Matching Algorithms for Causal Inference with Multiple Treatments

Anthony D. Scotina

International Conference on Health Policy Statistics

January 12, 2018
Motivation

- Estimating causal effects of multiple treatments/interventions
- Common in many studies. For example:
  1. Estimating the effects of nutrition label use on body mass index
  2. Evaluating treatment programs for adolescent substance abuse
  3. Evaluating the cardiovascular safety of multiple drug classes for type 2 diabetes
Randomized Design

- Ideal for estimating causal effects:
  - Treatment groups are guaranteed to be similar in terms of covariates, $X$.

- But...
  - Expensive
  - Unethical
  - Restricted population used in the experiments

- Sometimes, we need to rely on observational data!
The Assignment Mechanism

\[ P(W = w \mid Y(1), \ldots, Y(Z), X) \]

1. Individualistic: Treatment assignment for one unit does not depend on covariates or potential outcomes of other units

2. Unconfoundedness:
\[ P(W = w \mid Y(1), \ldots, Y(Z), X) = P(W = w \mid X) \]

3. Positivity: \( 0 < P(W = w \mid X) < 1 \) for all \( w \in \mathcal{W} \)
Steps in Implementing Matching Methods

Stuart (2010) –

1. Defining “closeness”: Use a **distance measure** in order to determine whether an individual is a good match for another.

2. Given the distance measure, implement a **matching method**.

3. Assessing the **quality** of the matched cohort.

4. Analysis of the **outcome** and estimation of the treatment effect.
1. Defining Closeness

- Multiple treatments: Match on the \textit{generalized propensity score} (GPS) vector,

\[
R(X_i) = \{P(W_i = 1 \mid X_i), \ldots, P(W_i = Z \mid X_i)\} = \{r(1, X_i), \ldots, r(Z, X_i)\}.
\]

- Some possible distance measures:
  
  (i) Exact (usually on $X$)

  (ii) Mahalanobis distance (of $R(X)$, or $X$)

  (iii) Linear GPS: For reference treatment $t$,

\[
|\text{logit}[r(t, X_i)] - \text{logit}[r(t, X_j)]|
\]
2. Implementing a Matching Method

Matching for ATT: $E[Y(j) - Y(k) \mid W = t]$, $(j, k) \in W^2, j \neq k$

- 1:1:1 nearest-neighbor matching (ex: for $Z = 3$ treatments)
  
  - Set a reference treatment, say, treatment 1.
  
  - For subject $i$ receiving reference treatment 1, select a subject from each of treatments 2 and 3 with the smallest distance from subject $i$.
  
  - Extract the matched triplet only if subject $i$ has a match in each of treatment groups 2 and 3.

- Some considerations:
  
  - Selecting the number of matches per subject
  
  - With or without replacement
2. Implementing a Matching Method – Vector Matching

- Lopez & Gutman (2017) – Match on a vector of generalized propensity scores (GPS)

- Stratify on $R(\mathbf{X}) = \{r(1, \mathbf{X}), \ldots, r(Z, \mathbf{X})\}$ using $k$-means clustering, match within strata.
  - Some possible matches may not be considered by VM because they are on the boundaries of clusters.

- Use the linear GPS, $|\text{logit}[r(t, \mathbf{X}_i)] - \text{logit}[r(t, \mathbf{X}_j)]|$ as the distance measure, where $t$ is the reference treatment.

- Vector matching (VM) has been shown to produce matched sets with low covariate bias for $Z = 3$ treatments.
2. Implementing a Matching Method – Proposed Matching

- **Fuzzy Matching (FM)**: Matching within *fuzzy clusters*, using the Mahalanobis distance of pairs of GPS vector components as the distance measure
  - Fuzzy clustering allows for subjects to belong to multiple clusters
  - Ex: A subject belonging to two clusters can be matched to a subject appearing in *either* of the two clusters.
  - Matching on pairs of components of $R(\mathbf{X})$ may be useful when the total number of components is large (i.e., large $Z$)

- **GPS Matching (GPS)**: Matching on the Mahalanobis distance of the GPS vector, $R(\mathbf{X})$

- **Covariate Matching (COV)**: Matching on the Mahalanobis distance of the covariates, $\mathbf{X}$
3. Assessing Quality of Matching

- How well does a matching method improve covariate balance between treatment groups?

- Calculate the **standardized bias** at each covariate $p$ for each pair of treatments $j$ and $k$,

$$SB_{pjk} = \frac{\bar{X}_{pj} - \bar{X}_{pk}}{\delta_{pt}},$$

where $\delta_{pt}$ is the standard deviation of $X_p$ among subjects receiving reference treatment $t$.

- Extract the maximum standardized bias at each covariate,

$$\text{Max2}SB_{p} = \max(|SB_{p12}|, |SB_{p13}|, |SB_{p23}|, \ldots).$$
Simulations

- Performance of VM, FM, GPS, COV
  - Looked at $Z = 5$ and $Z = 10$
  - Number of covariates $P \in \{5, 10, 20\}$

- Simulation factors: covariate distributions, number of covariates, treatment group sample size, and others

- We summarized $Max2SB_p$ by averaging over $p$:

  $$\overline{Max2SB} = \frac{1}{P} \sum_{p=1}^{P} Max2SB_p$$

  - Past literature advocates a cutoff of 0.20–0.25.
Results: \( Z = 5 \) Treatments
Results: \( Z = 10 \) Treatments
Discussion

- Matching on the GPS vector as a novel and effective approach to generating a well-balanced matched cohort

- Importance of study population and causal estimand

- Importance of setting
  - Number of covariates?
  - Number of treatment groups?
Contact

- Email: anthony_scotina@brown.edu
- Website: scotinastats.github.io
- Twitter: @ScotinaStats

Thank you!
Contact

- Email: anthony_scotina@brown.edu
- Website: scotinastats.github.io
- Twitter: @ScotinaStats

Any questions?