Optimal individualized decision rules using instrumental variable methods

Hongxiang Qiu[1]

Joint work with:
Marco Carone[1], Ekaterina Sadikova[2], Maria Petukhova[2], Ronald C. Kessler[2], Alex Luedtke[3]

[1]: Dept. of Biostatistics, University of Washington
[2]: Dept. of Health Care Policy, Harvard Medical School
[3]: Dept. of Statistics, University of Washington

Motivation

- Emerging area of individualized treatment rules (ITR).
- Previous methods assume no unmeasured confounding (e.g., Murphy, 2003; Robins, 2004; Zhao et al., 2012; Chakraborty and Moodie, 2013; Luedtke and van der Laan, 2016b).
  What if there is unmeasured confounding?

Motivation

- Emerging area of individualized treatment rules (ITR).
- Previous methods assume no unmeasured confounding (e.g., Murphy, 2003; Robins, 2004; Zhao et al., 2012; Chakraborty and Moodie, 2013; Luedtke and van der Laan, 2016b).
  What if there is unmeasured confounding?
- Instrumental variable (IV): another approach to identifying causal effects.
  Can we use an IV to estimate an optimal ITR?

Motivation

- Emerging area of individualized treatment rules (ITR).
- Previous methods assume no unmeasured confounding (e.g., Murphy, 2003; Robins, 2004; Zhao et al., 2012; Chakraborty and Moodie, 2013; Luedtke and van der Laan, 2016b).
  What if there is unmeasured confounding?
- Instrumental variable (IV): another approach to identifying causal effects.
  Can we use an IV to estimate an optimal ITR?
- Example:
  - IV: randomized treatment assignment
  - Treatment: actual treatment status
- Especially interested in settings with a treatment resource constraint (Luedtke and van der Laan, 2016a).

Motivation

- At times, direct intervention on treatment may be impossible or expensive.
- Individualized encouragement rule (IER): intervention on IV.
- Evaluate optimal rules: average benefit under optimal rule (compared to a reference rule).

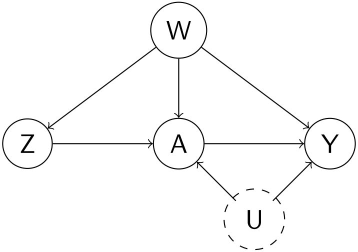
Table of Contents

Problem setup

- Observe iid $O = (W, Z, A, Y) \sim P_0$:
  - $W \in \mathcal{W}$: baseline covariates.
  - $Z \in \{0, 1\}$: binary IV.
  - $A \in \{0, 1\}$: binary treatment status (treatment vs control).
  - $Y \in \mathbb{R}$: outcome of interest (larger values are preferable).
- (Stochastic) individualized rule: $d : \mathcal{W} \to [0, 1]$ (prob of treatment)
- Counterfactuals:
  - $A(z)$: potential treatment status corresponding to $Z = z$
  - $Y(z, a)$: potential outcome corresponding to $(Z, A) = (z, a)$
- Given treatment resource constraint: $P_0(\text{receiving treatment}) \leq \kappa$.

IV conditions

- Relevance: $|P_0(A = 1 \mid Z = 1, W) - P_0(A = 1 \mid Z = 0, W)| > \delta^A$
- Exclusion restriction: $Y(0, a) = Y(1, a) =: Y(a)$
- Independence: $Z \perp\!\!\!\perp U \mid W$
- Z: IV/encouragement.
- U: unobserved confounder.



Case I: intervention on treatment

For an ITR $t : \mathcal{W} \to [0, 1]$, $Y(t) :=$ counterfactual outcome under t.

The optimal ITR t_0 solves

$$\textbf{maximize} \;\; \mathbb{E}[Y(t)] \;\; \textbf{subject to} \;\; \mathbb{E}[t(W)] \leq \kappa \; .$$

The impact of implementing the optimal ITR can be measured via its average treatment effect (ATE) relative to a given reference ITR t_r:

$$\mathbb{E}[Y(t_0) - Y(t_r)]$$

Case II: intervention on encouragement

For an IER $e : \mathcal{W} \to [0,1]$, $A(e) :=$ counterfactual treatment under e.

