratio $p/n = 0.2$, and increase the sample size n from 200 to 2000.

- 2 Sample $\tilde{Y}$ conditionally on (A, W) from the Gaussian law with (conditional) variance $1/25$ and expectation

$$f_0(A, W) \equiv \frac{2}{5}(1 + W_1 + W_2 + W_5 + W_6 + W_8 + A),$$

then define $Y \equiv \text{expit}(\tilde{Y})$.

C-TMLE to tune HAL-MLE of G_0 : Simulation Results

For all the estimators, we use main term linear regression with $W_3, \dots, W_{10}$ as biased initial estimator.

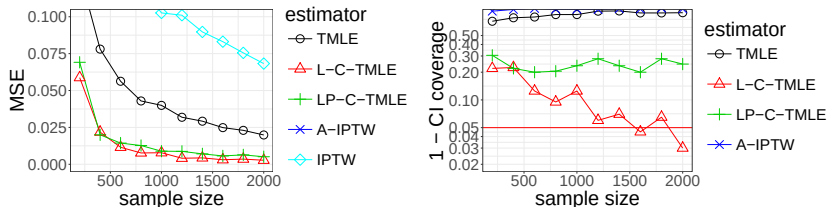


Figure: TMLE stands for TMLE with CV.LASSO as PS estimator; L-CTMLE is short for LASSO-C-TMLE, and L-P-CTMLE is a faster approximation of L-CTMLE.

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE**
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Super-efficient C-TMLE

- Consider the treatment specific mean example.
- Estimate $Q_0(A, W)$ with an $n^{-1/4}$ -estimator Q_n such as SL including HAL-MLE.
- Let g_0^r be the conditional distribution of A , given $Q_n(1, W)$, and let g_n^r be an $n^{-1/4}$ -estimator of g_0^r such as an HAL-MLE.
- Let Q_n^* be the TMLE update involving $A/g_n^r(A | W)$ as clever covariate.
- The super-efficient C-TMLE is defined as $\psi_n^* = \frac{1}{n} \sum_{i=1}^n Q_n^*(1, W_i)$.
- One can also use as off-set in g_n^r an initial estimator g_n^0 such a HAL-MLE only including basis functions with non zero coefficient in HAL-MLE of Q (CTMLE-OHAL).

Asymptotic linearity and super-efficiency of C-TMLE

- We have that ψ_n^* is an asymptotically linear estimator of ψ_0 with influence curve $D^*(Q_0, g_0^r)$, where $D^*(Q_0, g_0)$ is the efficient influence curve of the treatment-specific mean parameter.
- As a result, typically, due to inverse weighting by g_0^r results in smaller variance than inverse weighting by g_0 , ψ_n^* is super-efficient.

Simulation Results in Kang Shafer simulation

We use main-term linear regression to create biased initial estimator $\bar{Q}_n^0$ for TMLE/C-TMLE.

	TMLE-HAL	CTMLE-HAL	CTMLE-OHAL	oracle
N=500	7.06	6.34	2.94	3.64
N=1000	4.42	3.55	1.40	1.70
N=2000	2.94	1.85	0.83	0.87

Table: This table reports MSE for each estimator across 200 replications with different sample size. We can see CTMLE-HAL improves TMLE-HAL. CTMLE-OHAL further improves CTMLE-HAL with outcome adaptive regularization. It even achieves better performance than the oracle estimator, which is defined as TMLE with true PS.

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Nonparametric bootstrap of HAL-TMLE

- Fix M
- Draw 10,000 samples of size n from empirical measure P_n . For each bootstrap sample $P_n^\#$, recompute the HAL-TMLE(M), say $\mathbf{P}_n^{\#*}$.
- The HAL on bootstrap sample can be restricted to only include indicator basis functions that were selected by HAL-MLE(M) on original data.
- Use sampling distribution of $\psi_{n,M}^{\#*} = \Psi(\mathbf{Q}_{n,M}^{\#*})$, conditional on P_n , to construct 0.95-confidence interval.
- We propose to select M (larger than CV selector) by finding plateau in width of CI.

