Efficient Estimation of Stochastic Intervention Effects in Causal Mediation Analysis

Nima Hejazi

Thursday, 6th August 2020

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

nshejazi
nhejazi
nimahejazi.org
with Iván Díaz & Mark van der Laan

Structural causal model

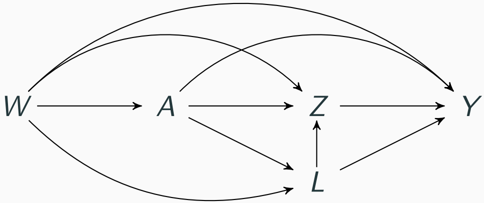
- Observed world

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

- Counterfactuals

$$Y(a) = f_Y(Z(a), L(a), a, W, U_Y)$$
$$Y(z, a) = f_Y(z, L(a), a, W, U_Y)$$

- Directed acyclic graph



Stochastic interventions

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

Generally, consider $A_\delta \sim g_\delta(a, w)$, where g_δ is a post-intervention distribution of A.

We will address 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 (2019))

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

Stochastic mediation effects

- For $A_\delta = d(A, W)$ or A_δ being a draw from $g_\delta(\cdot \mid W)$, Díaz and Hejazi (2020) defined the population intervention (in)direct effects:

$$\text{PIDE} = \mathbb{E}\{f_Y(Z, L, A, W, U_Y) - f_Y(Z, L_{A_\delta}, A_\delta, W, U_Y)\}$$
$$\text{PIIE} = \mathbb{E}\{f_Y(Z, L_{A_\delta}, A_\delta, W, U_Y) - f_Y(Z_{A_\delta}, L_{A_\delta}, A_\delta, W, U_Y)\}.$$

- Direct effect measures effect through paths *not* involving the mediator: $A \to Y$ and $A \to L \to Y$.
- Indirect effect measures the effect through paths involving the mediator: $A \to Z \to Y$ and $A \to L \to Z \to Y$.
- Not identified with intermediate confounder L:
  - Since L is a confounder of $\{Z, L\}$, adjustment required.
  - Since L is on the path from A to Y, adjustment disallowed.

Interventional stochastic mediation effects

- Stochastic intervention replaces a by $A_\delta \sim g_\delta(\cdot \mid w)$.
- Interventional effects involve stochastic interventions on Z, replacing z with G_δ, a random draw from distribution of Z_{A_δ} conditional on $\{A_\delta, W\}$.
- Interventional stochastic (in)direct effects:

$$\psi(\delta) = \overbrace{\mathbb{E}\{Y_{A,G} - Y_{A_\delta,G}\}}^{\text{DE}} + \overbrace{\mathbb{E}\{Y_{A_\delta,G} - Y_{A_\delta,G_\delta}\}}^{\text{IE}}.$$

Identification

- Common support: Assume $\mathrm{supp}\{g_\delta(\cdot \mid w)\} \subseteq \mathrm{supp}\{g(\cdot \mid w)\}$ $\forall w \in \mathcal{W}$.
- No unmeasured exposure-outcome confounder: Assume $Y_{a,z} \perp\!\!\!\perp A \mid W$.
- No unmeasured mediator-outcome confounder: Assume $Y_{a,z} \perp\!\!\!\perp Z \mid (L, A, W)$.
- No unmeasured exposure-mediator confounder: Assume $Z_a \perp\!\!\!\perp A \mid W$.

Identification

Under identification assumptions, the direct effect $\psi_{D,\delta}$ and indirect effect $\psi_{I,\delta}$ are identified and given, respectively, by

$$\psi_{D,\delta} = \theta_{1,0} - \theta_{2,\delta}$$
$$\psi_{I,\delta} = \theta_{2,\delta} - \theta_{1,\delta},$$

where

$$\theta_{1,\delta} = \int m(z,l,a,w)p(l \mid a,w)p(z \mid a,w)g_\delta(a \mid w)p(w)d\nu(a,z,l,w),$$
$$\theta_{2,\delta} = \int m(z,l,a,w)p(l \mid a,w)p(z \mid w)g_\delta(a \mid w)p(w)d\nu(a,z,l,w).$$

Compared to related (in)direct effects

- No cross-world independence assumptions, like the effects of Díaz and Hejazi (2020).
- Positivity guaranteed by definition, unlike the interventional effects of VanderWeele et al. (2014).
- Allows nonparametric effects for continuous exposures under intermediate confounding.
- Reduces to the stochastic mediation effects of Díaz and Hejazi (2020) in the absence of intermediate confounders L.