The optimal IER e_0 solves

$$\textbf{maximize} \;\; \mathbb{E}[Y(A(e))] \;\; \textbf{subject to} \;\; \mathbb{E}[A(e)] \leq \kappa \; .$$

The impact of implementing the optimal IER can be measured via its average encouragement effect (AEE) relative to a given reference IER e_r:

$$\mathbb{E}[Y(A(e_0)) - Y(A(e_r))]$$

Table of Contents

Additional notations

- IV $\rightarrow$ outcome effect (conditional average encouragement effect):

$$\Delta_0^Y(w) = \mathrm{E}_0[Y|Z=1, W=w] - \mathrm{E}_0[Y|Z=0, W=w]$$

- Treatment regression:

$$\mu_0^A(z,w) = \mathrm{E}_0[A \mid Z=z, W=w]$$

- IV $\rightarrow$ treatment effect:

$$\Delta_0^A(w) = \mu_0^A(1,w) - \mu_0^A(0,w) = \mathrm{E}_0[A \mid Z=1, W=w] - \mathrm{E}_0[A \mid Z=0, W=w]$$

- Wald estimand:

$$\Delta_0(w) = \frac{\Delta_0^Y(w)}{\Delta_0^A(w)} = \frac{\mathrm{E}_0[Y \mid Z=1, W=w] - \mathrm{E}_0[Y \mid Z=0, W=w]}{\mathrm{E}_0[A \mid Z=1, W=w] - \mathrm{E}_0[A \mid Z=0, W=w]}$$

Case I: intervention on treatment

Key identifying conditions (slightly relaxed version of Wang and Tchetgen Tchetgen (2018)):

- $Y(a) \perp\!\!\!\perp (A, Z) \mid (W, U)$.
- One of the following sets of conditions holds:
  (1) (a) (Uncorrelated IV) $\mathrm{Cov}(Y(0), Z \mid W) = 0$
  (b) (No unmeasured treatment-outcome effect modification)

$$\mathbb{E}[Y(1) - Y(0) \mid W, U] = \mathbb{E}[Y(1) - Y(0) \mid W]$$

  (2) (a) (Independent IV) $Z \perp\!\!\!\perp U \mid W$
  (b) (Independent compliance)

$$\begin{aligned}&\mathbb{E}[A(Z) \mid Z = 1, W, U] - \mathbb{E}[A(Z) \mid Z = 0, W, U] \\ &= \mathrm{E}_0[A \mid Z = 1, W] - \mathrm{E}_0[A \mid Z = 0, W]\end{aligned}$$

Case I: average treatment effect

ATE can be written as a summary of P_0.

Theorem (Identification of ATE)

- $\mathbb{E}\left[Y(1) - Y(0) \mid W\right] = \Delta_0(W)$
- $\mathbb{E}\left[Y(t) - Y(t_r)\right] = \mathrm{E}_0\left[\{t(W) - t_r(W)\}\,\Delta_0(W)\right]$ for any ITR t

The optimization problem in ITR t is equivalent to

$$\textbf{maximize}\ \ \mathrm{E}_0[t(W)\Delta_0(W)]\ \ \textbf{subject to}\ \ \mathrm{E}_0[t(W)] \leq \kappa\ .$$

Case I: optimal ITR

What does an optimal ITR look like?

Case I: optimal ITR

What does an optimal ITR look like?

Case I: optimal ITR

1. Sort subgroups according to $\Delta_0(W)$ (from high to low).
2. Assign treatment to those with highest (and positive) conditional ATE $\Delta_0(W)$ until treatment runs out.

$$t_0(w) = I\left\{\Delta_0(w) > \tau_0^T\right\}$$

Case II: intervention on encouragement

Key identifying condition: $Z \perp\!\!\!\perp (A(z), Y(a)) \mid W$

AEE can be written as a summary of P_0.

