

Part 2

The Propensity Score

Constructing a balanced sample

Recall the randomized trial setting in the case of a binary exposure.

- ▶ we obtain a random sample of size n of individuals from the target population, and measure their X values;
- ▶ according to some random assignment procedure, we intervene to assign treatment Z to individuals, and measure their outcome Y ;
- ▶ the link between X and Z is broken by the random allocation.

Constructing a balanced sample

Recall that this procedure led to the valid use of the estimators of the ATE based on (1) and (2).

The important feature of the randomized study is that we have, for confounders X (indeed all predictors)

$$f_{X|Z}(x|1) \equiv f_{X|Z}(x|0) \quad \text{for all } x,$$

or equivalently, in the case of a binary confounder,

$$\Pr[X = 1|Z = 1] = \Pr[X = 1|Z = 0].$$

Constructing a balanced sample

The distribution of X is balanced across the two exposure groups; this renders direct comparison of the outcomes possible. Probabilistically, X and Z are independent.

In a non-randomized study, there is a possibility that the two exposure groups are not balanced

$$f_{X|Z}(x|1) \neq f_{X|Z}(x|0) \quad \text{for some } x,$$

or in the binary case

$$\Pr[X = 1|Z = 1] \neq \Pr[X = 1|Z = 0].$$

If X influences Y also, then this imbalance renders direct comparison of outcomes in the two groups impossible.

Constructing a balanced sample

Whilst global balance may not be present, it may be that ‘local’ balance, within certain strata within the sample, may be present.

That is, for $x \in \mathcal{S}$ say, we might have balance; within $\mathcal{S}$, X is independent of Z .

$$f_{X|Z:\mathcal{S}}(x|1 : x \in \mathcal{S}) = f_{X|Z}(x|0 : x \in \mathcal{S})$$

Then, for individuals who have X values in $\mathcal{S}$, there is the possibility of direct comparison of the treated and untreated groups.

We might then restrict attention to causal statements relating to stratum $\mathcal{S}$.

Constructing a balanced sample

For discrete confounders, we might consider defining strata where the X values are precisely matched, and then comparing treated and untreated within those strata.

Consider matching strata $\mathcal{S}_1, \dots, \mathcal{S}_K$. We would then be able to compute the ATE by noting that

$$\mathbb{E}[Y(1) - Y(0)] = \sum_{k=1}^K \mathbb{E}[Y(1) - Y(0) | X \in \mathcal{S}_k] \Pr[X \in \mathcal{S}_k]$$

- ▶ $\mathbb{E}[Y(1) - Y(0) | X \in \mathcal{S}_k]$ may be estimated nonparametrically from the data by using (1) or (2) for data restricted to have $x \in \mathcal{S}_k$.
- ▶ $\Pr[X \in \mathcal{S}_k]$ may be estimated using the empirical proportion of x that lie in $\mathcal{S}_k$.

Constructing a balanced sample

For continuous confounders, we might consider the same strategy: consider matching strata $\mathcal{S}_1, \dots, \mathcal{S}_K$. Then the formula

$$\mathbb{E}[Y(1) - Y(0)] = \sum_{k=1}^K \mathbb{E}[Y(1) - Y(0) | X \in \mathcal{S}_k] \Pr[X \in \mathcal{S}_k]$$

still holds. However

- we must assume a model for how $\mathbb{E}[Y(1) - Y(0) | X \in \mathcal{S}_k]$ varies with x for $x \in \mathcal{S}_k$.

In both cases, inference is restricted to the set of X space contained in

$$\bigcup_{k=1}^K \mathcal{S}_k.$$

Constructing a balanced sample

In the continuous case, the above calculations depend on the assumption that the treatment effect is similar for x values that lie ‘close together’ in predictor (confounder) space. However

- I. Unless we can achieve exact matching, then the term ‘close together’ needs careful consideration.
- II. If X is moderate or high-dimensional, there may be insufficient data to achieve adequate matching to facilitate the estimation of the terms

$$\mathbb{E}[Y(1) - Y(0)|X \in \mathcal{S}_k];$$

recall that we need a large enough sample of treated and untreated subjects in stratum $\mathcal{S}_k$.

Nevertheless, matching in this fashion is an important tool in causal comparison.

Balance via the propensity score

We now introduce the important concept of the propensity score that facilitates causal comparison via a balancing approach.

Recall that our goal is to mimic the construction of the randomized study that facilitates direct comparison between treated and untreated groups. We may not be able to achieve this globally, but possibly can achieve it locally in strata of X space.

The question is how to define these strata.

