

SEMIPARAMETRIC INFERENCE AND DOUBLE ROBUSTNESS

Inverse Probability Weighting

We have derived an IPW estimator of

$$\mu(1) = \mathbb{E}_{Y|Z}^{\mathcal{E}}[Y|Z = 1]$$

in the binary exposure case as

$$\tilde{\mu}(1) = \frac{1}{n} \sum_{i=1}^n \frac{Z_i Y_i}{e(X_i)}.$$

The corresponding estimating equation is

$$\sum_{i=1}^n \left\{ \frac{Z_i Y_i}{e(X_i)} - \mu(1) \right\} = 0$$

and the corresponding influence function for this estimator is

$$\varphi_{IPW}(X, Y, Z) = \frac{ZY}{e(X)} - \mu(1).$$

Inverse Probability Weighting (cont.)

By design this is an unbiased estimator with variance

$$\text{Var}_{X,Y,Z}^{\mathcal{O}} \left[\frac{ZY}{e(X)} \right] = \mathbb{E}_{X,Y,Z}^{\mathcal{O}} \left[\frac{Z^2 Y^2}{\{e(X)\}^2} \right] - \{\mu(1)\}^2$$

divided by n , where

$$\mathbb{E}_{X,Y,Z}^{\mathcal{O}} \left[\frac{Z^2 Y^2}{\{e(X)\}^2} \right] = \mathbb{E}_{X,Z}^{\mathcal{O}} \left[\frac{Z \mu_2(X, 1)}{\{e(X)\}^2} \right] = \mathbb{E}_X^{\mathcal{O}} \left[\frac{\mu_2(X, 1)}{e(X)} \right]$$

where

$$\mu_2(X, 1) = \mathbb{E}_{Y|X,Z}^{\mathcal{O}} \left[Y^2 | X, Z = 1 \right].$$

Inverse Probability Weighting (cont.)

We now attempt to justify this from a semiparametric perspective, using the potential outcome notation.

If we were able to observe the counterfactual outcomes under the two cases $z = 1$ and $z = 0$, then the optimal estimator based on these 'full data' would be

$$\frac{1}{n} \sum_{i=1}^n Y_i(1)$$

based on the estimating equation

$$\sum_{i=1}^n (Y_i(1) - \mu(1)) = 0$$

Inverse Probability Weighting (cont.)

It can be shown that the class of *observed data* influence functions is the collection of all functions $\varphi(X, Y, Z)$ for which

$$\mathbb{E}[\varphi(X, Y, Z)|Y(1), Y(0), X] = Y(1) - \mu(1)$$

that is, $\varphi(X, Y, Z)$ takes the form

$$\varphi(X, Y, Z) = h(X, Y, Z) + L(X, Y, Z)$$

where $h(X, Y, Z)$ and $L(X, Y, Z)$ are any functions satisfying

$$\mathbb{E}[h(X, Y, Z)|Y(1), Y(0), X] = Y(1) - \mu(1)$$

$$\mathbb{E}[L(X, Y, Z)|Y(1), Y(0), X] = 0.$$

One such h function is given by the IPW estimating influence function

$$h(X, Y, Z) = \frac{ZY}{e(X)} - \mu(1)$$

provided $e(X) > 0$. For $L(X, Y, Z)$, as Z is binary, we must have that

$$L(X, Y, Z) = (1 - Z)L_0(X, Y) + ZL_1(X, Y).$$

Then

$$\begin{aligned}\mathbb{E}[L(X, Y, Z)|Y(1), Y(0), X] &= \\ &= \mathbb{E}[(1 - Z)L_0(X, Y) + ZL_1(X, Y)|Y(1), Y(0), X] \\ &= (1 - e(X))L_0(X, Y(0)) + e(X)L_1(X, Y(1)).\end{aligned}$$

This follows if we assume that conditional on $Y(z)$ the expectation of Y is $Y(z)$ independently of $Y(1 - z)$ for $z = 0, 1$.

This assumption is known as SUTVA (*stable unit treatment value assumption*).

Therefore

$$L_0(X, Y(0)) = -\frac{e(X)}{1 - e(X)} L_1(X, Y(1))$$

This identity demonstrates that $L_0(., .)$ and $L_1(., .)$ must in fact be functions of X alone, and hence

$$L_0(X) = -\frac{e(X)}{1 - e(X)} L_1(X)$$

so therefore $L(., ., .)$ must take the form

$$\begin{aligned} L(X, Y, Z) &= \left(-(1 - Z) \frac{e(X)}{1 - e(X)} + Z \right) L_1(X) \\ &= \left(\frac{Z - e(X)}{1 - e(X)} \right) L_1(X) \end{aligned}$$

for arbitrary $L_1(X)$. Combining terms, we have that the corresponding space of functions can be written

$$\Lambda \equiv \{(Z - e(X))h(X) : h(X) \text{ arbitrary}\}.$$

This space effectively corresponds to the nuisance tangent space.

Inverse Probability Weighting (cont.)

We now consider projecting $\varphi_{IPW}(X, Y, Z)$ onto Λ and taking the ‘residual’ in order to get the optimal (efficient) influence function. The projection is determined by a particular h , h_0 say, that satisfies the orthogonality condition

$$\mathbb{E} [\{\varphi_{IPW}(X, Y, Z) - (Z - e(X))h_0(X)\} \{(Z - e(X))h(X)\}] = 0$$

for all $h(X) \in \Lambda$.