Simulation 1: ATE

Setting:

- $W \sim N(0, 4^2, -10, 10)$ drawn i.i.d. from a truncated normal distribution, bounded within $[-10, 10]$.
- $A \sim \text{Bernoulli}(p(W))$ with probability $p(W)$ as a function of W bounded between $[0.3, 0.7]$, given by
$$p(W) = 0.3 + 0.1W\sin(0.1W) + \varepsilon, \varepsilon \sim N(0, 0.05^2)$$
- $Y = 3\sin(a_1 W) + A + \varepsilon_2$ is a sinusoidal function of W , where $\varepsilon_2 \sim N(0, 1)$.
- a_1 controls the frequency (and true sectional variation norm) of the sinusoidal function.

The value of the parameter of interest, the ATE, is 1.

- The experiment is repeated 500 times, and interval coverages are computed

Simulation 1: ATE

Results ($n = 500$):

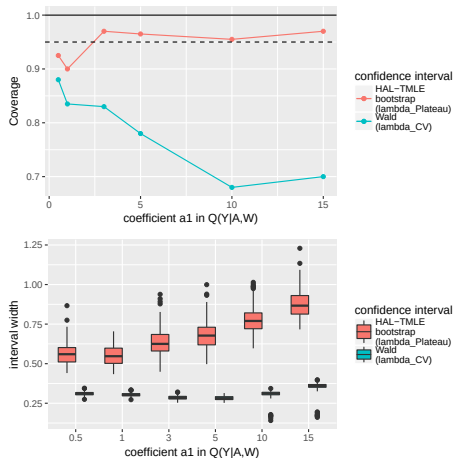
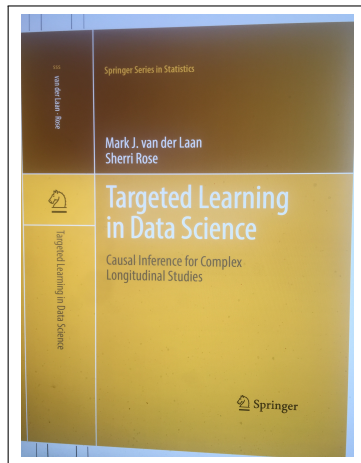
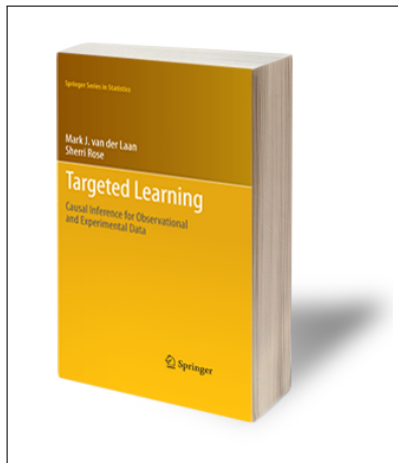


Figure: Coverage (upper) and interval width (lower) as a function of the a_1 coefficient (sectional variation norm) of the true distribution

Outline

- 1 TMLE
- 2 Highly Adaptive Lasso (HAL) MLE as initial estimator
- 3 HAL-TMLE: *Asymptotic* efficiency under strong positivity and finite sectional variation norm (vdL, 15)
- 4 Simulation to evaluate finite sample performance of HAL-TMLE of ATE ($d < 10$)
- 5 Undersmoothed HAL-MLE is efficient
- 6 Targeted HAL-MLE
- 7 Collaborative estimation of treatment/censoring mechanism in TMLE
- 8 Super-efficient C-TMLE
- 9 Nonparametric Bootstrap based Inference for HAL-TMLE
- 10 References

Targeted Learning (targetedlearningbook.com)



van der Laan & Rose, *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer, 2011.

- **General C-TMLE, and collaborative DR of CTMLE** van der Laan (2009), Gruber, van der Laan (2010), IJB.
- **Theoretical Results for continuous tuning with C-TMLE:** Ju, Cheng, Antoine Chambaz, and van der Laan M J.. *Collaborative targeted inference from continuously indexed nuisance parameter estimators*. arXiv preprint arXiv:1804.00102 (2018).
- **Application of scalable C-TMLE in Electronic Healthcare Database:** Ju, C., Wyss, R., Franklin, J. M., Schneeweiss, S., Haggstrom, J., van der Laan, M. J. (2017). *Collaborative-controlled LASSO for constructing propensity score-based estimators in high-dimensional data*. Statistical methods in medical research, 0962280217744588.

