Matching Continued

INFO/STSCI/ILRST 3900: Causal Inference

17 Oct 2024

Learning goals for today

At the end of class, you will be able to:

1. Understand propensity score matching and coarsened exact matching
2. Use matching methods to estimate causal effects

Matching: so far

Goal: Sample Average Treatment Effect on the Treated

$$E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$$

Potential Solution: Create a group of untreated individuals, $\mathcal{M}$, which have a **similar distribution of** L to the treated group

$$\frac{1}{n_m} \sum_{i \in \mathcal{M}} Y_i = \frac{1}{n_m} \sum_{i \in \mathcal{M}} Y_i^{a=0} \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0} \approx E(Y^{a=0} \mid A = 1)$$

How:

▶ Find untreated unit(s) which are similar to each treated unit

Matching: so far

Goal: Sample Average Treatment Effect on the Treated
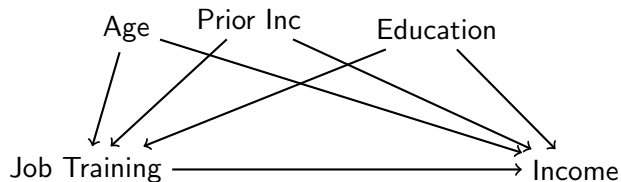
$$E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$$

Potential Solution: Create a group of untreated individuals, $\mathcal{M}$, which have a **similar distribution of** L to the treated group

$$\frac{1}{n_m} \sum_{i \in \mathcal{M}} Y_i = \frac{1}{n_m} \sum_{i \in \mathcal{M}} Y_i^{a=0} \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0} \approx E(Y^{a=0} \mid A = 1)$$
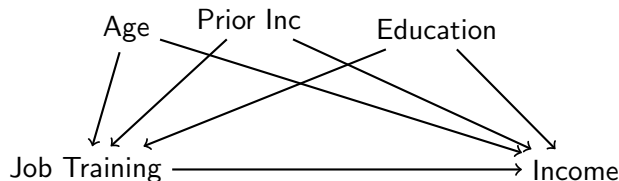
How:

▶ Find untreated unit(s) which are similar to each treated unit

▶ Define "similar"

Motivating example

Motivating example

- $\vec{L}$ denotes the set of covariates we want to condition on
- $\vec{L}_i$ denotes the covariates values for individual i
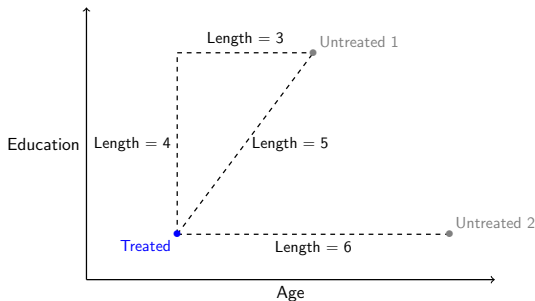
A common distance metric: Exact matching

A common distance metric: Exact matching

▶ Ideally, we find an exact match for each treated unit

$$d(i,j) = \begin{cases} 0 & \text{if } \vec{L}_i = \vec{L}_j \\ \infty & \text{if } \vec{L}_i \neq \vec{L}_j \end{cases}$$

Because of the curse of dimensionality, this often leads to **no matches at all**

What if $\vec{L}$ is multivariate? We need a **distance metric**

- Manhattan distance: $d(i,j) = \sum_p |L_{pi} - L_{pj}|$
- Euclidean distance: $d(i,j) = \sqrt{\sum_p (L_{pi} - L_{pj})^2}$
- Mahalanobis distance: Rescales and accounts for correlation

A common distance metric: Coarsened exact matching[1]

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]

- Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$

---

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]

- Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - Example: Age 15-20, 20-25, 25-30, etc

---

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - ▶ Example: Age 15-20, 20-25, 25-30, etc
- ▶ Match exactly on $\tilde{\vec{L}}$

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]

- Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - Example: Age 15-20, 20-25, 25-30, etc
- Match exactly on $\tilde{\vec{L}}$

$$d(i,j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

---

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]

- Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - Example: Age 15-20, 20-25, 25-30, etc
- Match exactly on $\tilde{\vec{L}}$

$$d(i,j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- Benefit: Very transparent

---

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - ▶ Example: Age 15-20, 20-25, 25-30, etc
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i,j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- ▶ Benefit: Very transparent
- ▶ Benefit: Directly targets balance in L

---

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - ▶ Example: Age 15-20, 20-25, 25-30, etc
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i,j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- ▶ Benefit: Very transparent
- ▶ Benefit: Directly targets balance in L
- ▶ Drawback: May not find a good match for all individuals