Balance via the propensity score

Recall that in the binary exposure case, balance corresponds to being able to state that within $\mathcal{S}$, X is independent of Z :

$$f_{X|Z:\mathcal{S}}(x|1 : x \in \mathcal{S}) = f_{X|Z}(x|0 : x \in \mathcal{S})$$

This can be achieved if $\mathcal{S}$ is defined in terms of a statistic, $e(X)$ say. That is, we consider the conditional distribution

$$f_{X|Z,e(X)}(x|z, e)$$

and attempt to ensure that, given $e(X) = e$, Z is independent of X , so that within strata of $e(X)$, the treated and untreated groups are directly comparable.

Balance via the propensity score

By Bayes theorem, for $z = 0, 1$, we have that

$$f_{X|Z,e(X)}(x|z, e) = \frac{f_{Z|X,e(X)}(z|x, e)f_{X|e(X)}(x|e)}{f_{Z|e(X)}(z|e)} \quad (4)$$

Now, as Z is binary, we must be able to write the density in the denominator as

$$f_{Z|e(X)}(z|e) = p(e)^z(1 - p(e))^{1-z} \quad z \in 0, 1$$

where $p(e)$ is a probability, a function of the fixed value e , and where $p(e) > 0$.

Balance via the propensity score

Therefore, in order to make the density $f_{X|Z,e(X)}(x|z, e)$ functionally independent of z , and achieve the necessary independence, we must have that

$$f_{Z|X,e(X)}(z|x, e) = p(e)^z(1 - p(e))^{1-z} \quad z \in 0, 1$$

also. But $e(X)$ is a function of X , so automatically we have that

$$f_{Z|X,e(X)}(z|x, e) \equiv f_{Z|X}(z|x).$$

Therefore, we require that

$$f_{Z|X}(z|x) = f_{Z|X}(z|x, e) = p(e)^z(1 - p(e))^{1-z} \equiv f_{Z|e(X)}(z|e)$$

for all relevant z, x .

Balance via the propensity score

This can be achieved by choosing the statistic

$$e(x) = f_{Z|X}(1|x) = \Pr_{Z|X}[Z = 1|x]$$

and setting $p(\cdot)$ to be the identity function, so that

$$f_{Z|X}(z|x) = e^z(1 - e)^{1-z} \quad z = 0, 1.$$

More generally, choosing $e(x)$ to be some monotone transform of $f_{Z|X}(1|x)$ would also achieve the same balance.

The corresponding random variable $e(X)$ defines the strata via which the causal calculation can be considered.

Balance via the propensity score

The function $e(x)$ defined in this way is the propensity score¹. It has the following important properties

- (i) as seen above, it is a balancing score; conditional on $e(X)$, X and Z are independent.
- (ii) it is a scalar quantity, irrespective of the dimension of X .
- (iii) in noting that for balance we require that

$$f_{Z|X}(z|x) \equiv f_{Z|e(X)}(z|e),$$

the above construction demonstrates that if $\tilde{e}(X)$ is another balancing score, then $e(X)$ is a function of $\tilde{e}(X)$;

- that is, $e(X)$ is the ‘coarsest’ balancing score.

¹ see Rosenbaum & Rubin (1983), *Biometrika*

Evaluating the propensity score

To achieve balance we must have

$$e(X) = \Pr[Z = 1|X]$$

correctly specified; that is, for confounders X , we must precisely specify the model $\Pr[Z = 1|X]$.

- ▶ If X comprises entirely discrete components, then we may be able to estimate $\Pr[Z = 1|X]$ entirely nonparametrically, and satisfactorily if the sample size is large enough.
- ▶ If X has continuous components, it is common to use parametric modelling, with

$$e(X; \alpha) = \Pr[Z = 1|X; \alpha].$$

Balance then depends on correct specification of this model.

Unconfoundedness given the propensity score

The assumption of ‘no unmeasured confounders’ amounts to assuming that the potential outcomes are jointly independent of exposure assignment given the confounders, that is

$$\{Y(0), Y(1)\} \perp Z \mid X$$

that is, in terms of densities

$$\begin{aligned} f_{Y(z), Z|X}(y, z|x) &= f_{Y(z)|X}(y|x) f_{Z|X}(z|x) \\ &= f_{Y|Z, X}(y|z, x) f_{Z|X}(z|x). \end{aligned}$$

Unconfoundedness given the propensity score

Now consider conditioning on propensity score $e(X)$ instead of X : we have by factorization that

$$f_{Y(Z),Z|e(X)}(y,z|e) = \frac{1}{f_{e(X)}(e)} \int_{\mathcal{S}_e} f_{Y(Z),Z,X}(y,z,x) \, dx$$

where $\mathcal{S}_e$ is the set of x values

$$\mathcal{S}_e \equiv \{x : e(x) = e\}$$

that yield a propensity score value equal to the value e .