- **Application of C-TMLE for Propensity Score Truncation:** Ju, Cheng, Joshua Schwab, and Mark J. van der Laan. *On Adaptive Propensity Score Truncation in Causal Inference*. arXiv preprint arXiv:1707.05861 (2017).
- **A Generally Efficient Targeted Minimum Loss Based Estimator based on the Highly Adaptive Lasso:** Mark J. van der Laan, *Int J Biostat*, 13(2), Oct 2017. PMID: 29023235
- **Finite sample inference for Targeted Learning:** M.J. van der Laan (2017), <https://arxiv.org/abs/1708.09502>

Three general methods for efficient estimation in literature

Undersmoothing/overfitting MLE does not always work, depends on sieve, and how to undersmooth is challenge.

Three general methods result in asymptotically efficient estimators, given good initial estimator $\mathbf{P}_n$:

- One-step estimator: $\psi_n^1 = \Psi(\mathbf{P}_n) + P_n D^*(\mathbf{P}_n)$.
- Estimating equation estimator: Assume estimating function representation $D^*(P) = D^*(\psi, \eta(P))$; let ψ_n solution of $P_n D^*(\psi, \eta(\mathbf{P}_n)) = 0$.
- TMLE: $\mathbf{P}_{n,\epsilon}$ least favorable submodel through initial $\mathbf{P}_n$; ϵ_n MLE; $P_n^* = \mathbf{P}_{n,\epsilon_n}$; TMLE is $\Psi(P_n^*)$.
- TMLE is general method that updates initial $\mathbf{P}_n$ into improved fit $\mathbf{P}_n^*$ that solves user supplied set of equations $P_n D(\mathbf{P}_n^*) \approx 0$, allowing for various additional properties beyond asymptotic efficiency.

Each one of the methods has a sample splitting analogue removing Donsker class condition.

Simulation Results

We follow the simulation from Kang, Shafer, 2007. The original pre-treatment covariates $(Z_{i1}, \dots, Z_{i4})$ are generated from uncorrelated standard normal distribution.

With the pre-treatment covariates, the treatment indicator is then generated from a Bernoulli distribution with:

$$P(A_i = 1|Z_i) = \text{expit}(-Z_{i1} + 0.5Z_{i2} - 0.25Z_{i3} - 0.1Z_{i4})$$

The continuous potential outcomes are generated by the linear combination of the pre-treatment covariates, and does not rely on the treatment A . More specific, the potential outcome is generated by:

$$Y_i^{(1)} = Y_i^{(0)} = 210 + 27.4Z_{i1} + 13.7Z_{i2} + 13.7Z_{i3} + 13.7Z_{i4} + \epsilon$$
$$\epsilon \sim N(0, 1)$$

Thus the value of the parameter of interest, the ATE, is 0

Simulation Results

Only transformed covariates W are provided:

$$W_{i1} = \exp(Z_{i1}/2)$$

$$W_{i2} = z_{i2}/(1 + \exp(Z_{i1}) + 10)$$

$$W_{i3} = (Z_{i1}Z_{i3}/25 + 0.6)^3$$

$$W_{i4} = (Z_2 + Z_4 + 20)^2.$$

In our experiment, we further include an instrumental variable W_{i5} , and the treatment mechanism is changed into:

$$P(A_i = 1|Z_i, W_i) = \text{expit}((-Z_{i1} + 0.5Z_{i2} - 0.25Z_{i3} - 0.1Z_{i4})/2 + W_{i5})$$

Simulation Results in Kang Shafer simulation

We use main-term linear regression to create biased initial estimator $\bar{Q}_n^0$ for TMLE/C-TMLE.

	TMLE-HAL	CTMLE-HAL	CTMLE-OHAL	oracle
N=500	7.06	6.34	2.94	3.64
N=1000	4.42	3.55	1.40	1.70
N=2000	2.94	1.85	0.83	0.87

Table: This table reports MSE for each estimator across 200 replications with different sample size. We can see CTMLE-HAL improves TMLE-HAL. CTMLE-OHAL further improves CTMLE-HAL with outcome adaptive regularization. It even achieves better performance than the oracle estimator, which is defined as TMLE with true PS.