Efficient influence function

- The efficient influence function (EIF) characterizes asymptotic behavior of all regular, asymptotically linear estimators.
- We require the EIF for $\theta_{1,\delta}$ and $\theta_{2,\delta}$, which we denote $D^1_{P,\delta}(o)$ and $D^2_{P,\delta}(o)$, respectively.
- EIF takes the form $D^j_{P,\delta}(o) = S^j_{P,\delta}(o) - S^{j,A}_{P,\delta}(o)$, for orthogonal scores given by $S^j_{P,\delta}(o)$ and $S^{j,A}_{P,\delta}(o)$.
- The exact form of $S^{j,A}_{P,\delta}(o)$ varies by the type of stochastic intervention (modified treatment policies, exponential tilting).
- Re-parameterizations can simplify the estimation process when either L or Z is low-dimensional.

Efficient estimation

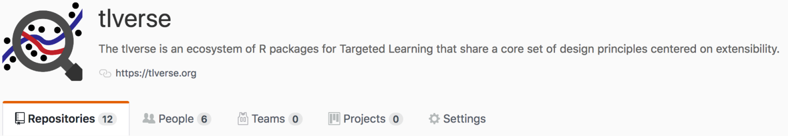
- We can use the EIF in the nonparametric model to construct efficient estimators of $\psi_{D,\delta}$ and $\psi_{I,\delta}$.
- Either one-step estimation or targeted minimum loss estimation (TMLE); denote estimators.
- Unlike one-step estimation, TMLE constructs substitution estimators, respecting bounds by updating in model space.
- Avoid entropy conditions by cross-validation (Zheng and van der Laan 2011, Chernozhukov et al. 2018).
- Multiple robustness: consistency $\hat{\theta}_{1,\delta}$ and $\hat{\theta}_{2,\delta}$ across six nuisance parameter configurations.

Weak convergence

- $D^1_{P,\delta}(o)$ and $D^2_{P,\delta}(o)$ do not depend on all of P, only on nuisance parameters $\eta = (m, g, b, \bar{u}, v, d, e, s, q)$.
- Construct cross-validated estimates of η — i.e., for computing $\hat{\eta}(O_i)$, use training data not containing O_i.
- Assume convergence of certain second-order terms — e.g., $\|\hat{m} - m\|\{\ldots\} = o_P(n^{-1/2})$, $\|\hat{g} - g\|\{\ldots\} = o_P(n^{-1/2})$, $\|\hat{b} - b\|\{\ldots\} = o_P(n^{-1/2})$.
- Then, $\sqrt{n}\{\hat{\psi}_{D,\delta} - \psi_{D,\delta}\} \rightsquigarrow N(0; \mathrm{var}\{D^1_{\eta,0}(O) - D^2_{\eta,\delta}(O)\})$ and $\sqrt{n}\{\hat{\psi}_{I,\delta} - \psi_{I,\delta}\} \rightsquigarrow N(0; \mathrm{var}\{D^2_{\eta,\delta}(O) - D^1_{\eta,\delta}(O)\})$.
- Wald-type confidence intervals may be generated based on estimation of the variance terms.

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 `sl3` R package for ensemble machine learning.
  - Relies on the `tmle3` framework for the TMLE implementation.
  - Cross-fitting implementation via the `origami` R package.
- `sl3`, `tmle3`, and `origami` are the 3 core engines of the **`tlverse`** software ecosystem (`https://tlverse.org`).
  - Handbook: `https://tlverse.org/tlverse-handbook`



Final remarks

- Nonparametric efficient estimation of stochastic interventional (in)direct effects with flexible regression and cross-validation.
    - Avoid reliance on misspecified parametric models.
    - Cross-validation minimizes assumptions on estimators.
- `R` package: `https://github.com/nhejazi/medshift`
- For stochastic mediation, see Díaz and Hejazi (2020): `https://doi.org/10.1111/rssb.12362`
- This paper: soon to appear on the arXiv (end of the month).

Thank you!

https://nimahejazi.org

https://github.com/nhejazi

https://twitter.com/nshejazi

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. M. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Díaz, I. and Hejazi, N. S. (2020). Causal mediation analysis for stochastic interventions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):661–683.

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. (2019). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, 114(526):645–656.

VanderWeele, T. J., Vansteelandt, S., and Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology (Cambridge, Mass.)*, 25(2):300.

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

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).