Unconfoundedness given the propensity score

Now we have by unconfoundedness given X that

$$f_{Y(z),Z,X}(y,z,x) = f_{Y(z)|X}(y|x)f_{Z|X}(z|x)f_X(x)$$

and on the set $\mathcal{S}_e$, we have

$$f_{Z|X}(z|x) = e^z(1-e)^{1-z} \equiv f_{Z|e(X)}(z|e).$$

Unconfoundedness given the propensity score

Therefore, recalling the $\mathcal{S}_e$ is defined via the fixed constant e ,

$$\begin{aligned}\int_{\mathcal{S}_e} f_{Y(z),Z,X}(y,z,x) \, dx &= \int_{\mathcal{S}_e} f_{Y(z)|X}(y|x) e^z (1-e)^{1-z} f_X(x) \, dx \\ &= e^z (1-e)^{1-z} \int_{\mathcal{S}_e} f_{Y(z)|X}(y|x) f_X(x) \, dx \\ &= f_{Z|e(X)}(z|e) f_{Y(z)|e(X)}(y|e).\end{aligned}$$

Hence

$$f_{Y(z),Z|e(X)}(y,z|e) = \frac{1}{f_{e(X)}(e)} f_{Z|e(X)}(z|e) f_{Y(z)|e(X)}(y|e)$$

and so

$$Y(z) \perp Z \mid e(X) \quad \text{for all } z.$$

Estimation using the propensity score

We now consider the same stratified estimation strategy as before, but using $e(X)$ instead X to stratify.

Consider strata $\mathcal{S}_1, \dots, \mathcal{S}_K$ defined via $e(X)$. In this case, recall that

$$0 < e(X) < 1$$

so we might consider an equal quantile partition, say using quintiles.

Then we have

$$\mathbb{E}[Y(1) - Y(0)] = \sum_{k=1}^K \mathbb{E}[Y(1) - Y(0) | e(X) \in \mathcal{S}_k] \Pr[e(X) \in \mathcal{S}_k]$$

still holds approximately if the $\mathcal{S}_k$ are small enough.

Estimation using the propensity score

This still requires us to be able to estimate

$$\mathbb{E}[Y(1) - Y(0)|e(X) \in \mathcal{S}_k]$$

which requires us to have a sufficient number of treated and untreated individuals with $e(X) \in \mathcal{S}_k$ to facilitate the ‘direct comparison’ within this stratum.

If the expected responses are constant across the stratum, the formulae (1) and (2) may be used.

Matching

The derivation of the propensity score indicates that it may be used to construct matched individuals or groups that can be compared directly.

- ▶ if two individuals have precisely the same value of $e(x)$, then they are exactly matched;
- ▶ if one of the pair is treated and the other untreated, then their outcomes can be compared directly, as any imbalance between their measured confounder values has been removed by the fact that they are matched on $e(x)$;
- ▶ this is conceptually identical to the standard procedure of matching in two-group comparison.

Matching

For an exactly matched pair (i_1, i_0) , treated and untreated respectively, the quantity

$$y_{i_1} - y_{i_0}$$

is an unbiased estimate of the ATE

$$\mathbb{E}[Y(1) - Y(0)];$$

more typically we might choose m such matched pairs, usually with different $e(x)$ values across pairs, and use the estimate

$$\frac{1}{m} \sum_{i=1}^m (y_{i_1} - y_{i_0})$$

Matching

Exact matching is difficult to achieve, therefore we more commonly attempt to achieve approximate matching

- ▶ May match one treated to M untreated ($1 : M$ matching)
- ▶ caliper matching;
- ▶ nearest neighbour/kernel matching;
- ▶ matching with replacement.

Most standard software packages have functions that provide automatic matching using a variety of methods.

Beyond binary exposures

The theory developed above extends beyond the case of binary exposures.

Recall that we require balance to proceed with causal comparisons; essentially, with strata defined using X or $e(X)$, the distribution of X should not depend on Z .

We seek a scalar statistic such that, conditional on the value of that statistic, X and Z are independent. In the case of general exposures, we must consider balancing scores that are functions of both Z and X .

Beyond binary exposures

For a balancing score $b(Z, X)$, we require that

$$X \perp Z \mid b(Z, X).$$

We denote $B = b(Z, X)$ for convenience.

