Model-based causal estimation: From outcome models to treatment weighting

Ian Lundberg
Sociol 114
soc114.github.io

Winter 2025

Learning goals for today

At the end of class, you will be able to

- estimate average causal effects with a parametric model
  - for the outcome $\mathrm{E}(Y \mid A, \vec{L})$
  - for the treatment $\mathrm{P}(A \mid \vec{L})$

Optional reading:

- Hernán and Robins 2020 Chapter 12.1–12.5, 13, 15.1

Nonparametric estimation

Causal assumptions

$$\begin{tikzcd} \vec{L} \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

Nonparametric estimator

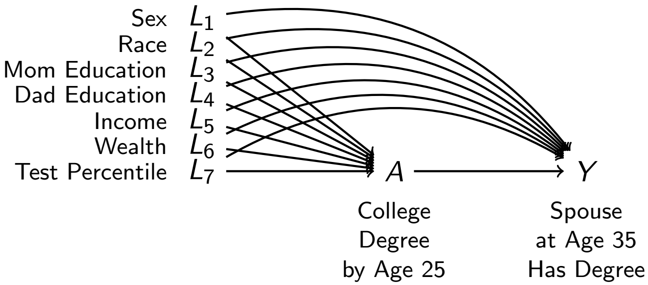
$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathrm{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

For every unit i,

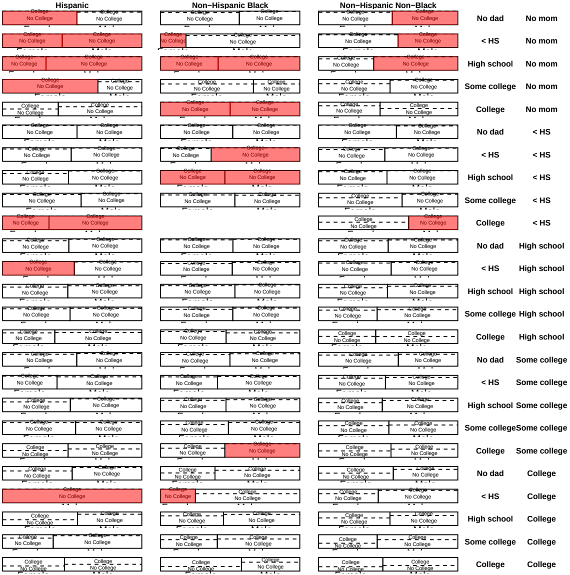
- find units who look like them on confounders $\vec{L}$
- who actually got treatment $A = a$
- take the average among those units

Then average over all units

Nonparametric estimation breaks down



Nonparametric estimation breaks down



Parametric estimation: Outcome model

Causal assumptions

$$\begin{tikzcd} \vec{L} \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

Parametric estimator

$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n}\sum_{i=1}^{n}\hat{\mathrm{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

Parametric estimation: Outcome model

Causal assumptions

$$
\begin{tikzcd}
\vec{L} \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

Parametric estimator

$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathrm{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

Where $\hat{E}$ is a model-based prediction

$$\hat{\mathrm{E}}(Y \mid \vec{L}, A) = \hat{\alpha} + \vec{L}'\hat{\vec{\gamma}} + A\hat{\beta}$$

Parametric estimation: Outcome model

Causal assumptions

$$\begin{tikzcd} \vec{L} \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

Parametric estimator

$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n}\sum_{i=1}^{n}\hat{\mathrm{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

Where $\hat{E}$ is a model-based prediction

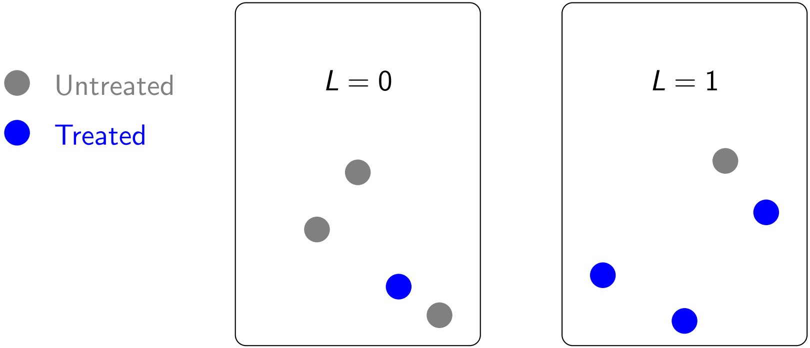
$$\hat{\mathrm{E}}(Y \mid \vec{L}, A) = \hat{\alpha} + \vec{L}'\hat{\vec{\gamma}} + A\hat{\beta}$$