Theorem (Identification of AEE)

- $\mathbb{E}[Y(A(1)) - Y(A(0)) \mid W] = \Delta_0^Y(W)$
- $\mathbb{E}\left[Y(A(e)) - Y(A(e_r))\right] = \mathrm{E}_0\left[\{e(W) - e_r(W)\}\,\Delta_0^Y(W)\right]$ for any IER e

Case II: optimal IER

The optimization problem in IER e is equivalent to

$$\begin{aligned} \textbf{maximize} \quad & \mathrm{E}_0[e(W)\Delta_0^Y(W)] \\ \textbf{subject to} \quad & \mathrm{E}_0[e(W)\mu_0^A(1,W) + (1-e(W))\mu_0^A(0,W)] \leq \kappa \ , \end{aligned}$$

namely,

$$\begin{aligned} \textbf{maximize} \quad & \mathrm{E}_0[e(W)\Delta_0^Y(W)] \\ \textbf{subject to} \quad & \mathrm{E}_0[e(W)\Delta_0^A(W)] \leq \kappa - \mathrm{E}_0[\mu_0^A(0,W)] \ , \end{aligned}$$

Assume that $\Delta_0^A > 0$ and $\kappa - \mathrm{E}_0[\mu_0^A(0,W)] > 0$.

Case II: optimal IER

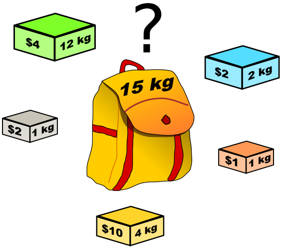
What does an optimal IER look like?

Case II: optimal IER

What does an optimal IER look like?

View as a fractional knapsack problem:

- Subgroup with same W: item
- Conditional AEE: $\Delta_0^Y(W)$ = value
- Additional proportion treated: $\Delta_0^A(W)$ = weight
- Max additional proportion treated: $\kappa - \mathrm{E}_0[\mu_0^A(0, W)]$ = total weight capacity
- $\xi_0(W) := \Delta_0^Y(W)/\Delta_0^A(W)$ = unit value



Case II: optimal IER

1. Sort subgroups according to $\xi_0(W)$ from high to low
2. Assign encouragement to treatment to those with highest (and positive) $\xi_0(W)$ until treatment runs out

$$e_0(w) = I\left(\xi_0(w) > \tau_0^E\right)$$

Table of Contents

Overview of targeted minimum-loss based estimation

- Goal: estimate $\Phi(P_0)$ (ATE, AEE)
- Seemingly natural approach:
  - Estimate P_0 with $\tilde{P}_n$
  - Plug in: estimate $\Phi(P_0)$ with $\Phi(\tilde{P}_n)$
- Problem: typically inefficient and not asymptotically normal.
- Efficient and asymptotically normal plug-in estimator using TMLE.
- Construct Wald CI based on asymptotic variance.

Overview of proposed procedure

1. Flexibly estimate relevant regression functions.
2. Estimate optimal ITR/IER with sample analogue.
3. Targeted estimation of ATE/AEE. The resulting estimator is asymptotically normal under conditions.

Simulation

Estimands:

- Resource constrained AEE: AEE of e_0 with $\kappa = 0.68$
- ATE: ATE of t_0 without resource constraint ($\kappa = 1$)
- Resource constrained ATE: ATE of t_0 with $\kappa = 0.25$

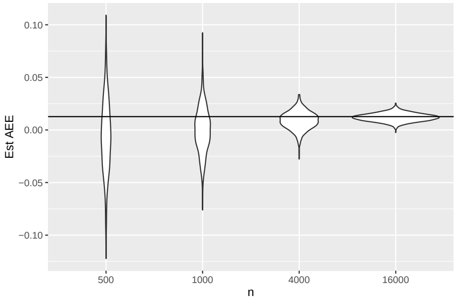
Data-generating mechanism has strong unmeasured treatment-outcome confounding.

