Augmented Inverse Probability Weighting (AIPW)

Here we briefly outline the construction of our estimators and required assumptions. For a didactic explanation of these kinds of methods see Schuler & van der Laan (Schuler and van der Laan, 2022) .

Setup Let Y, D, A and X represent our outcome (possibly missing), an indicator of outcome observed, an indicator of treatment, and a vector of covariates. Let $Y(a)$ represent the potential outcome that would obtain if treatment were forced to $A = a$. Define $\psi_a^* = E[Y(a)]$ to be the counterfactual population average outcomes. The causal average treatment effect (ATE) is defined as $\psi^* = \psi_1^* - \psi_0^*$. For notational convenience define

- $\mu_a(X) = E[Y(a)|X]$ (conditional potential outcome means)
- $\pi_a(X) = P\{A = a|X\}$ (propensity scores)
- $\delta(X) = P\{D = 1|X\}$ (conditional probability of being observed)

Identification The measured outcome $Y = D \times Y(A)$ depends on which treatment is given and on the observation indicator (without loss of generality we arbitrarily say $Y = 0$ when it goes unobserved). We can identify the causal ATE under the following assumptions:

- $P\{A = a|X = x\} > 0 \ \ \forall\ a, x$ (treatment positivity)
- $(Y(a) \perp A)|X$ (treatment unconfounded)
- $P\{D = 1|A = a, X = x\} > 0 \ \ \forall a, x$ (missingness positivity)
- $D \perp Y(a)|X, A$ (missingness unconfounded)

Given these, standard conditioning arguments show that $E[Y|D = 1, A = a, X] = \mu_a(X)$. Define the statistical counterfactual mean $\psi_a = E[E[Y|D = 1, A = 1, X]]$ and define the *statistical* ATE to be $\psi = \psi_1 - \psi_0$ (which is equal to the *causal* ATE ψ^* when our identifying assumptions hold).

Inference Standard derivation techniques (Kennedy, 2022) show that the efficient influence function for ψ_a is

$$\phi_a = \frac{D1_a(A)}{\delta(X)\pi_a(X)}(Y - \mu_a(X)) + (\mu_a(X) - \psi_a)$$

and the efficient influence function for ψ is therefore $\phi_1 - \phi_0$.

We can thus obtain an efficient estimating equations (i.e. AIPW-style) estimator

$$\hat{\psi}_a = \frac{1}{n}\sum_i \frac{D_i 1_a(A_i)}{\hat{\delta}(X_i)\hat{\pi}_a(X_i)}(Y_i - \hat{\mu}_a(X_i)) + \hat{\mu}_a(X_i)$$

where the hat quantities are cross-fit estimates of their true counterparts. We obtain our estimate of the ATE by taking $\hat{\psi} = \hat{\psi}_1 - \hat{\psi}_0$.

By standard theory, this estimator is asymptotically normal with asymptotic sampling variance $V[\phi]$. We can therefore obtain a consistent estimate $\hat{\sigma}^2_\infty$ by taking the empirical sample variance of the estimated influence function $\hat{\phi} = \hat{\phi}_1 - \hat{\phi}_0$ where

$$\hat{\phi}_a = \frac{D1_a(A)}{\hat{\delta}(X)\hat{\pi}_a(X)}(Y - \hat{\mu}_a(X)) + (\hat{\mu}_a(X) - \hat{\psi}_a).$$

An estimate of the finite-sample sampling variance is therefore $\hat{\sigma}^2 = \hat{\sigma}^2_\infty / n$, which we can use to build confidence intervals (e.g. 95% CI is $\hat{\psi} \pm 1.96 \times \hat{\sigma}$) and compute p-values (use a Z-test to compare the estimated effect to the null $H_0 : \hat{\psi} \sim \mathcal{N}(0, \hat{\sigma}^2)$).

Difference in ATEs Let G represent a moderator of interest, which is one of the covariates in $X = [X_1 \ldots G \ldots X_p]$. Let $\psi^*_{a,g} = E[Y(a)|G = g]$ and define a difference in causal ATEs between two groups $G = 0$ and $G = 1$ to be $(\psi^*_{a=1,g=1} - \psi^*_{a=0,g=1}) - (\psi^*_{a=1,g=0} - \psi^*_{a=0,g=0})$. This transparently and nonparametrically represents a measure of the extent to which G moderates the causal effect of A on Y.

Identification proceeds along the same lines as the standard ATE. Again using standard techniques we obtain that the efficient influence function for this estimand is $\phi = (\phi_{a=1,g=1} - \phi_{a=0,g=1}) - (\phi_{a=1,g=0} - \phi_{a=0,g=0})$ where

$$\phi_{a,g} = \frac{1_g(G)}{\gamma_g} \left[\frac{D1_a(A)}{\delta(X)\pi_a(X)}(Y - \mu_a(X)) + (\mu_a(X) - \psi_{a,g}) \right]$$

and we define the population group proportion $\gamma_g = P\{G = g\}$. The appropriate efficient estimating equations estimator and inference follow immediately in similar fashion to the above.

References

1. Kennedy, E.H., 2022. *Semiparametric doubly robust targeted double machine learning: A review*. Carnegie Mellon University, Pittsburgh, PA.
2. Schuler, A., van der Laan, M., 2022. *Introduction to modern causal inference.* https://alejandroschuler.github.io/mci/4a08c1afbfb545f0bbdc4668de4da329.html.