Causal mediation analysis for stochastic interventions

Nima Hejazi

Wednesday 29th April, 2020

Graduate Group in Biostatistics, and
Center for Computational Biology,
University of California, Berkeley

nshejazi
nhejazi
nimahejazi.org
bit.ly/2020_berkeley_medshift
joint work with Iván Díaz & Mark van der Laan

Structural causal model

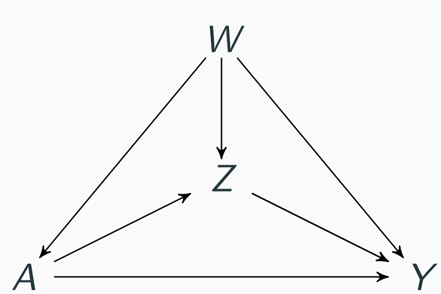
- Observed world

$$W = f_W(U_W)$$
$$A = f_A(W, U_A)$$
$$Z = f_Z(W, A, U_Z)$$
$$Y = f_Y(W, A, Z, U_Y)$$

- Intervened world

$$W = f_W(U_W)$$
$$A = a$$
$$Z(a) = f_Z(W, a, U_Z)$$
$$Y(a) = f_Y(W, a, Z(a), U_Y)$$

- Also consider counterfactual $Y(a, z) = f_Y(W, a, z, U_Y)$
- Note that $Y(a, Z(a)) = Y(a)$



Natural (in)direct effects

Counterfactuals:

$$Y(1) = f_Y(W, 1, Z(1), U_Y); \qquad Y(0) = f_Y(W, 0, Z(0), U_Y)$$

Average treatment effect:

$$\begin{aligned}\psi_{\text{ATE}} &= \mathbb{E}\{Y(1) - Y(0)\} \\ &= \underbrace{\mathbb{E}\{Y(1, Z(1)) - Y(1, Z(0))\}}_{\text{natural indirect effect}} + \underbrace{\mathbb{E}\{Y(1, Z(0)) - Y(0, Z(0))\}}_{\text{natural direct effect}}\end{aligned}$$

Problems

- Focuses on binary exposures and static interventions
- Needs cross-world counterfactual independencies
- Requires identification of the exposure-mediator effect
- Requires positivity of exposure and mediator mechanisms

Stochastic interventions

Definition
Stochastic interventions yield a post-intervention exposure that is itself a random variable after conditioning on covariates W.

We will study two types of stochastic interventions:

- Modified treatment policies $d(A, W)$.
- Exponential tilting.

Modified treatment policies

- Consider a hypothetical world where the treatment received is some function $d(A, W)$ of the treatment *actually* received A and covariates W.

Example (Haneuse and Rotnitzky (2013))

- What is the impact of operating time on outcomes for patients undergoing surgical resection for non-small-cell lung cancer.
- We can answer this using a hypothetical intervention $d(A, W) = A - \delta$ for user-supplied δ.
- Denote the post-intervention exposure $A_\delta = d(A, W)$
- Non-parametric definition of effects with an interpretation that is familiar to users of OLS regression adjustment.
- Requires assuming piecewise smooth invertibility of $d(\cdot, w)$.

Exponential tilting

- Consider an intervention that changes the exposure distribution conditional on covariates from $g(a \mid w)$ to $g_\delta(a \mid w)$, where

$$g_\delta(a \mid w) \propto \exp(\delta a) g(a \mid w)$$

- Denote by A_δ a draw from post-intervention distribution g_δ.

Example (Kennedy (2018))

- *Incremental propensity score interventions.* For binary A,

$$g_\delta(1 \mid w) = \frac{\delta g(1 \mid w)}{\delta g(1 \mid w) + 1 - g(1 \mid w)},$$

- Here, $\text{odds}\{A_\delta = 1 \mid W = w\} = \delta \, \text{odds}\{A = 1 \mid W = w\}$.