Use sample splitting when estimating the optimal rule

- avoid a main source of finite-sample positive bias
- possible finite-sample negative bias
- valid 97.5% confidence lower bound, even under poorly estimated optimal rule

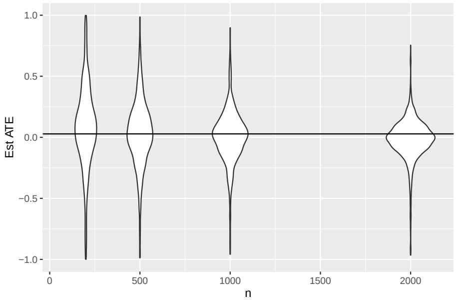
Simulation: AEE with resource constraints

Performance measure	Sample size	IV
95% Wald CI coverage	500	71%
	1000	74%
	4000	84%
	16000	90%
97.5% confidence lower bound coverage	500	96%
	1000	98%
	4000	98%
	16000	98%



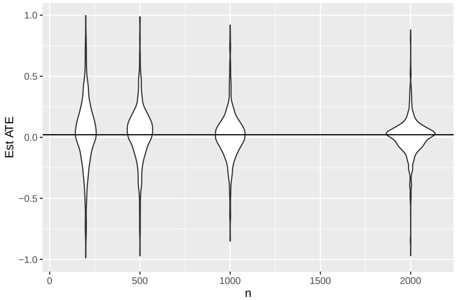
Simulation: ATE

Performance measure	Sample size	IV	Confounder
95% Wald CI coverage	200	97%	3%
	500	95%	$<$ 1%
	1000	93%	$<$ 1%
	2000	92%	$<$ 1%
97.5% confidence lower bound coverage	200	$>$ 99%	3%
	500	$>$ 99%	$<$ 1%
	1000	$>$ 99%	$<$ 1%
	2000	$>$ 99%	$<$ 1%



Simulation: ATE with resource constraints

Performance measure	Sample size	IV	Confounder
95% Wald CI coverage	200	96%	83%
	500	95%	84%
	1000	93%	87%
	2000	94%	88%
97.5% confidence lower bound coverage	200	$>$ 99%	97%
	500	$>$ 99%	97%
	1000	$>$ 99%	96%
	2000	$>$ 99%	92%



Discussion

- Estimators of optimal individualized treatment/encouragement rule using an IV under treatment resource constraints.
- Inference on average causal effects of the optimal rule under a locally nonparametric model.
- Cui and Tchetgen Tchetgen (2020) studied a similar problem with IV.
  - Cui and Tchetgen Tchetgen needed not consider intervention on encouragement or resource constraints.
  - Weaker identifying conditions for optimal ITR (not ATE).
  - Cui and Tchetgen Tchetgen's optimal ITR among compliers = our optimal IER
- Han (2020) and Cui and Tchetgen Tchetgen (2021) proposed even weaker identifying conditions for optimal ITR.
- Sun et al. (2021) studied a similar problem with random treatment cost and treatment cost constraint.
  - Our methods for IER and AEE can be readily adapted to this setting.
  - We additionally provide statistical inference on AEE.

References

Work based on

Qiu, H., M. Carone, E. Sadikova, M. Petukhova, R. C. Kesslar, and A. Luedtke (2020). Optimal individualized decision rules using instrumental variable methods. *JASA 116*(553), 174–191.

Qiu, H., M. Carone, A. Luedtke (publicly available soon). Estimation of optimal individualized encouragement rules and statistical inference about their effects under treatment resource constraints

References

Chakraborty, B. and E. E. Moodie (2013). *Statistical Methods for Dynamic Treatment Regimes*. Statistics for Biology and Health. New York, NY: Springer New York.

Cui, Y. and E. Tchetgen Tchetgen (2020). A Semiparametric Instrumental Variable Approach to Optimal Treatment Regimes Under Endogeneity. *Journal of the American Statistical Association*.

