ESTIMATION WITH PROPENSITY SCORE

Propensity score used to estimate treatment effects in observational studies in four ways:

1. regression
2. subclassification
3. weighting
4. matching

ESTIMATION WITH PROPENSITY SCORE

Assuming unconfoundedness, linear regression leads to biased and inconsistent estimates of ATE if not correctly specified.

While it may be difficult to estimate high-dimensional regression non-parametrically, previous lesson indicates it is sufficient to regress the outcomes on the propensity score and treatment assignment, then average over the distribution of the propensity score. If regression model is correctly specified, this yields consistent estimate of ATE.

In practice propensity score is unknown. Therefore, this strategy requires estimation and specification of two models: one for the outcome, one for the propensity score. We shall see it is more common to use the propensity score as an additional regressor along with the covariates X, or in conjunction with the regression function to obtain a "doubly robust" estimator.

SUB-CLASSIFICATION WITH PROPENSITY SCORE

Sub-classification forms blocks s based on the covariate values, and then the data are analyzed as in a block randomized experiment. Within each block s, there are $n_s = n_{0s} + n_{1s}$ units, where n_{0s} is the number of control units and n_{1s} is the number of treated units.

Let $\bar{Y}_{1s}$ denote the average value of the outcome for the treated units in stratum s, and let $\bar{Y}_{1s} - \bar{Y}_{0s}$ denote the estimate of the ATE for that stratum. The ATE is estimated as:

$$\hat{\text{ATE}} = \sum_{s=1}^{S} \frac{n_s}{n} (\bar{Y}_{1s} - \bar{Y}_{0s})$$

Replacing n_s/n with the within stratum treated proportions n_{1s}/n_1 yields an estimate of the ATT.

SUB-CLASSIFICATION WITH PROPENSITY SCORE

Rosenbaum and Rubin show that for two units with the same value of the propensity score, $e(\underline{X})$, one treated and one not, the difference in outcomes is unbiased for ATE($e(\underline{X})$).

This justifies sub-classification based on the propensity score. As the propensity score is a many-to-one function of the original covariates, this may help to address the problems noted above.

SUB-CLASSIFICATION WITH PROPENSITY SCORE

To implement sub-classification on the propensity score, which is typically unknown in observational studies, we replace it with an estimate $\hat{e}(\underline{X})$. Let $0 < \hat{e}_1 < \ldots < \hat{e}_{S-1} < 1$, and form S strata $(0, \hat{e}_1), \ldots, [\hat{e}_{S-1}, 1)$, where each unit i belongs to the stratum containing $\hat{e}(\underline{x}_i)$.

Grouping together observations with different propensity scores introduces bias. If S is large enough, the estimator should have minimal bias. However, each subclass will then have fewer observations, which increases the variability of the estimator.

It is difficult to give general analytic results, which will depend on the form of the regression functions in both treatment and control groups, and also on how balanced the covariates are in the subclasses. But in practice, the use of 5-10 subclasses is often recommended and for some reasonable examples, this can reduce the bias (relative to using the estimator $\bar{Y}_1 - \bar{Y}_0$) by more than 90

SUB-CLASSIFICATION WITH PROPENSITY SCORE

Suppose one starts by estimating the propensity score model using logistic regression or a probit model. (see Imbens and Rubin 2015 for more details)

In practice, one may want to include not only main effects of covariates, but a number of interactions as well.

Other methods may also be used to estimate the propensity score model, e.g., generalized boosted models (McCaffrey et al. 2004).

SUB-CLASSIFICATION WITH PROPENSITY SCORE

After model has been fitted, it often happens that at "higher" ("lower") values of the estimated propensity score, there are no, or relatively "few", control (treatment) group observations. This is referred to as insufficient "overlap".

One way to deal with this is to choose $\hat{e}_1$ and/or $\hat{e}_{S-1}$ so that the intervals $(0, \hat{e}_1)$ and/or $[\hat{e}_{S-1}, 1)$ contain an "adequate" number of observations from both the treatment group and the control group. But then the first and/or last sub-class might include cases from the treatment and control groups that are quite different, resulting in increased estimation bias.