Population intervention effect

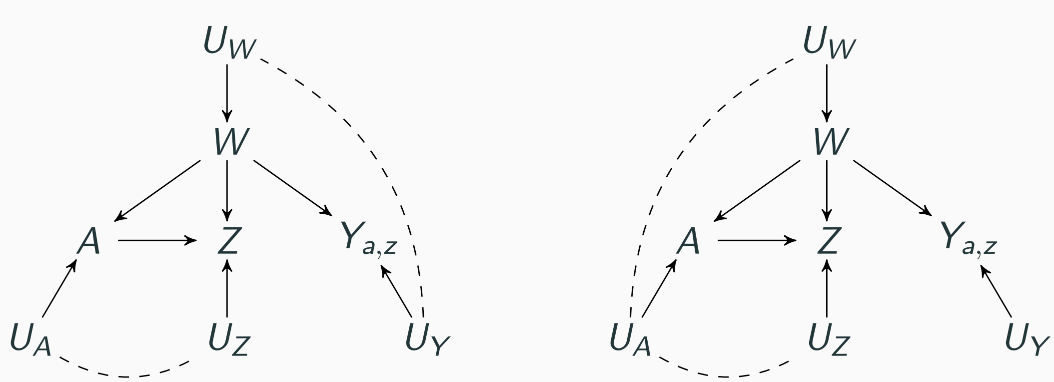
- For $A_\delta = d(A, W)$ (modified treatment policy), or A_δ being a draw from $g_\delta(\cdot \mid W)$, define the population intervention effect (PIE) of A on Y as

$$\begin{aligned}\psi_{\text{PIE}}(\delta) &= \mathbb{E}\{Y(A_\delta) - Y\} \\ &= \underbrace{\mathbb{E}\{Y(A_\delta, Z(A_\delta)) - Y(A_\delta, Z)\}}_{\text{indirect effect}} + \underbrace{\mathbb{E}\{Y(A_\delta, Z) - Y(A, Z)\}}_{\text{direct effect}}.\end{aligned}$$

- Identification and estimation of $\mathbb{E}\{Y(A_\delta, Z(A_\delta)\}$ has been done (e.g., Díaz and van der Laan 2012, Kennedy 2018).
- Díaz and Hejazi (2020) focus on the decomposition term $\theta(\delta) = \mathbb{E}\{Y(A_\delta, Z)\}$

Identification

- *Common support:* $\text{supp}\{g_\delta(\cdot \mid w)\} \subseteq \text{supp}\{g(\cdot \mid w)\}$
- *Conditional exchangeability:* $\mathbb{E}\{Y(a, z) \mid A, W, Z\} = \mathbb{E}\{Y(a, z) \mid W, Z\}$



Identification

$$\theta(\delta) = \int m(a, z, w) g_\delta(a \mid w) p(z, w) \mathrm{d}\nu(a, z, w),$$

where $m(a, z, w) = \mathbb{E}(Y \mid A = a, Z = z, W = w)$.

Compared to the natural (in)direct effects:

- No cross-world independence assumptions.
- No need to identify exposure-mediator effect.
- Positivity guaranteed by definition.
- Allows nonparametric effects for continuous exposures.

G-computation and IPW estimation

$$\begin{aligned} \theta(\delta) &= \mathbb{E}\left\{\int m(a, Z, W) g_\delta(a \mid W) \mathrm{d}\nu(a)\right\} \\ &= \mathbb{E}\left\{\frac{g_\delta(A \mid W)}{e(A \mid Z, W)} Y\right\}, \end{aligned}$$

where e is the pdf of $A \mid Z, W$:

$$e(a \mid z, w) = \frac{g(a \mid w) p(z \mid a, w)}{p(z \mid w)}.$$

- The above formulas may be used to construct the classical G-computation and IPW estimators.
- If the nuisance parameters estimators are data-adaptive, G-computation and IPW generally fail to be $n^{1/2}$-consistent.