Consider the conditional distribution $f_{Z|X,B}(z|x,b)$: we wish to demonstrate that

$$f_{Z|X,B}(z|x,b) = f_{Z|B}(z|b) \quad \text{for all } z, x, b.$$

That is, we require that B completely characterizes the conditional distribution of Z given X .

Beyond binary exposures

This can be achieved by choosing the statistic

$$b(z, x) = f_{Z|X}(z|x)$$

in line with the choice in the binary case.

The balancing score defined in this way is termed the

Generalized Propensity Score

which is a balancing score for general exposures.

Beyond binary exposures

Note, however, that this choice that mimics the binary exposure case is not the only one that we might make. The requirement

$$f_{Z|X,B}(z|x,b) = f_{Z|B}(z|b)$$

for all relevant z, x is met if we define $b(Z, X)$ to be any sufficient statistic that characterizes the conditional distribution of Z given X .

It may be possible, for example, to choose functions purely of X .

Beyond binary exposures

Example: Normally distributed exposures

Suppose that continuous valued exposure Z is distributed as

$$Z|X = x \sim \text{Normal}(x\alpha, \sigma^2)$$

for row-vector confounder X . We have that

$$f_{Z|X}(z|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (z - x\alpha)^2 \right\}$$

Beyond binary exposures

Example: Normally distributed exposures

We might therefore choose

$$b(Z, X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (Z - X\alpha)^2 \right\}.$$

However, the linear predictor

$$b(X; \alpha) = X\alpha$$

also characterizes the conditional distribution of Z given X ; if we know that $x\alpha = b$, then

$$Z|X = x \equiv Z|B = b \sim \text{Normal}(b, \sigma^2).$$

In both cases, parameters α are to be estimated.

Beyond binary exposures

The generalized propensity score inherits all the properties of the standard propensity score;

- ▶ it induces balance;
- ▶ if the potential outcomes and exposure are independent given X under the unconfoundedness assumption, they are also independent given $b(Z, X)$.

However, how exactly to use the generalized propensity score in causal adjustment for continuous exposures is not clear.

Propensity Score Regression

Up to this point we have considered using the propensity score for stratification, that is, to produce directly comparable groups of treated and untreated individuals.

Causal comparison can also be carried out using regression techniques: that is, we consider building an estimator of the APO by regressing the outcome on a function of the exposure and the propensity score.

Regressing on the propensity score is a means of controlling the confounding.

Propensity Score Regression

If we construct a model

$$\mathbb{E}[Y|Z = z, b(Z, X) = b] = \mu(z, b)$$

then because potential outcomes $Y(\mathbf{z})$ and Z are independent given $b(Z, X)$, we have

$$\mathbb{E}[Y(\mathbf{z})|b(Z, X) = b] = \mathbb{E}[Y|Z = \mathbf{z}, b(\mathbf{z}, X) = b] = \mu(\mathbf{z}, b)$$

and therefore

$$\mathbb{E}[Y(\mathbf{z})] = \mathbb{E}_{b(\mathbf{z}, X)}[\mathbb{E}[Y|Z = \mathbf{z}, b(\mathbf{z}, X)]] = \mathbb{E}_{b(\mathbf{z}, X)}[\mu(\mathbf{z}, b(\mathbf{z}, X))].$$

Propensity Score Regression

That is, to estimate the APO, we might

- ▶ fit the propensity score model $b(Z, X)$ to the observed exposure and confounder data by regressing Z on X ;
- ▶ fit the conditional outcome model $\mu(z, b)$ using the fitted $b(Z, X)$ values, $\hat{b}(z_i, x_i)$;
- ▶ for each z of interest, estimate the APO by

$$\frac{1}{n} \sum_{i=1}^n \hat{\mu}(z, \hat{b}(z, x_i)).$$

Propensity Score Regression

If the propensity function $b(Z, X) \equiv b(X)$, we proceed similarly, and construct a model

$$\mathbb{E}[Y|Z = z, b(X) = b] = \mu(z, b)$$

then

$$\mathbb{E}[Y(z)|b(X) = b] = \mathbb{E}[Y|Z = z, b(X) = b] = \mu(z, b)$$

and therefore

$$\mathbb{E}[Y(z)] = \mathbb{E}_{b(X)}[\mathbb{E}[Y|Z = z, b(X)]] = \mathbb{E}_{b(X)}[\mu(z, b(X))].$$

Propensity Score Regression

To estimate the APO:

- ▶ fit the propensity score model $b(X)$ to the observed exposure and confounder data by regressing Z on X ;
- ▶ fit the conditional outcome model $\mu(z, b)$ using the fitted $b(X)$ values, $\hat{b}(x_i)$;
- ▶ for each z of interest, estimate the APO by

$$\frac{1}{n} \sum_{i=1}^n \hat{\mu}(z, \hat{b}(x_i)).$$

Example: Binary Exposure

We specify

- ▶ $e(X; \alpha) = \Pr[Z = 1|X, \alpha]$ then regress Z on X to obtain $\hat{\alpha}$ and fitted values $\hat{e}(X) \equiv e(X; \hat{\alpha})$.
- ▶ $\mathbb{E}[Y|Z = z, e(X) = e; \beta] = \mu(z, e; \beta)$ and estimate this model by regressing y_i on z_i and $\hat{e}(x_i)$. For example, we might have that

$$\mathbb{E}[Y|Z = z_i, e(X_i) = e_i; \beta] = \beta_0 + \beta_1 z_i + \beta_2 e_i.$$

This returns $\hat{\beta}$.

We finally compute the predictions under this model, and average them to obtain the APO estimate

$$\hat{\mathbb{E}}[Y(\mathbf{z})] = \frac{1}{n} \sum_{i=1}^n \mu(\mathbf{z}, \hat{e}(x_i); \hat{\beta}).$$

Example: Continuous Exposure

In the case of a continuous exposure, we have a parametric probability density for the exposure

$$b(Z, X; \alpha) = f_{Z|X}(Z|X; \alpha)$$

for which we estimate α by regressing Z on X to obtain $\hat{\alpha}$ and fitted values $\hat{b}(Z, X) \equiv b(Z, X; \hat{\alpha})$.

Then we specify outcome model

$$\mathbb{E}[Y|Z = z, b(X) = b; \beta] = \mu(z, b; \beta)$$

and estimate this model by regressing y_i on z_i and $\hat{b}(z_i, x_i)$. Again, we might have that

$$\mathbb{E}[Y|Z = z_i, b(Z_i, X_i) = b_i; \beta] = \beta_0 + \beta_1 z_i + \beta_2 b_i.$$

This returns $\hat{\beta}$.

Example: Binary Exposure

We then compute the predictions under this model, and average them to obtain the APO estimate

$$\widehat{\mathbb{E}}[Y(\mathbf{z})] = \frac{1}{n} \sum_{i=1}^n \mu(\mathbf{z}, \widehat{b}(\mathbf{z}, x_i); \widehat{\beta}).$$

Note that here the propensity terms that enter into μ are computed at the target $\mathbf{z}$ values

not the observed exposure values.

Propensity Score Regression

These procedures require us to make two modelling choices:

- ▶ the propensity model, $b(Z, X)$ or $b(X)$;
- ▶ the outcome mean model $\mu(z, b)$.

Unfortunately, both models must be correctly specified for consistent inference.

Misspecification of the outcome mean model will lead to bias; this model needs to capture the outcome to exposure and propensity function relationship correctly.

Weighting approaches

For a causal quantity of interest, we focus on the APO

$$\mathbb{E}[Y(\mathbf{z})] = \int y f_{Y(\mathbf{z}),X}(y, x) \, dy \, dx$$

that is, the average outcome, over the distribution of the confounders and predictors, if we hypothesize that the intervention sets the exposure to $\mathbf{z}$.

We now study methods that utilize the components already described, including the propensity score, but in a different fashion;

- ▶ instead of accounting for confounding by balancing through matching, we aim to achieve balance via weighting

Average potential outcome

If we could intervene at the population level to set $Z = z$ for all individuals independently of their X value, we might rewrite this as

$$\mathbb{E}[Y(z)] = \int y \mathbb{1}_z(z) f_{Y(z),X}(y, x) \, dy \, dz \, dx$$

and take a random sample from the population with density

$$\mathbb{1}_z(z) f_{Y(z),X}(y, x) \equiv \mathbb{1}_z(z) f_{Y|Z,X}(y|z, x) f_X(x).$$

We could then construct the ‘Monte Carlo’ estimator

$$\widehat{\mathbb{E}}[Y(z)] = \frac{1}{n} \sum_{i=1}^n Y_i$$

as $Z_i = z$ for all i .