Another approach is to trim the sample by excluding observations with estimated values below and above some thresholds and estimate the treatment effect of interest on this region of "common support". For example, Imbens and Rubin (2015) exclude observations with estimated propensity scores less than

SUB-CLASSIFICATION WITH PROPENSITY SCORE

Next, a pre-specified number of subclasses is formed using propensity score intervals of equal length and in each interval, a test is conducted to assess whether or not the mean propensity score is different in the treatment and control groups.

In those intervals where the null hypothesis of no difference is rejected, the interval is split until the null is not rejected or until further splitting would result in the situation where an interval fails to contain both treatment and control observations.

SUB-CLASSIFICATION WITH PROPENSITY SCORE

After the number and spacing of intervals has been determined, using the fact that the propensity score is a balancing score, the covariate distributions should be balanced across the treatment and control groups in each sub-class.

To check this, Imbens and Rubin (2015) recommend using the "normalized" difference:

$$\sum_{s=1}^{S} \frac{n_s}{n} \frac{\bar{X}_{1ks} - \bar{X}_{0ks}}{((s_{0k}^2 + s_{1k}^2)/2)^{1/2}}$$

where

$\bar{X}_{1ks}$ ($\bar{X}_{0ks}$) is the treatment (control) group mean for covariate k in subclass s,

$\bar{X}_{1k}$ ($\bar{X}_{0k}$) is the treatment (control) group mean for covariate k,

$s_{1k}^2 = (n_1 - 1)^{-1} \sum_{i=1}^{n} Z_i (X_{ik} - \bar{X}_{1k})^2$ is the estimate of the variance in the treatment group, and

s_{0k}^2 is defined analogously in the estimated variance in the control

SUB-CLASSIFICATION WITH PROPENSITY SCORE

For example, the propensity score model might be estimated using non-parametric logistic regression or machine learning methods for classification, e.g., generalized boosted models, as in McCaffrey et al. (2004).

Recall that in a completely randomized experiment, we saw that using linear regression to adjust for differences between the treatment and control groups resulted in an unbiased estimator of the ATE with smaller variance than the estimator $\bar{Y}_1 - \bar{Y}_0$.

Since the randomized block experiment is a randomized experiment within blocks and sub-classification is an attempt to mimic a block randomized experiment, this suggests using linear regression within blocks to adjust for differences in balance between covariates in the treatment and control groups.

SUB-CLASSIFICATION WITH PROPENSITY SCORE

Thus, consider the following regression

$$Y_i(Z_i) = \alpha^*_{s(i)} + \tau^*_{s(i)} Z_i + \underline{\beta}^{*\prime}_{s(i)} \underline{X}_i + v_i$$

where

$s(i)$ is the sub-class to which i has been allocated

$E(v_i \mid Z_i = z, \underline{X}_i = \underline{x}, s(i)) = 0$

$\alpha^*_{s(i)}$ and $\underline{\beta}^{*\prime}_{s(i)}$ are the intercept and regression coefficients in subclass $s(i)$, and

$\tau^*_{s(i)}$ is the ATE in sub-class $s(i)$

SUB-CLASSIFICATION WITH PROPENSITY SCORE

As before, using the fact that the OL S residuals and weighted residuals sum to 0, it follows:

$$\bar{Y}_{1s} - \bar{Y}_{0s} = \hat{\tau}_s^* + \hat{\underline{\beta}}_s^* (\underline{\bar{X}}_{1s} - \underline{\bar{X}}_{0s})$$

The ATE is estimated as $\sum_{s=1}^{S} \frac{n_s}{n} \hat{\tau}_s^*$

The ATT as $\sum_{s=1}^{S} \frac{n_{1s}}{n_1} \hat{\tau}_s^*$, where $n_1 = \sum_{s=1}^{S} n_{1s}$

The variance of the estimated ATE as $\sum_{s=1}^{S} (\frac{n_s}{n})^2 \hat{V}(\hat{\tau}_s^*)$, where $\hat{V}(\hat{\tau}_s^*)$ is the estimated variance of the OLS estimator of $\hat{\tau}_s^*$

The variance of the estimated ATT as $\sum_{s=1}^{S} (\frac{n_{1s}}{n_1})^2 \hat{V}(\hat{\tau}_s^*)$