Efficient one-step estimator

- *Main idea:* Find $D_P(o)$ such that:

$$\theta_{\hat{P}}(\delta) - \theta_P(\delta) = -\mathbb{E}\{D_{\hat{P}}(O)\} + O(||\hat{P} - P||^2)$$

- De-bias $\theta_{\hat{P}}(\delta)$ by computing

$$\tilde{\theta}(\delta) = \theta_{\hat{P}}(\delta) + \frac{1}{n}\sum_{i=1}^{n} D_{\hat{P}}(O_i)$$

- In the nonparametric model there is only one such D_P; it is referred to as *canonical gradient* or *efficient influence function*.

Efficient Influence Function for $\theta(\delta)$

- For simplicity, let $A \in \{0, 1\}$ and consider incremental propensity score interventions (Kennedy 2018), EIF given as

$$D^{Z,W}_{\eta,\delta}(o) = \int m(z, a, w) g_\delta(a \mid w) d\kappa(a)$$

$$D^{Y}_{\eta,\delta}(o) = \frac{\mathbf{g}_\delta(\mathbf{a} \mid \mathbf{w})}{\mathbf{e}(\mathbf{a} \mid \mathbf{z}, \mathbf{w})}\{y - m(z, a, w)\},$$

$$D^{A}_{\eta,\delta}(o) = \frac{\delta\phi(\mathbf{w})\{a - g(1 \mid w)\}}{\{\delta\mathbf{g}(\mathbf{1} \mid \mathbf{w}) + \mathbf{g}(\mathbf{0} \mid \mathbf{w})\}^{\mathbf{2}}},$$

where $\phi(w) = \mathbb{E}\{m(1, Z, W) - m(0, Z, W) \mid W = w\}$.

- For an unabridged treatment, see Díaz and Hejazi (2020).
- Original work does not construct a TML estimator.

Weak convergence for exponential tilting

- D_P does not depend on all of P, just on $\eta = (m, g, e, \phi)$.
- Construct cross-validated estimates of m, g, e, ϕ — i.e., for computing $\hat{m}(O_i)$, use training data not containing O_i.
- Assume convergence of certain second-order terms — e.g., $\|\hat{m} - m\|\|\hat{g} - g\| = o_P(n^{-1/2})$.
- Then, $\sqrt{n}\{\tilde{\theta}(\delta) - \theta(\delta)\} \rightsquigarrow N(0; \text{var}\{D_\eta(O)\})$.
- This result can be made uniform in intervals $[\delta_l, \delta_u]$
- Uniform result can be used to test the hypothesis of no direct effect (Kennedy 2018): $H_0 : \sup_{\delta \in \Delta} \theta(\delta) = \mathbb{E}(Y)$.

Multiple robustness

- *Modified treatment policies:* Assume piecewise smooth invertibility of $d(\cdot, w)$, and define $r(z \mid w) = p(z \mid w)$. Consistency requires:
  1. $g_1 = g$ and either $e_1 = e$ or $m_1 = m$, or
  2. $m_1 = m$ and either $g_1 = g$ or $r_1 = r$.

  Intuition: use change of variable formula to get

$$\theta(\delta) = \mathbb{E}\left\{\int m(d(A, W), z, W) r(z \mid, W) \mathrm{d}\nu(z)\right\}.$$

- *Exponential tilt:* consistency requires that $g_1 = g$ and either $e_1 = e$ or $m_1 = m$.

TML Estimator for $\theta(\delta)$

- We can construct a TML estimator by using the EIF to update initial estimates of nuisance parameters:

$$\hat{\theta}_{\text{TMLE}}(\delta) = \int \frac{1}{n} \sum_{i=1}^{n} \hat{m}^{\star}_{j(i)}(Z, a, W) \hat{g}^{\star}_{\delta, j(i)}(a \mid W) d\kappa(a).$$