---

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching[1]
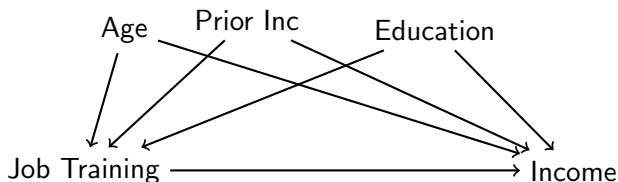
- Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - Example: Age 15-20, 20-25, 25-30, etc
- Match exactly on $\tilde{\vec{L}}$

$$d(i,j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- Benefit: Very transparent
- Benefit: Directly targets balance in L
- Drawback: May not find a good match for all individuals

---

[1]Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. Political Analysis, 20(1), 1-24.
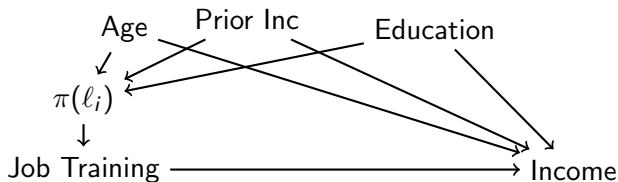
A common distance metric: Propensity scores

A common distance metric: Propensity scores
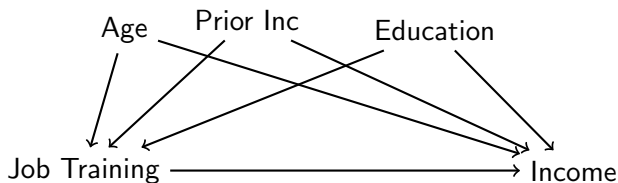
Suppose $\vec{L}$ only affects A through a probability of treatment

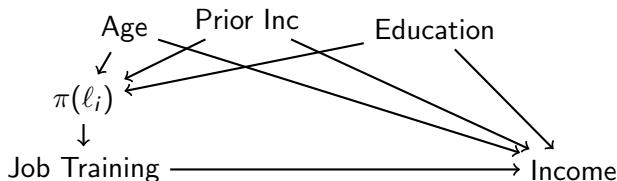$$\pi_i = \pi(\vec{\ell}_i) = P(A_i = 1 \mid \vec{L} = \vec{\ell})$$

A common distance metric: Propensity scores

Suppose $\vec{L}$ only affects A through a probability of treatment

$$\pi_i = \pi(\vec{\ell_i}) = P(A_i = 1 \mid \vec{L} = \vec{\ell})$$

A common distance metric: Propensity scores

Suppose $\vec{L}$ only affects A through a probability of treatment

$$\pi_i = \pi(\vec{\ell_i}) = P(A_i = 1 \mid \vec{L} = \vec{\ell})$$

Conditional exchangeability holds given $\pi(\ell_i)$

Why propensity scores are nice

Why propensity scores are nice

- Can match on propensity scores directly instead of L

Why propensity scores are nice

- ▶ Can match on propensity scores directly instead of L
  - ▶ Easy to reason about

Why propensity scores are nice

- Can match on propensity scores directly instead of L
  - Easy to reason about
  - Can directly visualize the univariate matches

Why propensity scores are nice

- ▶ Can match on propensity scores directly instead of *L*
  - ▶ Easy to reason about
  - ▶ Can directly visualize the univariate matches
- ▶ Intuitive: Prioritizes covariates that predict treatment
- ▶ Mathematical guarantees on average

Why propensity scores are nice

- ▶ Can match on propensity scores directly instead of L
  - ▶ Easy to reason about
  - ▶ Can directly visualize the univariate matches
- ▶ Intuitive: Prioritizes covariates that predict treatment
- ▶ Mathematical guarantees on average
  - ▶ If our DAG is correct

Why propensity scores are nice

- Can match on propensity scores directly instead of L
  - Easy to reason about
  - Can directly visualize the univariate matches
- Intuitive: Prioritizes covariates that predict treatment
- Mathematical guarantees on average
  - If our DAG is correct
  - If our matches are good

Why propensity scores are nice

- ▶ Can match on propensity scores directly instead of L
  - ▶ Easy to reason about
  - ▶ Can directly visualize the univariate matches
- ▶ Intuitive: Prioritizes covariates that predict treatment
- ▶ Mathematical guarantees on average
  - ▶ If our DAG is correct
  - ▶ If our matches are good
  - ▶ We should **on average** get a matched group which looks like the the treatment group

$$P(L \mid \pi_i, A_i = 1) = P(L \mid \pi_i, A_i = 0)$$

Multivariate distances: Recap

When matching on multivariate $\vec{L}$, you have to define the distance between each pair of confounder values $\vec{\ell}_j$ and $\vec{\ell}_i$

- ▶ Manhattan distance
- ▶ Euclidean distanace
- ▶ Mahalanobis distance
- ▶ Coarsened exact distance
- ▶ Propensity score distance

There is no right answer! Depends on the setting.

