Motivating targeted learning

Anders Munch

April 25, 2024

A parameter of interest

A target parameter is a functional

$$\Psi\colon \mathcal{P} \longrightarrow \mathbb{R}.$$

Common case that

$$\Psi(P) = P[\varphi(\,\cdot\,; \nu(P))] = \int \varphi(x; \nu(P)) P(\mathrm{d}x),$$

where ν is a function-valued nuisance parameter such as a conditional expectation or a conditional probability.

A parameter of interest – our setting

Population of women at their second birth [Wikkelsø et al., 2014].

The data consists of observations $X = (Y, A, W)$, where $Y \in \{0, 1\}$ denotes PPH, $A \in \{0, 1\}$ denotes planned c-section, and $W \in \mathbb{R}^p$ is a vector with information collected at the start of the second pregnancy, including information from the first birth.

Our parameter of interest is

$$\begin{aligned}\Psi(P) &= \mathbb{E}_P[\mathbb{E}_P[Y \mid A = 1, W] - \mathbb{E}_P[Y \mid A = 0, W]] \\ &= P[\nu(1, \cdot; P) - \nu(0, \cdot; P)],\end{aligned}$$

where

$$\nu(a, w; P) = \mathbb{E}_P[Y \mid A = a, W = w].$$

If W includes all confounders of Y and A, and there is uniform positive probability of treatment, Ψ(P) can be interpreted as the average treatment effect (ATE) of a planned c-section on PPH. This is known a the g-formula.

Parametric modeling versus ML

Parametric approach

Fit a parametric model for ν (e.g., logistic regression). Estimate Ψ with

$$\hat{\Psi}_n^{\mathrm{glm}} = \mathbb{P}_n[\hat{\nu}_n^{\mathrm{glm}}(1,\cdot) - \hat{\nu}_n^{\mathrm{glm}}(0,\cdot)] = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\nu}_n^{\mathrm{glm}}(1, W_i) - \hat{\nu}_n^{\mathrm{glm}}(0, W_i)\right\}.$$

Nonparametric (machine learning) approach

Fit a machine learning algorithm to learn ν (e.g., random forest). Estimate Ψ with

$$\hat{\Psi}_n^{\mathrm{RF}} = \mathbb{P}_n[\hat{\nu}_n^{\mathrm{RF}}(1,\cdot) - \hat{\nu}_n^{\mathrm{RF}}(0,\cdot)] = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\nu}_n^{\mathrm{RF}}(1, W_i) - \hat{\nu}_n^{\mathrm{RF}}(0, W_i)\right\}.$$

What would you prefer in our setting?

Targeted estimators

Semi-parametric efficiency theory tells us that an initial estimator can be improved by adding an augmentation term:

$$\hat{\Psi}_n = \hat{\Psi}_n^{\mathsf{RF}} + \mathbb{P}_n[\dot{\psi}(\cdot, \hat{P}_n)].$$

The function $\dot{\psi}(\cdot, P)$ is the efficient influence function (aka canonical gradient) of Ψ.

The term $P[\dot{\psi}(\cdot, \hat{P}_n)]$ can be interpreted as the first order bias due to the estimation of ν with $\hat{\nu}_n^{\mathsf{RF}}$, and we approximate this with $\mathbb{P}_n[\dot{\psi}(\cdot, \hat{P}_n)]$.

Targeted estimators – our setting

The EIF for our Ψ is well-known [e.g., Kennedy, 2016, 2022, Hines et al., 2022] and equals

$$\dot{\psi}(X;P) = \nu(W,1;P) - \nu(W,0;P) - \Psi(P) \\ + \frac{A}{\pi(W;P)}(Y - \nu(W,1;P)) - \frac{1-A}{1-\pi(W;P)}(Y - \nu(W,0;P)),$$

where π is the propensity score,

$$\pi(w;P) = P(A = 1 \mid W = w).$$

A targeted estimator of Ψ is then

$$\hat{\Psi}_n = \frac{1}{n}\sum_{i=1}^{n}\Big\{\hat{\nu}_n^{\mathrm{RF}}(1,W_i) - \hat{\nu}_n^{\mathrm{RF}}(0,W_i) + \frac{A_i}{\hat{\pi}_n(W_i;P)}(Y_i - \nu(W_i,1;P)) \\ - \frac{1-A_i}{1-\hat{\pi}_n(W_i;P)}(Y_i - \nu(W_i,0;P))\Big\},$$

where $\hat{\pi}_n$ is an estimator of π.

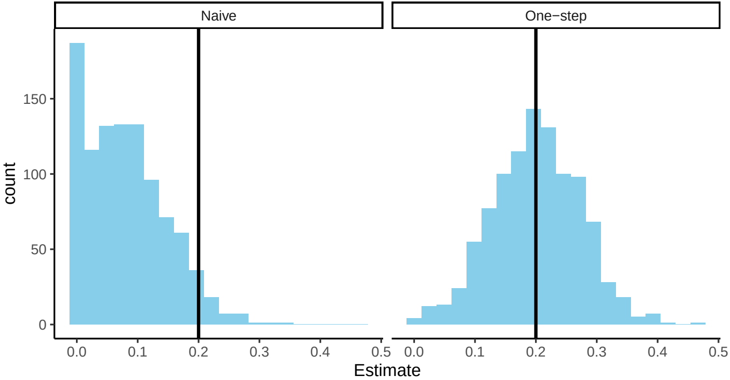
Simulation study

1. Generate a simulated data set of 800 individuals with 10 covariates.
2. Fit a ridge regression for the outcome using cross-validation to select to penalty parameter
3. Use the fit from step 2 to and the g-formula to estimate the ATE. We refer to this estimator as `naive`.
4. Fit another ridge regression for the propensity score using cross-validation to select to penalty parameter
5. Use the efficient influence function and the estimators from step 2 and 4 to target/debias the estimator calculated in step 3. We refer to this estimator as `one-step`.

Repeat steps 1-5 1000 times to obtain 1000 samples of the `naive` estimator and the `one-step` estimator.

Examine performance of the two estimators by comparing these 1000 random samples to the true ATE.

Result of simulation study



Statistical inference

To quantify uncertainty we calculate confidence intervals. This procedure relies on an estimate of the asymptotic variance of our estimator.

Naive ML approach

- The use of cross-validation (and other data-adaptive algorithms) complicates calculation of the asymptotic variance of the `naive` estimator.
- Bootstrap is not feasible in practice (and works poorly with cross-validation)

One step estimator

- Closed form expression for the asymptotic variance which can be estimated based on the models we have already fitted.

References

O. Hines, O. Dukes, K. Diaz-Ordaz, and S. Vansteelandt. Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 76(3): 292–304, 2022.

E. H. Kennedy. Semiparametric theory and empirical processes in causal inference. In *Statistical causal inferences and their applications in public health research*, pages 141–167. Springer, 2016.

E. H. Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*, 2022.

A. J. Wikkelsø, S. Hjortøe, T. A. Gerds, A. M. Møller, and J. Langhoff-Roos. Prediction of postpartum blood transfusion–risk factors and recurrence. *The Journal of Maternal-Fetal & Neonatal Medicine*, 27(16):1661–1667, 2014.