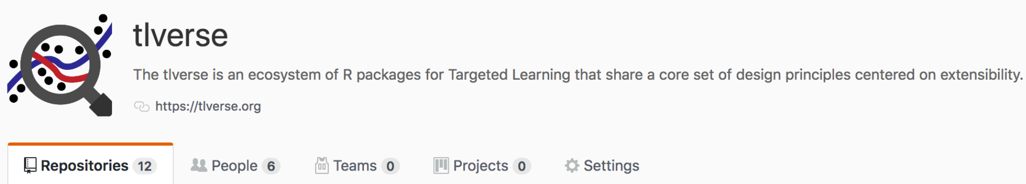
- TMLE constructs a substitution estimator, respecting bounds.
- Avoid entropy conditions by cross-validation (Zheng and van der Laan 2011, Chernozhukov et al. 2016), so let $j(i)$ be the index of the validation set containing observation i.
- Use universal least favorable submodels (van der Laan and Gruber 2016) for the targeting step.

Targeting Step of TMLE for $\theta(\delta)$

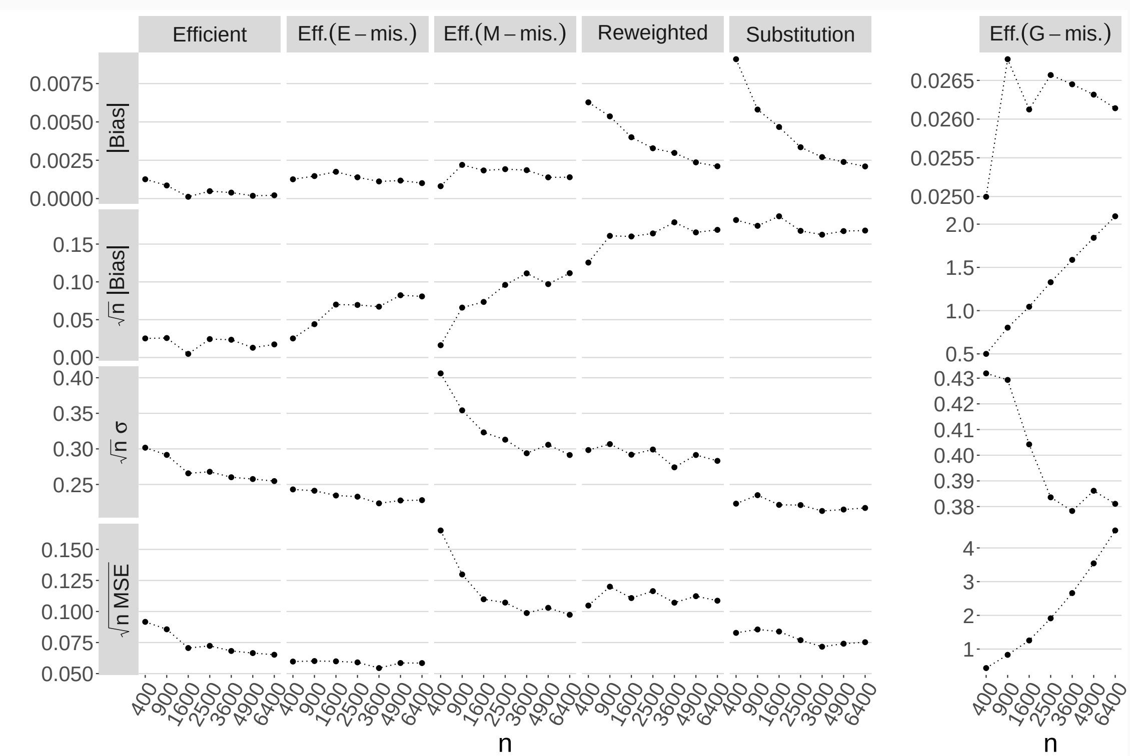
- $\hat{g}^{\star}_{\delta}(a \mid w)$ generated via *targeting* fluctuation that tilts initial estimates towards solutions of the score $\frac{1}{n}\sum_{i=1}^{n} D^A(O_i) = 0$: $\text{logit}(\hat{g}_{\delta,k\xi}) = \text{logit}(\hat{g}_{\delta,(k-1)\xi}) + \xi^{\text{lfm}}_{\Delta g} \mathbf{H}^{\mathbf{A}}_{\mathbf{(k-1)}\xi}$
  - Take $\hat{g}_{\delta,k\xi}$ in final step as $\hat{g}^{\star}_{\delta}(a \mid w)$.
  - Use the term before the residual $a - g(1 \mid w)$ in D^A as the covariate in this regression (treating initial estimate as offset).
- Similarly for $\hat{m}^{\star}(z, a, w)$ but to solve $\frac{1}{n}\sum_{i=1}^{n} D^Y(O_i) = 0$: $\text{logit}(\hat{m}_{k\xi}) = \text{logit}(\hat{m}_{(k-1)\xi}) + \xi^{\text{lfm}}_{\Delta m} \mathbf{H}^{\mathbf{Y}}_{\mathbf{(k-1)}\xi}$.
  - Take $\hat{m}_{k\xi}$ in final step as $\hat{m}^{\star}_{j(i)}(Z, a, W)$.
  - Use the term before the residual $y - m(z, a, w)$ in D^Y as the covariate in this regression (treating initial estimate as offset).

