IDS 702: Module 6.7

Causal inference using propensity scores

Dr. Olanrewaju Michael Akande

CAUSAL INFERENCE USING PROPENSITY SCORES

Propensity score analysis (in observational studies) typically involves two stages:

- Stage 1. Estimate the propensity score: by a logistic regression model or machine learning methods.

- Stage 2. Given the estimated propensity score, estimate the causal effects through one of these methods:

  - Stratification
  - Matching
  - Regression
  - Weighting (which we will not cover)
  - Mixed combinations of the above

The general idea is to use the estimated propensity scores to correct for lack of balance between groups, then go on to estimate the causal effect using the "balanced" data.

STAGE 1: ESTIMATING THE PROPENSITY SCORE

STAGE 1: ESTIMATING THE PROPENSITY SCORE

- The main purpose of estimating propensity score is to ensure overlap and balance of covariates between treatment groups, instead of "finding a perfect fit" of propensity score.

- As long as the important covariates are balanced, model overfitting is not a concern; underfitting can be a problem however.

- Essentially any balancing score (not necessarily propensity score) would be good enough for practical use.

STAGE 1: ESTIMATING THE PROPENSITY SCORE

- A standard procedure for estimating propensity scores includes:

  1. initial fit;

  2. discarding outliers (with too large or too small propensity scores);

  3. check covariate balance; and

  4. re-fit if necessary.

STAGE 1: ESTIMATING THE PROPENSITY SCORE

- **Step 1.** Estimate propensity score using a logistic regression:

$$W_i | X_i \sim \text{Bernoulli}(\pi_i); \quad \log\left(\frac{\pi_i}{1 - \pi_i}\right) = X_i \boldsymbol{\beta}.$$

Include all covariates in this initial model or do a stepwise selection on the covariates and interactions to get an initial estimate of the propensity scores. That is,

$$\hat{e}^0(X_i) = \frac{e^{X_i \hat{\boldsymbol{\beta}}}}{1 + e^{X_i \hat{\boldsymbol{\beta}}}}.$$

Can also use machine learning methods.

IDS 702

STAGE 1: ESTIMATING THE PROPENSITY SCORE

- **Step 2.** Check overlap of propensity score between treatment groups. If necessary, discard the observations with non-overlapping propensity scores.

- **Step 3.** Assess balance given by initial model in Step 1.

- **Step 4.** If one or more covariates are seriously unbalanced, include some of their higher order terms and/or interactions to re-fit the propensity score model and repeat Steps 1-3, until most covariates are balanced.

Note: There are situations where some important covariates will still not be completely balanced after repeated trials. Then they should be taken into account in Stage 2 (outcome stage) of propensity score analysis.

IDS 702

STAGE 1: ESTIMATING THE PROPENSITY SCORE

- In practice, balance checking in the PS estimation stage can be done via sub-classification/stratification, matching or weighting.

  - sub-classification/stratification: check the balance of all important covariates within K blocks of $\hat{e}^0(X_i)$ based on its quantiles.

  - matching: check the balance of all important covariates in the matched sample.

  - in weighting, check the balance of the weighted covariates between treatment and control groups.

- The workflow is the same: fit initial model, check balance (sub-classification, matching or weighting), then refit.