$$\begin{tikzcd} \vec{L} \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

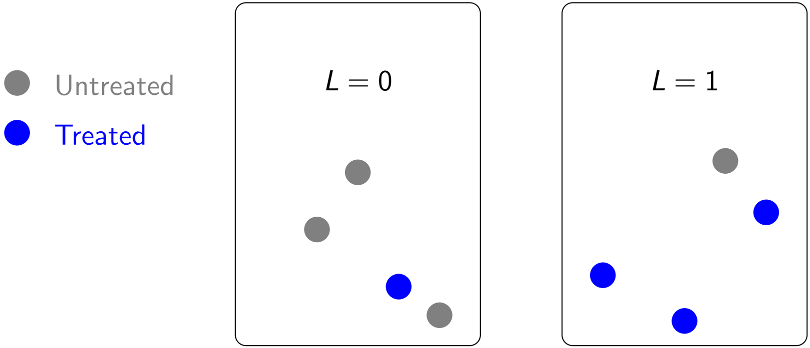
$$\begin{tikzcd} \vec{L} \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

$$\begin{tikzcd} \vec{L} \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

Inverse probability of treatment weighting

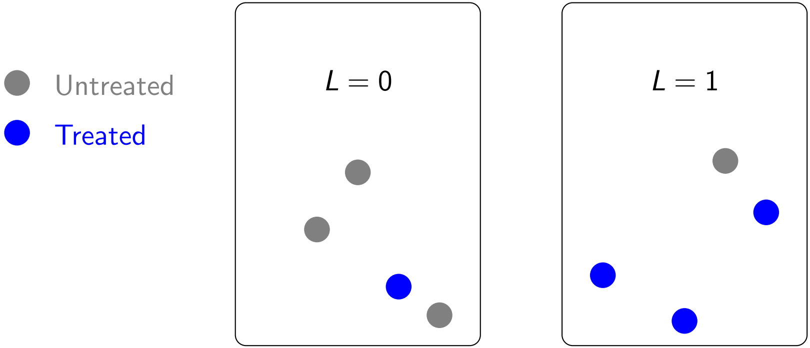


Inverse probability of treatment weighting



Propensity score: $\pi_i = \mathrm{P}(A = A_i \mid L = L_i)$

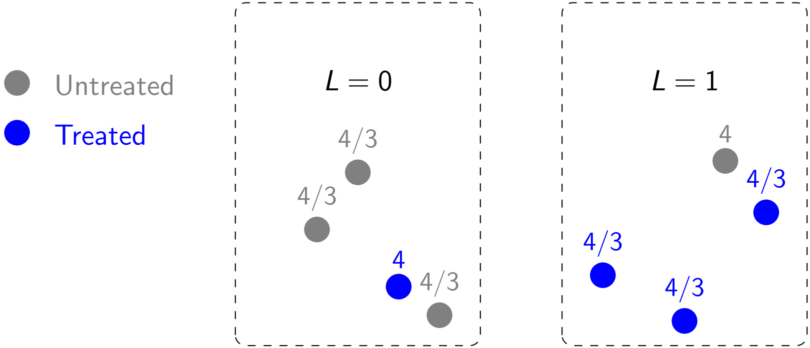
Inverse probability of treatment weighting



Propensity score: $\pi_i = \mathrm{P}(A = A_i \mid L = L_i)$

Inverse probability weight: $w_i = \frac{1}{\pi_i}$

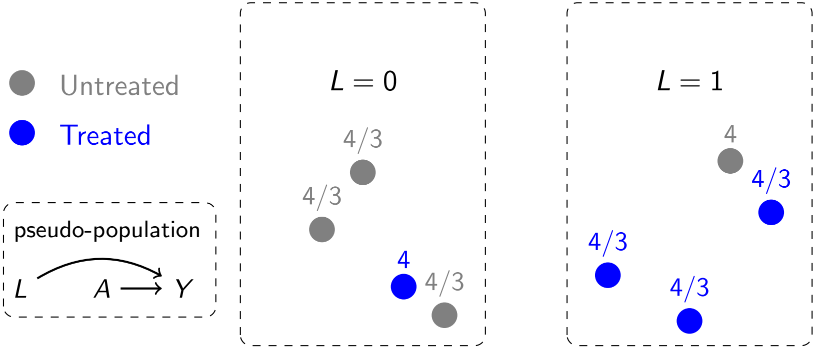
Inverse probability of treatment weighting