Software implementation

- The medshift R package (Hejazi and Díaz 2020) implements TML estimator with state-of-the-art machine learning.
  - Access all estimators via the eponymous `medshift()` function.
  - Uses the `s13` R package for ensemble machine learning.
  - Relies on the `tmle3` framework for the TMLE implementation.
  - Cross-fitting implementation via the `origami` R package.
- `s13`, `tmle3`, and `origami` are the 3 core engines of the **`tlverse`** software ecosystem (`https://tlverse.org`).
  - Our handbook: `https://tlverse.org/tlverse-handbook`



Numerical studies



Some notes and next steps

- Nonparametric efficient estimation of effects using data-adaptive regression and cross-validation
  - Avoid reliance on misspecified parametric models.
  - Cross-validation helps keep the function classes unrestricted.
- Working on adaptations to mediator-outcome confounders affected by treatment.
- R package: `https://github.com/nhejazi/medshift`
- Integrated in the `tlverse` targeted learning ecosystem
- Paper (JRSS-B): `https://doi.org/10.1111/rssb.12362`
- arXiv pre-print: `https://arxiv.org/abs/1901.02776`
- We would love to hear your input:
  - ild2005@med.cornell.edu
  - nhejazi@berkeley.edu

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., et al. (2016). Double machine learning for treatment and causal parameters. *arXiv preprint arXiv:1608.00060*.

Díaz, I. and Hejazi, N. S. (2020). Causal mediation analysis for stochastic interventions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Díaz, I. and van der Laan, M. J. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549.

Haneuse, S. and Rotnitzky, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in medicine*, 32(30):5260–5277.

Hejazi, N. S. and Díaz, I. (2020). *medshift: Causal mediation analysis for stochastic interventions in R*. R package version 0.1.4.

Kennedy, E. H. (2018). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, (just-accepted).

van der Laan, M. and Gruber, S. (2016). One-step targeted minimum loss-based estimation based on universal least favorable one-dimensional submodels. *The international journal of biostatistics*, 12(1):351–378.

Zheng, W. and van der Laan, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning*, pages 459–474. Springer.

Thank you.

Slides: bit.ly/2020_berkeley_medshift

https://nimahejazi.org

https://github.com/nhejazi

https://twitter.com/nshejazi

Appendix

Haneuse and Rotnitzky (2013)

- *Proposal:* Characterization of stochastic interventions as *modified treatment policies* (MTPs).
- Assumption of *piecewise smooth invertibility* allows for the intervention distribution of any MTP to be recovered:

$$g_{0,\delta}(a \mid w) = \sum_{j=1}^{J(w)} I_{\delta,j}\{h_j(a,w), w\} g_0\{h_j(a,w) \mid w\} h_j'(a,w)$$

- Such intervention policies account for the natural value of the intervention A directly yet are interpretable as the imposition of an altered intervention mechanism.
- Identification conditions for assessing the parameter of interest under such interventions appear technically complex (at first).