Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference

Ho et al. (2007)

Gento Kato
gkato@ucdavis.edu

POL290G Causal Inference
(Lauren Peritz)

October 30, 2017 @ University of California, Davis

Matching is the method suggested to improve the **causal inference**. But, its results are often **misinterpreted**.

Authors introduce matching as the *nonparametric* method to **preprocess** data to improve the subsequent estimation of causal effect using *parametric* techniques.

The article describe various ways to avoid misinterpretations and apply matching *correctly*.

Introduction

Review

Matching Theory

Matching Application

Review: Potential Outcome & Treatment Effect

Potential Outcome Representation (in this article):

- $T_i \in \{1, 0\}$: the treatment for the unit i.
- $y_i(k) = y_i(T_i = k)$: the potential outcome of i when $T_i = k$. The realization of the corresponding random variable $Y_i(k)$.
- X_i: the characteristics of i that ensure *conditional independence* between $y_i(k)$ and T_i. (No selection bias after controling for X)

Average Treatment Effect on the Treated (**ATT**):

$$
\begin{aligned}
ATT &\equiv \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} \{T_i \times E[Y_i(1) - Y_i(0)|X_i]\} \\
&= \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} \{T_i \times [E[Y_i(1)|X_i] - E[Y_i(0)|X_i]]\} \\
&= \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} \{T_i \times [\mu_1(X_i) - \mu_0(X_i)]\}
\end{aligned}
$$

Review: ATT in Parametric Context

Remember, ATT (using observed data) is:

$$\begin{aligned} ATT &= \frac{1}{\sum_{i=1}^{n} T_i} \sum_{i=1}^{n} \{ T_i \times [\mu_1(X_i) - \mu_0(X_i)] \} \\ &= \mu_1(T_i = 1, X_i) - \mu_0(T_i = 0, X_i) \\ &= \mu_1(X_i) - \mu_0(X_i) \end{aligned}$$

Using parametric method to estimate $\mu_{T_i}(X_i)$:

$$\begin{aligned} \mu_1(X_i) &\equiv E[Y_i(1) | T_i = 1, X_i] = g(\alpha + \beta + X_i\gamma) \\ \mu_0(X_i) &\equiv E[Y_i(0) | T_i = 0, X_i] = g(\alpha + X_i\gamma) \end{aligned}$$

Where $g(\cdot)$ is a functional form ($g(c) = c$ if linear), and X_i is a distribution with probability density $p(\mu_{T_i}(X_i), \theta)$.

Review: Randomized Experiment

The ATT in randomized experiment is simply:

$$\mu_1 - \mu_0 \textbf{ or } g(\alpha + \beta) - g(\alpha)$$

We can drop X_i, because...

1. The observed units are **randomly sampled**.
2. The value of T_i is **randomly assigned** (independent of Y_i).
3. The sample has **large N**.

(1) and (2) to avoid selection bias, and (3) to reduce the possibility of error by chance.

The Problem of Unable to Drop X_i

In observed data, we cannot drop X_i from the model of causal inference. This causes problems...

1. **The curse of dimensionality**: By avoiding assumptions about γX_i's functional form, as many parameter as values of X_i and all possible interactions need to be included in the model. This is not feasible given the data limitation.
2. **Model dependence**: In practice, one need to make a lot of assumptions about the functional form of parameters. The results become sensitive to model specifications.

Knowing X_i is not enough. The issue persists unless **correct functional form of all relationships** b/w X_i & Y_i is known.

Matching as Nonparametric Preprocessing

Idea: Preprocess the data so that T_i **and** X_i **are completely unrelated** (or at least minimally related). In new data...

- No (or less) need to model full parametric relationship between Y_i and X_i.
- Elimination of (or reduction in) model dependence.

Matching: Achieve the nonparametric **balance** in data.

$$\tilde{p}(X|T=1) = \tilde{p}(X|T=0)$$

This balance minimizes/eliminates γ in $g(\alpha + \beta T_i + X_i\gamma)$, which reduces the equation to $g(\alpha + \beta T_i)$.

Achieving Balance through Matching

Two components of finding balance:

1. Enduring **common support** by eliminating observations with X_i densities that are not overlapping between observed treatment and control units.
2. Enduring overlapping **densities to have same heights** by additional selection.

The flexibility: Allows **replacement** or **double-matching**.

The only issue: **Dropping too much** n

- Increase variance by small n.
- May increase bias in ATT by dropping treated units.

Matching Methods

Exact Matching:

- "[M]atch all control units with exactly the same covariate values" (217)
- About **distributions**, not about **one-to-one pairing**.
- This is the best, if sufficient treatment units are preserved.

Propensity Score Matching:

- Estimate propensity score $p(T_i = 1|X_i)$, often using logistic regression (functional forms assumed).
- Match each treated unit to the control unit with the most similar value of the propensity score.
- If the procedure balances X between treated & control, use it. If not, try other model specifications.

Assessment of Balance

It is hard to examine multidimensional density. Usually examine various low-dimensional summaries.

The balance test fallacy:

- "[B]alance is a characteristic of the observed sample, not some hypothetical population." (221)
- More balance is *always* better, not above some threshold.
- $\Downarrow n$ automatically leads to $p \Uparrow$. Harder to reject balance.

Better evaluations:

- Compare means.
- Assess QQ plot to compare distributions.

Failure in balancing implies that the data set "is too fragile for making robust causal inferences by any means" (223).

Apply Parametric Analysis to Matched Data

Conventional Practice: Just applying *difference in means* is an *unfortunate* practice.

- T_i and X_i are still related (unless exactly matched)
- The omitted variable bias is still present.

Better practice: "[U]se the **same parametric analysis** on the preprocessed data as would have been used to analyze the original raw data set" (223). Same variance estimator can be used for uncertainty estimates.

Introduction

Review

Matching Theory

Matching Application

Carpenter (2002)

Variables:

- i: 408 new drugs considered by FDA.
- Y_i: Approval time (months from the submission).
- T_i: Dem. Majority in Senate (1); Not (0)
- X_i: 18 variables of clinical and epidemiology factors and firm characteristics.

Matching (Table 1):

- Propensity score $p(T_i|X_i)$ estimated with linear predictors.
- 15 control units and 2 units out of the common support discarded.
- Nearest neighbor matching (without replacement). Exact restriction for six binary variables.
- Discard 10 treated units and 92 control units (including those without common support).

Result (Figure 2): Reduction in model dependence.

Koch (2002)

Variables:

- i: 1203 Republican male candidates.
- Y_i: Voter evaluation of ideology.
- T_i: Have Visibility (0, N=853), Not (1, N=350)
- X_i: 5. candidate ideology, voter perception of party ideology, respondent ideology, candidate FT, political awareness.

Matching (Figure 3):

- Propensity score $p(T_i|X_i)$ estimated with linear predictors.
- 350 matches to treatment units.

Result (Figure 4): Reduction in model dependence.

References

Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3):199–236.

Thank you for listening!