PROPENSITY SCORE ANALYSIS WORKFLOW
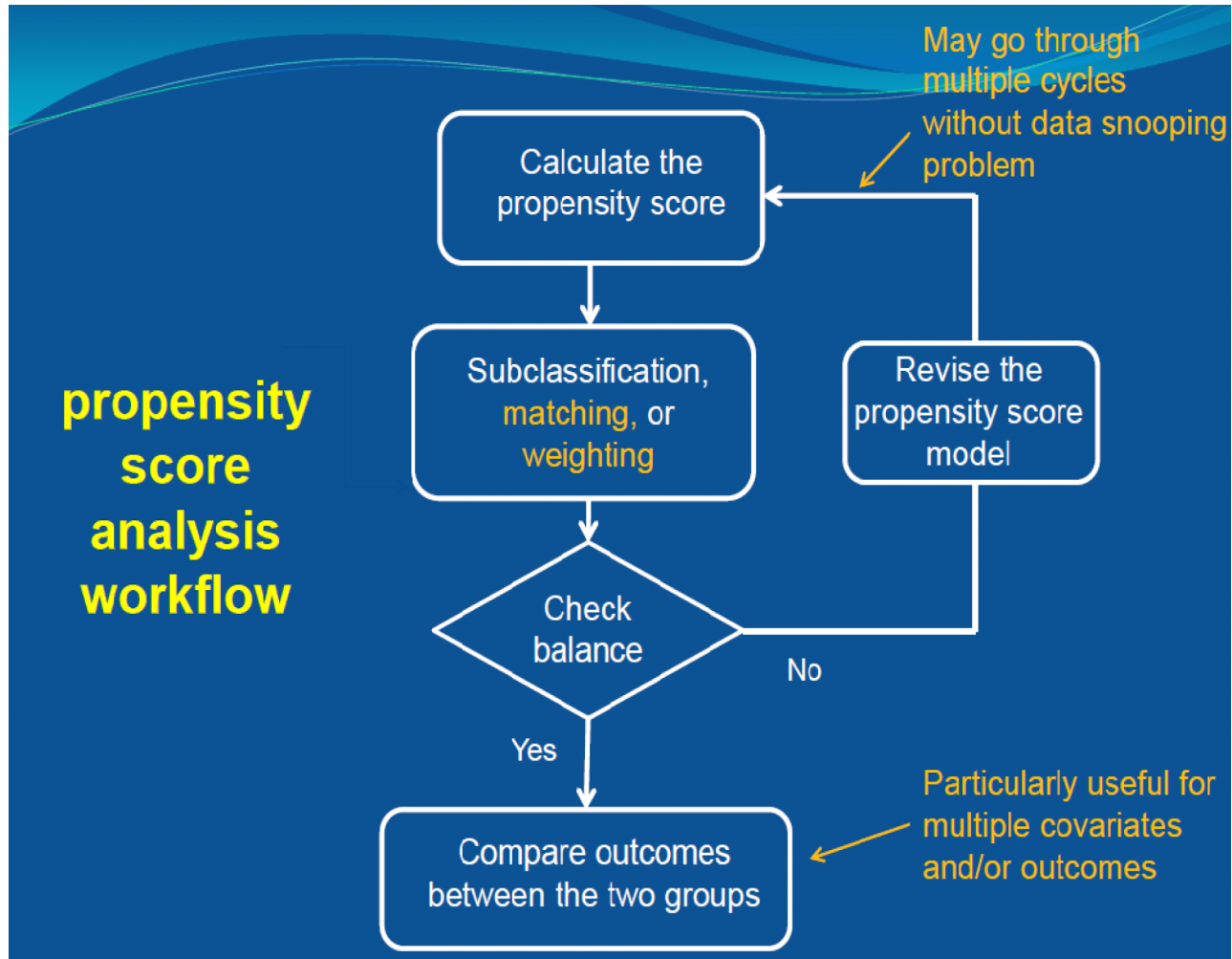
STAGE 2: ESTIMATING THE CAUSAL EFFECT

STAGE 2: STRATIFICATION

- Given the estimated propensity score, we can estimate the causal estimands through sub-classification/stratification, weighting or matching.

- Let's start with stratification.

- Recall that the result of 5 strata of a single covariate removes 90% bias.

- Stratification using propensity score as the summary score should have approximately the same effects.

STAGE 2: STRATIFICATION

- Divide the subjects in to K strata by the corresponding quantiles of the estimated propensity scores.

- ATE: estimate ATE within each stratum and then average by the block size. That is,

$$\hat{\tau}^{ATE} = \sum_{k=1}^{K} \left(\bar{Y}_{k,1} - \bar{Y}_{k,0} \right) \frac{N_{k,1} + N_{k,0}}{N},$$

  with $N_{k,1}$ and $N_{k,0}$ being the numbers of units in class k under treated and control, respectively.

- ATT: weight within-block ATE by proportion of treated units $N_{k,1}/N_1$.

- A variance estimator for $\hat{\tau}^{ATE}$ is

$$\mathbb{Var}\left[\hat{\tau}^{ATE} \right] = \sum_{k=1}^{K} \left(\mathbb{Var}[\bar{Y}_{k,1}] - \mathbb{Var}[\bar{Y}_{k,0}] \right) \left(\frac{N_{k,1} + N_{k,0}}{N} \right)^2,$$

  or use bootstrap.

IDS 702

PROPENSITY SCORE STRATIFICATION: REMARKS

- 5 blocks is usually not enough, consider higher number such as 10.

- Stratification is a coarsened version of matching.

- Empirical results from real applications and situations: usually not as good as matching or weighting.

- Good for cases with extreme outliers (smoothing): less sensitive, but also less efficient.

- Can be combined with regression: first estimate causal effects using regression within each block and then average the within-subclass estimates.

IDS 702

Stage 2: matching

- In propensity score matching, potential matches are compared using (estimated) propensity score.

- 1-to-n closest neighbor matching is common when the control group is large compared to treatment group.

- In most software packages, the default is actually 1-to-1 closest neighbor matching.

- Pros: robust, matched pairs (so you can do within pair analysis).

- Sometimes, dimension reduction via the propensity score may be too drastic, recent methods advocate matching on the multivariate covariates directly.

- Nonetheless, this is what we will focus on for our minimum wage data.

IDS 702

STAGE 2: REGRESSION

- Remember the key propensity score property:

$$Y_i(0), Y_i(1) \perp W_i | X_i \;\; \Rightarrow \;\; Y_i(0), Y_i(1) \perp W_i | e(X_i)$$

- Idea: in a regression estimator, adjusting for $e(X)$ instead of the whole X; thus in regression models of $Y(w)$ use $e(X)$ as the single predictor.

- Clearly, modeling $\Pr(Y(w)|\hat{e}(X))$ is simpler than modeling $\Pr(Y(w)|X)$; effectively more data to estimate essential parameters due to the dimension reduction.

- However,

  - we lose interpretation of the effects of individual covariates, e.g. age, sex; and

  - reduction to the one-dimensional propensity score may be too drastic.

STAGE 2: REGRESSION

- Idea: instead of using the estimated $\hat{e}(X)$ as the single predictor, use it as an additional predictor in the model. That is, $\mathbb{Pr}(Y(w)|X, \hat{e}(X))$.

- Turns out that $\mathbb{Pr}(Y(w)|X, \hat{e}(X))$ gives both efficiency and robustness.

- Also, if we are unable to achieve full balance on some of the predictors, using $\mathbb{Pr}(Y(w)|X, \hat{e}(X))$ will help further control for those unbalance predictors.

- Empirical evidences (e.g. simulations) support this claim.

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!