Average potential outcome: Experimental study

In a randomized (experimental) study, suppose that exposure $Z = \mathbf{z}$ is assigned with probability determined by $f_Z(\mathbf{z})$. Then

$$\begin{aligned}\mathbb{E}[Y(\mathbf{z})] &= \frac{\int y \mathbb{1}_{\mathbf{z}}(z) f_{Y(\mathbf{z}),X}(y, x) f_Z(z) \, dy \, dz \, dx}{\int \mathbb{1}_{\mathbf{z}}(z) f_{Y(\mathbf{z}),X}(y, x) f_Z(z) \, dy \, dz \, dx} \\ &= \frac{\int y \mathbb{1}_{\mathbf{z}}(z) f_{Y|Z,X}(y|z, x) f_X(x) f_Z(z) \, dy \, dz \, dx}{\int \mathbb{1}_{\mathbf{z}}(z) f_Z(z) \, dz}\end{aligned}$$

Average potential outcome: Experimental study

This suggests the Monte Carlo estimators

$$\widehat{\mathbb{E}}[Y(\mathbf{z})] = \frac{\sum_{i=1}^n \mathbb{1}_{\mathbf{z}}(Z_i)Y_i}{\sum_{i=1}^n \mathbb{1}_{\mathbf{z}}(Z_i)} \quad \text{or} \quad \widehat{\mathbb{E}}[Y(\mathbf{z})] = \frac{1}{nf_Z(\mathbf{z})} \sum_{i=1}^n \mathbb{1}_{\mathbf{z}}(Z_i)Y_i.$$

Average potential outcome: Observational study

Denote by $P_{\mathcal{E}}$ the probability measure for samples drawn under the experimental measure corresponding to the density

$$f_{Y|Z,X}^{\mathcal{E}}(y|z, x) f_X^{\mathcal{E}}(x) f_Z^{\mathcal{E}}(z).$$

Now consider the case where the data arise from the observational (non-experimental) measure $P_{\mathcal{O}}(\mathrm{d}y, \mathrm{d}z, \mathrm{d}x)$.

We have

$$\begin{aligned}\mathbb{E}[Y(\mathbf{z})] &= \frac{1}{f_Z^{\mathcal{E}}(\mathbf{z})} \int y \mathbf{1}_{\mathbf{z}}(z) P_{\mathcal{E}}(\mathrm{d}y, \mathrm{d}z, \mathrm{d}x) \\ &= \frac{1}{f_Z^{\mathcal{E}}(\mathbf{z})} \int y \mathbf{1}_{\mathbf{z}}(z) \underbrace{\frac{P_{\mathcal{E}}(\mathrm{d}y, \mathrm{d}z, \mathrm{d}x)}{P_{\mathcal{O}}(\mathrm{d}y, \mathrm{d}z, \mathrm{d}x)}}_{\textcircled{1}} P_{\mathcal{O}}(\mathrm{d}y, \mathrm{d}z, \mathrm{d}x).\end{aligned}$$

Average potential outcome: Observational study

In terms of densities ① becomes

$$\begin{aligned} & \frac{f_{Y|Z,X}^{\mathcal{E}}(y|z,x) f_Z^{\mathcal{E}}(z) f_X^{\mathcal{E}}(x)}{f_{Y|Z,X}^{\mathcal{O}}(y|z,x) f_{Z|X}^{\mathcal{O}}(z|x) f_X^{\mathcal{O}}(x)} \\ &= \frac{f_{Y|Z,X}^{\mathcal{E}}(y|z,x)}{f_{Y|Z,X}^{\mathcal{O}}(y|z,x)} \times \frac{f_Z^{\mathcal{E}}(z)}{f_{Z|X}^{\mathcal{O}}(z|x)} \times \frac{f_X^{\mathcal{E}}(x)}{f_X^{\mathcal{O}}(x)} \end{aligned}$$

- for the first term, we assume that

$$\frac{f_{Y|Z,X}^{\mathcal{E}}(y|z,x)}{f_{Y|Z,X}^{\mathcal{O}}(y|z,x)} = 1 \quad \text{for all } y, z, x;$$

this is essentially a no unmeasured confounders assumption.

- the third term equals 1 by assumption.

Experimental vs observational sampling

The second term

$$\frac{f_Z^{\mathcal{E}}(z)}{f_{Z|X}^{\mathcal{O}}(z|x)}$$

constitutes a weight that appears in the integral that yields the desired APO; the term

$$\frac{1}{f_{Z|X}^{\mathcal{O}}(z|x)}$$

accounts for the imbalance that influences the confounding and measures the difference between the observed sample and a hypothetical idealized randomized sample.