Now, as Z is binary, the form of any influence function $\varphi(X, Y, Z)$ is

$$\varphi(X, Y, Z) = (1 - Z)\varphi_0(X, Y) + Z\varphi_1(X, Y)$$

say. By linearity, we may consider the two components separately when constructing the projection.

Consider first $Z\varphi_1(X, Y)$. We have

$$\begin{aligned}\mathbb{E} [\{Z\varphi_1(X, Y) - (Z - e(X))h_0(X)\} \{(Z - e(X))h(X)\}] \\ = \mathbb{E} [Z(Z - e(X))h(X)\varphi_1(X, Y)] - \\ \mathbb{E} [(Z - e(X))^2 h_0(X)h(X)]\end{aligned}$$

Inverse Probability Weighting (cont.)

For the first term, by iterated expectation over Y given X and Z , then over Z given X

$$\begin{aligned}\mathbb{E}[Z(Z - e(X))h(X)\varphi_1(X, Y)] \\ = \mathbb{E}[e(X)(1 - e(X))h(X)\mathbb{E}[\varphi_1(X, Y)|Z = 1, X]]\end{aligned}$$

Similarly, again by iterated expectation over Z given X , we have for the second term

$$\mathbb{E}\left[(Z - e(X))^2 h_0(X)h(X)\right] = \mathbb{E}[e(X)(1 - e(X))h_0(X)h(X)] .$$

Therefore, the orthogonality condition can be re-written

$$\mathbb{E} [e(X)(1 - e(X))h(X) \{ \mathbb{E}[\varphi_1(X, Y)|Z = 1, X] - h_0(X) \}] = 0$$

This must hold for arbitrary $h(X)$, and therefore as by assumption $e(X) > 0$ with probability 1, we must have that

$$\mathbb{E}[\varphi_1(X, Y)|Z = 1, X] - h_0(X) \equiv 0$$

that is

$$h_0(X) \equiv \mathbb{E}[\varphi_1(X, Y)|Z = 1, X].$$

Therefore the projection of the influence function $Z\varphi_1(X, Y)$ is

$$(Z - e(X))\mathbb{E}[\varphi_1(X, Y)|Z = 1, X].$$

A similar result holds for the projection of the second component $(1 - Z)\varphi_0(X, Y)$.

Inverse Probability Weighting (cont.)

For the IPW influence function $\varphi_{IPW}(X, Y, Z)$, consider the projection of

$$Z\varphi_1(X, Y) = Z \frac{Y}{e(X)}$$

as the projection of the constant $\mu(1)$ is merely the constant itself – in this case $\varphi_0(X, Y) \equiv 0$.

In this case, projection of $Z\varphi_1(X, Y)$ is therefore

$$\begin{aligned}(Z - e(X))\mathbb{E} \left[\frac{Y}{e(X)} \middle| Z = 1, X \right] &= \frac{(Z - e(X))}{e(X)} \mathbb{E} [Y|Z = 1, X] \\ &= \frac{(Z - e(X))}{e(X)} \mu(X, 1)\end{aligned}$$

say.

Inverse Probability Weighting (cont.)

We therefore have that the efficient influence function is the ‘residual’

$$\begin{aligned}\varphi^{\text{eff}}(X, Y, Z) &= \frac{ZY}{e(X)} - \frac{(Z - e(X))}{e(X)}\mu(X, 1) - \mu(1) \\ &= \frac{Z(Y - \mu(X, 1))}{e(X)} + \mu(X, 1) - \mu(1).\end{aligned}$$

which corresponds to the AIPW estimator

$$\tilde{\mu}_{AIPW}(1) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{Z_i}{e(X_i)} (Y_i - \mu(X_i, 1)) + \mu(X_i, 1) \right\}.$$

Equivalent results hold for estimating $\mu(0)$

$$\tilde{\mu}_{AIPW}(0) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1 - Z_i}{1 - e(X_i)} (Y_i - \mu(X_i, 0)) + \mu(X_i, 0) \right\},$$

and for $\delta = \mu(1) - \mu(0)$

$$\tilde{\delta}_{AIPW} = \tilde{\mu}_{AIPW}(1) - \tilde{\mu}_{AIPW}(0).$$

Inverse Probability Weighting (cont.)

Writing the estimator for $\mu(z)$

$$\begin{aligned}\tilde{\mu}_{AIPW}(z) &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\mathbb{1}_{\{z\}}(Z_i)}{\{e(X_i)\}^z \{1 - e(X_i)\}^{1-z}} (Y_i - \mu(X_i, z)) + \mu(X_i, z) \right\} \\ &\equiv \frac{1}{n} \sum_{i=1}^n \{W_{zi}(Y_i - \mu(X_i, z)) + \mu(X_i, z)\}\end{aligned}$$

say, we may also construct the alternative estimator

$$\tilde{\mu}_{AIPW}(z) = \sum_{i=1}^n W_{zi}^* (Y_i - \mu(X_i, z)) + \frac{1}{n} \sum_{i=1}^n \mu(X_i, z)$$

where

$$W_{zi}^* = \frac{W_{zi}}{\sum_{j=1}^n W_{zj}}.$$

Inverse Probability Weighting (cont.)

These estimators are *doubly robust*, that is, unbiased if at least one of the proposed models for

- $e(X)$
- $\mu(X, Z)$

is correctly specified.

In practice, parametric models for these two components must be specified, that is,

$$e(X) \equiv e(X; \alpha) \qquad \mu(X, Z) \equiv \mu(X, Z; \beta).$$

Provided consistent estimation of α or β can be carried out, the resulting estimators of $\mu(z)$ are consistent also.