SUB-CLASSIFICATION WITH PROPENSITY SCORE

The reason we cannot estimate the ATE using $\bar{Y}_1 - \bar{Y}_0$ when there are confounders is because the distribution of the confounders is different in the treatment and control groups.

In sub-classification, we formed blocks using the propensity score, looked at the average difference within block, and then weighted each block using the proportion of the sample within a block. We can rewrite the sub-classification estimator of the ATE as

$$\hat{\text{ATE}} = \sum_{s=1}^{S} \frac{n_s}{n}(\bar{Y}_{1s} - \bar{Y}_{0s}) = n^{-1} \sum_{s=1}^{S} [\frac{n_s}{n_{1s}} Z_{si'} Y_{si'} - \frac{n_s}{n_{0s}}(1 - Z_{si'}) Y_{si'}]$$

which makes it evident that the sub-classification estimator essentially re-weights the treated observations within a block by n_s/n_{1s}, the inverse of the treatment probability within the block, and the control observations by n_s/n_{0s}, the inverse of the within block probability of not receiving treatment.

WEIGHTING WITH PROPENSITY SCORE

This suggests more generally weighting the treated observations by the propensity score and the untreated observations by the probability of not receiving treatment, and forming the weighting estimator:

$$\hat{\text{ATE}} = n^{-1} \sum_{i=1}^{n} [\frac{Z_i Y_i}{e(\underline{X}_i)} - \frac{(1 - Z_i) Y_i}{1 - e(\underline{X}_i)}]$$

Propensity score weighting reduces to propensity score sub-classification when the propensity score within block s is actually n_{1s}/n_s, from which it is evident that sub-classification is a crude version of weighting, where the weights are replaced by an approximate propensity score applied to the observations in block s, $s = 1, \ldots, S$. This demonstrates that weighting using the propensity score creates distributions in the treatment group and the control group that are the same, and also that weighting is theoretically superior to sub-classification, as the latter generally leads to a biased estimate of the ATE.

WEIGHTING WITH PROPENSITY SCORE

To see a more formal justification for weighting, we note that

$$Z_i Y_i = Z_i(Z_i Y_i(1) + (1 - Z_i) Y_i(0) = Z_i Y_i(1)$$

and thus

$$\begin{aligned} E[\frac{Z_i Y_i}{e(\underline{X}_i)}] &= EE[\frac{Z_i}{e(\underline{X}_i)} Y_i(1) \mid \underline{X}_i] \\ &= E[(\frac{E(Z_i)}{e(\underline{X}_i)} \mid \underline{X}_i) E(Y_i(1) \mid \underline{X}_i)] \\ &= E(1 \times E(Y_i(1) \mid \underline{X}_i) \\ &= E(Y_i(1)) \end{aligned}$$

That is, the weighting estimator will be unbiased for the ATE.

WEIGHTING WITH PROPENSITY SCORE

In practice, however, a number of issues arise. First, for any given sample, neither $n^{-1}\sum_{i=1}^{n} Z_i/e(\underline{X}_i)$ nor $n^{-1}\sum_{i=1}^{n}(1-Z_i)/(1-e(\underline{X}_i))$ will generally add to 1. This is easily fixed by normalizing

$$\hat{\text{ATE}} = n^{-1}\sum_{i=1}^{n}[\frac{Z_i Y_i}{e(\underline{X}_i)} - \frac{(1-Z_i)Y_i}{1-e(\underline{X}_i)}]$$

as

$$\hat{\text{ATE}}' = \sum_{i=1}^{n}\frac{Z_i Y_i}{e(\underline{X}_i)}/\sum_{i=1}^{n}\frac{Z_i}{e(\underline{X}_i)} - \sum_{i=1}^{n}\frac{(1-Z_i)Y_i}{1-e(\underline{X}_i)}/\sum_{i=1}^{n}\frac{1-Z_i}{1-e(\underline{X}_i)}$$

WEIGHTING WITH PROPENSITY SCORE

More seriously, in observational studies, the propensity score is typically not known and must be estimated. In practice, $e(\underline{X}_i)$ in $\hat{\text{ATE}}$ or $\hat{\text{ATE}}'$ is replaced with an estimate $\hat{e}(\underline{X}i)$.

If the model for the propensity score is misspecified, this can create severe bias in the estimated ATE, especially at large and small values of $\hat{e}(\underline{X})$. For example, if the propensity score is .05, and is estimated as .01 instead, the treatment observations are weighted 5 times more heavily than should be the case. Similarly, if the propensity score is .95 and is estimated as .99, the control observations are weighted 5 times more heavily than should be the case. This reduces the advantage of weighting over sub-classification.

In both sub-classification and weighting, when the propensity score is very small (large), there are often an insufficient number of treatment (control) group observations, which creates additional

WEIGHTING WITH PROPENSITY SCORE

The weighting estimator is also readily adopted to estimation of the ATT: $E(Y(1) - Y(0) \mid Z = 1)$. First, $\bar{Y}_1$ is unbiased and consistent for $E(Y(1) \mid Z = 1)$.

To estimate $E(Y(0) \mid Z = 1)$ we need to re-weight the control group observations to the distribution of the propensity score in the treatment group. Assuming also that the propensity score is estimated, the estimator is:

$$\hat{ATT} = \bar{Y}_1 - \sum_{i=1}^{n} \frac{(1 - Z_i)\hat{e}(\underline{X}_i)Y_i}{1 - \hat{e}(\underline{X}_i)} \Big/ \sum_{i=1}^{n} \frac{(1 - Z_i)\hat{e}(\underline{X}_i)}{1 - \hat{e}(\underline{X}_i)}$$

Weighting estimators can be implemented using least squares. For example, to estimate $\hat{\text{ATE}}$ the weights are $(\hat{e}(\underline{X}_i))^{-1/2}$ and for the untreated observations, $(1 - \hat{e}(\underline{X}_i))^{-1/2}$.

WEIGHTING WITH PROPENSITY SCORE

One might also add covariates to the regression. A seeming disadvantage is that it is then necessary to estimate two models, one for the propensity score, one for the regression. However, Robins and Ritov (1997) showed that if either the model for the propensity score or the regression function was specified properly, the estimator for the ATE is consistent. They refer to this property as double robustness.

Weighting estimators that use the propensity score also feature prominently in the literature on longitudinal causal inference, where the goal is to estimate the effect of a treatment regimen $Z_1 = z_1, Z_2 = z_2, \ldots, Z_T = z_T$ vs. an alternative regimen $Z_1^* = z_1^*, Z_2^* = z_2^*, \ldots, Z_T^* = z_T^*$, where Z_t (Z_t^*) denotes the treatment assignment at time $t, t = 1, \ldots, T$. As an example, one might want to ask whether an outcome is improved by taking a pill twice daily vs. once a day in the morning over a period of $T/2$ days.