Multivariate distances: Recap

When matching on multivariate $\vec{L}$, you have to define the distance between each pair of confounder values $\vec{\ell}_j$ and $\vec{\ell}_i$

- ▶ Manhattan distance
- ▶ Euclidean distanace
- ▶ Mahalanobis distance
- ▶ Coarsened exact distance
- ▶ Propensity score distance

There is no right answer! Depends on the setting.

- ▶ Propensity scores are most popular
- ▶ Sometimes they are substantively meaningful

Multivariate distances: Recap

When matching on multivariate $\vec{L}$, you have to define the distance between each pair of confounder values $\vec{\ell}_j$ and $\vec{\ell}_i$

- ▶ Manhattan distance
- ▶ Euclidean distanace
- ▶ Mahalanobis distance
- ▶ Coarsened exact distance
- ▶ Propensity score distance

There is no right answer! Depends on the setting.

- ▶ Propensity scores are most popular
- ▶ Sometimes they are substantively meaningful
- ▶ Balance only occurs on average

Evaluate the matched sets

Whatever method, you should check that it worked

- Compare means of $\vec{L}$ (propensity scores) across groups
- Possibly compare interaction cells; e.g., race $\times$ age

Evaluate the matched sets

Whatever method, you should check that it worked

- ▶ Compare means of $\vec{L}$ (propensity scores) across groups
- ▶ Possibly compare interaction cells; e.g., race $\times$ age
- ▶ Visually assess distribution

Overlap

▶ Lack of overlap may indicate violation of positivity assumption

$$P(A = a \mid L = \ell) > 0 \text{ for all a}$$

▶ Ex: Sarah has no MD training. Would Sarah earn more money if she were a surgeon?

$$P(A = \text{Surgeon} \mid \text{No MD}) = 0$$

▶ If no good match exists, could be that $P(A = 0 \mid L = \ell) = 0$

Matching: A word of warning[2]

[2]Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. Annual Review of Political Science, 12(1), 487-508.

Matching: A word of warning[2]

$$L \longrightarrow A \longrightarrow Y$$

Matching works!

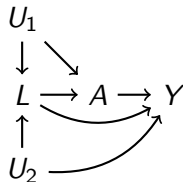$$L \longrightarrow A \longrightarrow Y$$

Matching: A word of warning[2]

Matching works!

No help!

$$U$$

$$L \longrightarrow A \longrightarrow Y \qquad L \longrightarrow A \longrightarrow Y$$
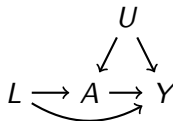
Matching: A word of warning[2]
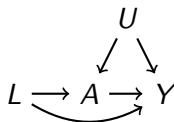
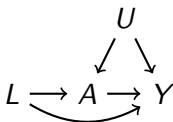**Matching works!**   **No help!**   **No help!**

[2]Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. Annual Review of Political Science, 12(1), 487-508.
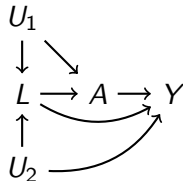
Matching: A word of warning[2]

Matching works!   **No help!**   **No help!**

Matching is an estimation strategy.

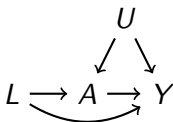[2]Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. Annual Review of Political Science, 12(1), 487-508.
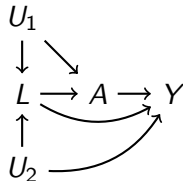
Matching: A word of warning[2]

Matching works!   **No help!**   **No help!**

Matching is an estimation strategy.
It does not solve identification problems.

---

[2]Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. Annual Review of Political Science, 12(1), 487-508.

Estimating a causal effect

▶ If we've matched everything well, we can compare the means

---

[3]On the statistical role of inexact matching in observational studies. Guo and Rothenhäusler (2023)

Estimating a causal effect

- If we've matched everything well, we can compare the means
  - Treated group (with a match)
  - Matched control group

---

[3]On the statistical role of inexact matching in observational studies. Guo and Rothenhäusler (2023)

Estimating a causal effect

- If we've matched everything well, we can compare the means
  - Treated group (with a match)
  - Matched control group
- We can be extra careful by combining regression + matching
  - If everything is perfect, both should be fine on their own

---

[3]On the statistical role of inexact matching in observational studies. Guo and Rothenhäusler (2023)

Estimating a causal effect

- If we've matched everything well, we can compare the means
  - Treated group (with a match)
  - Matched control group
- We can be extra careful by combining regression + matching
  - If everything is perfect, both should be fine on their own
  - Combining can reduce bias
  - Reduces model sensitivity[3]

---

[3]On the statistical role of inexact matching in observational studies. Guo and Rothenhäusler (2023)

Learning goals for today

At the end of class, you will be able to:

1. Understand propensity score matching and coarsened exact matching

2. Use matching methods to estimate causal effects