Propensity score: $\pi_i = \mathrm{P}(A = A_i \mid L = L_i)$

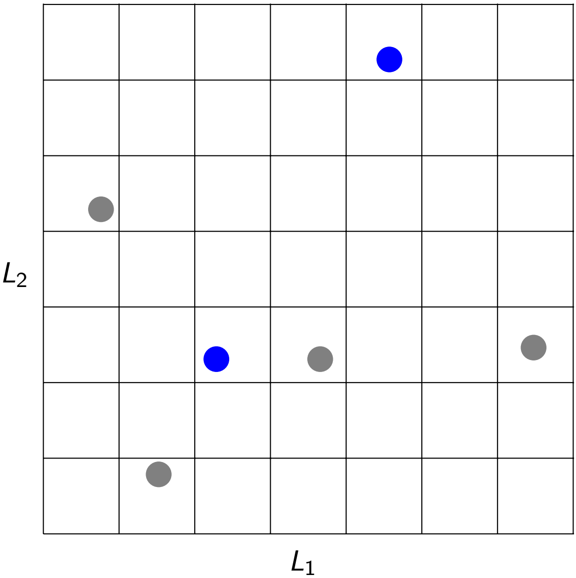
Inverse probability weight: $w_i = \frac{1}{\pi_i}$

Inverse probability of treatment weighting



Propensity score: $\pi_i = \mathrm{P}(A = A_i \mid L = L_i)$

Inverse probability weight: $w_i = \frac{1}{\pi_i}$


L_2
L_1

Model the treatment assignment

$$\hat{\text{P}}(A = 1 \mid \vec{L}) = \text{logit}^{-1}\left(\hat{\alpha} + \hat{\vec{\gamma}}\vec{L}\right)$$

Predict the propensity score for each unit

$$\hat{\pi}_i = \begin{cases} \text{logit}^{-1}\left(\hat{\alpha} + \hat{\vec{\gamma}}\vec{L}\right) & \text{if } A_i = 1 \\ 1 - \text{logit}^{-1}\left(\hat{\alpha} + \hat{\vec{\gamma}}\vec{L}\right) & \text{if } A_i = 0 \end{cases}$$

Estimate by inverse probability weighting

$$\hat{\text{E}}(Y^a) = \frac{1}{N}\sum_{i:A_i=a}\frac{Y_i}{\hat{\pi}_i}$$

Learning goals for today

At the end of class, you will be able to

- estimate average causal effects with a parametric model
  - for the outcome $\text{E}(Y \mid A, \vec{L})$
  - for the treatment $\text{P}(A \mid \vec{L})$

Optional reading:

- Hernán and Robins 2020 Chapter 12.1–12.5, 13, 15.1

Problem: Extreme weights create high variance

Problem: Extreme weights create high variance

Suppose a stratum $\vec{L} = \vec{\ell}$ contains

- 100 untreated units
- 1 treated unit

Problem: Extreme weights create high variance

Suppose a stratum $\vec{L} = \vec{\ell}$ contains

- 100 untreated units
- 1 treated unit

The treated unit gets a weight of 100.

Problem: Extreme weights create high variance

Suppose a stratum $\vec{L} = \vec{\ell}$ contains

- 100 untreated units
- 1 treated unit

The treated unit gets a weight of 100.

The estimate depends heavily on which treated unit happens to be included in the sample $\rightarrow$ high-variance estimator

Problem: Extreme weights create high variance

Suppose a stratum $\vec{L} = \vec{\ell}$ contains

- 100 untreated units
- 1 treated unit

The treated unit gets a weight of 100.

The estimate depends heavily on which treated unit happens to be included in the sample $\rightarrow$ high-variance estimator

Two solutions

1. Trim the weights
2. Truncate the weights

Problem: Extreme weights create high variance

Suppose a stratum $\vec{L} = \vec{\ell}$ contains

- 100 untreated units
- 1 treated unit

The treated unit gets a weight of 100.

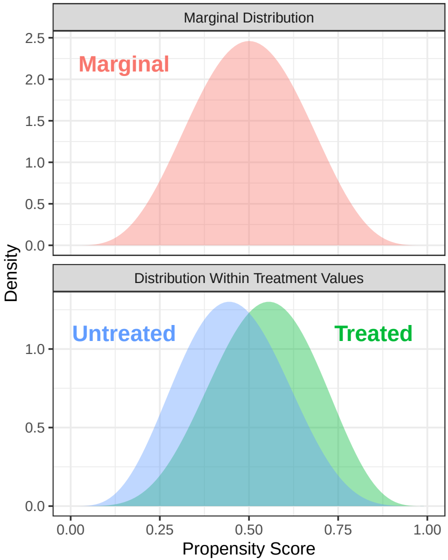
The estimate depends heavily on which treated unit happens to be included in the sample $\rightarrow$ high-variance estimator

Two solutions

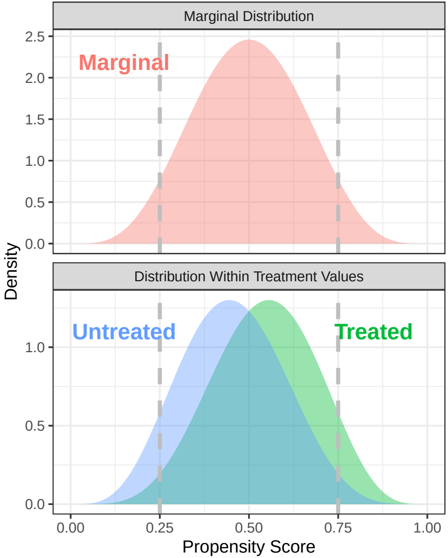
1. Trim the weights
2. Truncate the weights