Estimation

This suggests the (nonparametric) estimators

$$\widehat{\mathbb{E}}[Y(\mathbf{z})] = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{1}_{\mathbf{z}}(Z_i) Y_i}{f_{Z|X}^{\mathcal{O}}(Z_i|X_i)} \quad (\text{IPW0})$$

which is unbiased, or

$$\widehat{\mathbb{E}}[Y(\mathbf{z})] = \frac{\sum_{i=1}^n \frac{\mathbb{1}_{\mathbf{z}}(Z_i) Y_i}{f_{Z|X}^{\mathcal{O}}(Z_i|X_i)}}{\sum_{i=1}^n \frac{\mathbb{1}_{\mathbf{z}}(Z_i)}{f_{Z|X}^{\mathcal{O}}(Z_i|X_i)}} \quad (\text{IPW})$$

which is consistent, each provided $f_{Z|X}^{\mathcal{O}}(\cdot|\cdot)$ correctly specifies the conditional density of Z given X for all (z, x) .

Inverse weighting and the propensity score

Note 7.

Inverse weighting constructs a pseudo-population in which there are no imbalances on confounders between the exposure groups. The pseudo-population is balanced, as required for direct comparison of treated and untreated groups.

Note 8.

The term in the denominator of the components of the sum is $f_{Z|X}^{\mathcal{O}}(Z_i|X_i)$, that is, the probability model that captures the conditional model for Z_i given X_i . If Z_i is binary, this essentially reduces to

$$e(X_i)^{Z_i}(1 - e(X_i))^{1-Z_i}$$

where $e(.)$ is the propensity score as defined previously.

Positivity

Note 9.

It is evident that we must have

$$f_{Z|X}^{\mathcal{O}}(Z_i|X_i) > 0$$

with probability 1 for this calculation to be valid.

This is commonly assumed, and is termed the positivity or experimental treatment assignment assumption.

Note 10.

The inverse weighting procedure can also be justified from a weighted likelihood perspective.

Estimation via Augmentation

We may write

$$\mathbb{E}[Y(\mathbf{z})] = \mathbb{E}[Y(\mathbf{z}) - \mu(\mathbf{z}, X)] + \mathbb{E}[\mu(\mathbf{z}, X)]$$

where $\mu(\mathbf{z}, x) = \mathbb{E}[Y|Z = \mathbf{z}, X = x]$.

We have the alternate estimator

$$\hat{\mathbb{E}}[Y(\mathbf{z})] = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{1}_{\mathbf{z}}(Z_i)(Y_i - \mu(Z_i, X_i))}{f_{Z|X}^{\mathcal{O}}(Z_i|X_i)} + \frac{1}{n} \sum_{i=1}^n \mu(\mathbf{z}, X_i) \quad (\text{AIPW})$$

and

$$\text{Var}_{\text{AIPW}} \leq \text{Var}_{\text{IPW}}.$$

Furthermore, (AIPW) is doubly robust (i.e. consistent even if one of $f_{Z|X}^{\mathcal{O}}(\mathbf{z}|x)$ and $\mu(\mathbf{z}, x)$ is misspecified).

Properties under misspecification

Implementing (AIPW) relies on specification of the components

$$f_{Z|X}^{\mathcal{O}}(z|x) \quad \mu(z, x).$$

Suppose that, in reality, the correct specifications are

$$\tilde{f}_{Z|X}(z|x) \quad \tilde{\mu}(z, x).$$

Then the bias of (AIPW) is

$$\mathbb{E} \left[\frac{(f_{Z|X}^{\mathcal{O}}(z|X) - \tilde{f}_{Z|X}(z|X))(\mu(z, X) - \tilde{\mu}(z, X))}{f_{Z|X}^{\mathcal{O}}(z|X)} \right] \quad (5)$$

which is zero if

$$f_{Z|X}^{\mathcal{O}} \equiv \tilde{f}_{Z|X} \quad \text{or} \quad \mu(z, x) \equiv \tilde{\mu}(z, x)$$

that is, (AIPW) is doubly robust.

Properties under misspecification

Asymptotically, for estimators that are sample averages, the variance of the estimator converges to zero under standard conditions.

Therefore in large samples it is the magnitude of the bias as given by (5) that determines the quality of the estimator.

- equation (5) demonstrates that misspecification in the functions $\mu(z, x)$ and $f_{Z|X}^{\mathcal{O}}$ play equal roles in the bias.