Cui, Y. and E. Tchetgen Tchetgen (2021). On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable. *Statistics & Probability Letters*, 109180.

Han, S. (2020). Comment: Individualized Treatment Rules Under Endogeneity. *Journal of the American Statistical Association*.

Luedtke, A. R. and M. J. van der Laan (2016a). Optimal Individualized Treatments in Resource-Limited Settings. *International Journal of Biostatistics 12*(1), 283–303.

Luedtke, A. R. and M. J. van der Laan (2016b). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics 44*(2), 713–742.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 331–355.

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. pp. 189–326. Springer, New York, NY.

Sun, H., S. Du, and S. Wager (2021). Treatment Allocation under Uncertain Costs. *arXiv preprint arXiv:2103.11066v1*.

References

Wang, L. and E. Tchetgen Tchetgen (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(3), 531–550.

Zhao, Y., D. Zeng, A. J. Rush, and M. R. Kosorok (2012). Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association 107*(499), 1106–1118.

Thank you!

Questions?

Additional notations

For any distribution P on observed data*, define

- $\mu_P^Z : w \mapsto \mathrm{E}_P(Z \mid W = w)$ *(IV propensity)*
- $\mu_P^A : (z, w) \mapsto \mathrm{E}_P\left(A \mid Z = z, W = w\right)$ *(Treatment regression)*
- $\mu_P^Y : (z, w) \mapsto \mathrm{E}_P\left(Y \mid Z = z, W = w\right)$ *(Outcome regression)*
- $\Delta_P^A : w \mapsto \mu_P^A(1, w) - \mu_P^A(0, w)$ *(IV $\rightarrow$ treatment effect)*
- $\Delta_P^Y : w \mapsto \mu_P^Y(1, w) - \mu_P^Y(0, w)$ *(IV $\rightarrow$ outcome effect)*
- $\Delta_P : w \mapsto \Delta_P^Y(w)/\Delta_P^A(w)$ *(Wald estimand)*

*For P_0, we use 0 instead of P_0 in the subscript.

Formal optimal ITR

Theorem (Identifying optimal ITR)

With $\eta_0^T := \inf\{\eta : P_0(\Delta_0(W) > \eta) \leq \kappa\}$ *and the threshold* $\tau_0^T := \max\{\eta_0^T, 0\}$, *the optimal ITR is*

$$t_0(w) := \begin{cases} \frac{\kappa - P_0\{\Delta_0(W) > \tau_0^T\}}{P_0\{\Delta_0(W) = \tau_0^T\}} & : \Delta_0(w) = \tau_0^T > 0, \\ & \quad P_0\{\Delta_0(W) = \tau_0^T\} > 0 \\ I\{\Delta_0(w) > \tau_0^T\} & : \textit{otherwise.} \end{cases}$$

Formal optimal IER

Theorem (Identifying optimal IER)

With $\eta_0^E := \inf \left\{\eta : \mathrm{E}_0\left[I(\xi_0(W) > \eta)\Delta_0^A(W)\right] \leq \kappa - \mathrm{E}_0[\mu_0^A(0, W)]\right\}$ *and* $\tau_0^E := \max\left\{\eta_0^E, 0\right\}$, *the optimal IER is*

$$e_0(w) = \begin{cases} \frac{\kappa - \mathrm{E}_0[\mu_0^A(0,W)] - \mathrm{E}_0\left[I(\xi_0(W) > \tau_0^E)\Delta_0^A(W)\right]}{\mathrm{E}_0\left[I(\xi_0(W) = \tau_0^E)\Delta_0^A(W)\right]} & : \xi_0(w) = \tau_0^E > 0, \\ & \quad \mathrm{E}_0\left[I(\xi_0(W) = \tau_0^E)\bar{\mu}_0^A(W)\right] > 0 \\ I\left(\xi_0(w) > \tau_0^E\right) & : \textit{otherwise.} \end{cases}$$