Parametric modelling: two-stage approach

In the formulation, the nonparametric models

$$f_{Z|X}^{\mathcal{O}}(z|x) \quad \mu(z, x)$$

are commonly replaced by parametric models

$$f_{Z|X}^{\mathcal{O}}(z|x; \alpha) \quad \mu(z, x; \beta) = \int y f_{Y|Z,X}^{\mathcal{O}}(y|z, x; \beta) \, dy.$$

Parameters (α, β) are estimated from the observed data by regressing

- ▶ Stage I: Z on X using $(z_i, x_i), i = 1, \dots, n$,
- ▶ Stage II: Y on (Z, X) using $(y_i, z_i, x_i), i = 1, \dots, n$

and using plug-in version of (IPW) and (AIPW).

The estimated propensity score

Note 11.

It is possible to conceive of situations where the propensity-type model

$$f_{Z|X}^{\mathcal{O}}(z|x) \quad \text{or} \quad f_{Z|X}^{\mathcal{O}}(z|x; \alpha)$$

is known precisely and does not need to be estimated.

This is akin to the randomized study where the allocation probabilities are fixed by the experimenter. It can be shown that using estimated quantities

$$\hat{f}_{Z|X}^{\mathcal{O}}(z|x) \quad \text{or} \quad f_{Z|X}^{\mathcal{O}}(z|x; \hat{\alpha})$$

yields lower variances for the resulting estimators than if the known quantities are used.

Alternative view of augmentation

Scharfstein et al. (1999), Bang & Robins (2005) write the estimating equation yielding (AIPW) as

$$\sum_{i=1}^n \frac{\mathbb{1}_{\mathbf{z}}(Z_i)}{f_{Z|X}^{\mathcal{O}}(Z_i|X_i)} (Y_i - \mu(Z_i, X_i)) + \sum_{i=1}^n \{\mu(\mathbf{z}, X_i) - \mu(\mathbf{z})\} = 0$$

Alternative view of augmentation

The first summation is a component of the score obtained when performing OLS regression for Y with mean function

$$\mu(z, x) = \mu_0(z, x) + \epsilon \frac{\mathbb{1}_z(z)}{f_{Z|X}^{\mathcal{O}}(z|x)}.$$

and $\mu_0(z, x)$ is a conditional mean model, and ϵ is a regression coefficient associated with the derived predictor

$$\frac{\mathbb{1}_z(z)}{f_{Z|X}^{\mathcal{O}}(z|x)}.$$

Alternative view of augmentation

Therefore, estimator (AIPW) can be obtained by regressing Y on (X, Z) for fixed z using the mean specification $\mu(z, x)$, and forming the estimator

$$\frac{1}{n} \sum_{i=1}^n \left\{ \mu_0(Z_i, X_i) + \hat{\epsilon} \frac{\mathbf{1}_z(Z_i)}{f_{Z|X}^{\mathcal{O}}(Z_i|X_i)} \right\}.$$

In a parametric model setting, this becomes

$$\frac{1}{n} \sum_{i=1}^n \left\{ \mu_0(Z_i, X_i; \hat{\beta}) + \hat{\epsilon} \frac{\mathbf{1}_z(Z_i)}{f_{Z|X}^{\mathcal{O}}(Z_i|X_i; \hat{\alpha})} \right\}$$

where α is estimated from Stage (I), and β is estimated along with ϵ in the secondary regression.

Augmentation and contrasts

The equivalent to (AIPW) for estimating the ATE for binary treatment

$$\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

is merely $\hat{\mathbb{E}}[Y(1)] - \hat{\mathbb{E}}[Y(0)]$ or

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left[\frac{\mathbb{1}_1(Z_i)}{f_{Z|X}^{\mathcal{O}}(1|X_i)} - \frac{\mathbb{1}_0(Z_i)}{f_{Z|X}^{\mathcal{O}}(0|X_i)} \right] (Y_i - \mu(Z_i, X_i)) \\ + \frac{1}{n} \sum_{i=1}^n \{ \mu(1, X_i) - \mu(0, X_i) \}. \end{aligned}$$

Augmentation and contrasts

Therefore we can repeat the above argument and base the contrast estimator on the regression of Y on (X, Z) using the mean specification

$$\mu(z, x) = \mu_0(z, x) + \epsilon \left[\frac{\mathbb{1}_1(z)}{f_{Z|X}^{\mathcal{O}}(1|x)} - \frac{\mathbb{1}_0(z)}{f_{Z|X}^{\mathcal{O}}(0|x)} \right]$$

or

$$\mu(z, x) = \mu_0(z, x) + \left[\epsilon_1 \frac{\mathbb{1}_1(z)}{f_{Z|X}^{\mathcal{O}}(1|x)} - \epsilon_0 \frac{\mathbb{1}_0(z)}{f_{Z|X}^{\mathcal{O}}(0|x)} \right].$$