Both solutions accept bias in order to reduce variance

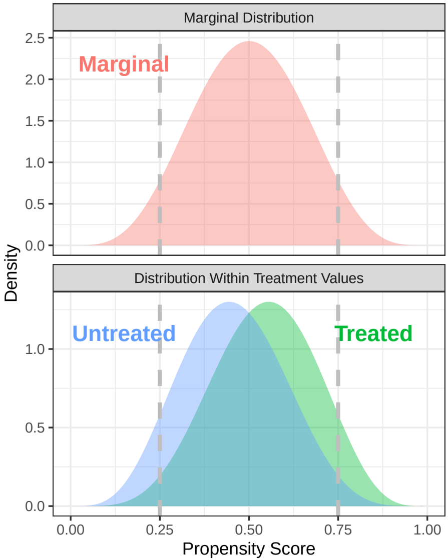
Accepting bias to reduce variance: Trimming



Accepting bias to reduce variance: Trimming

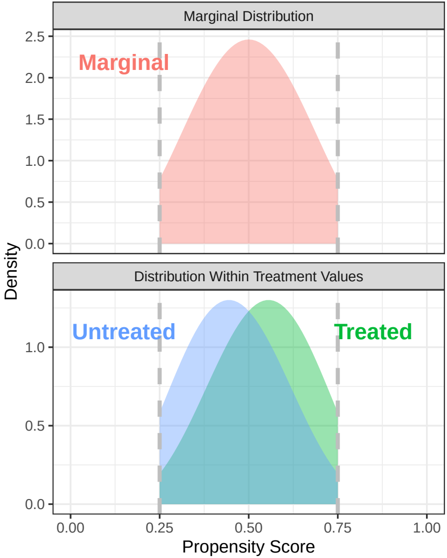


Accepting bias to reduce variance: Trimming



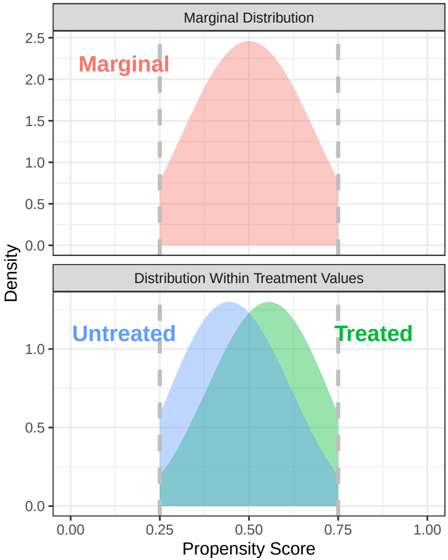
Drop units with extreme weights

Accepting bias to reduce variance: Trimming



Drop units with extreme weights

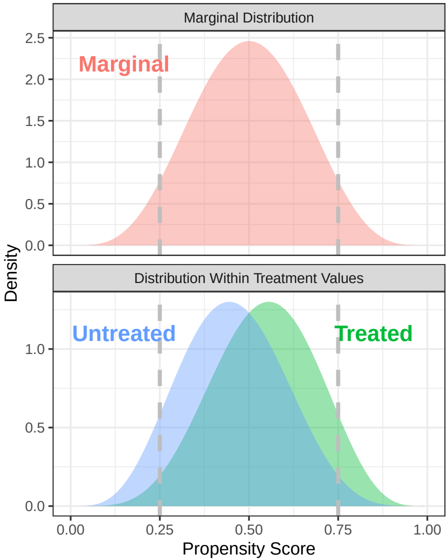
Accepting bias to reduce variance: Trimming



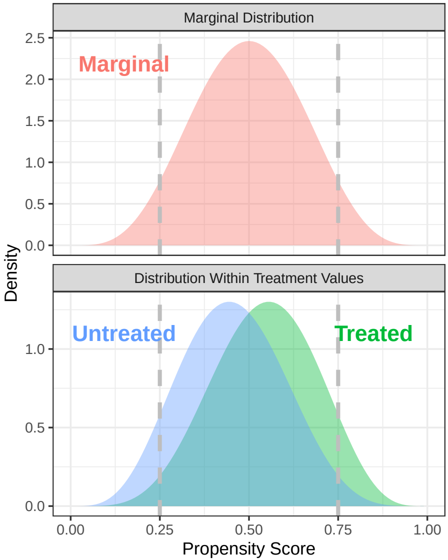
Drop units with extreme weights

Changes target population — Biased for full population

Accepting bias to reduce variance: Weight truncation

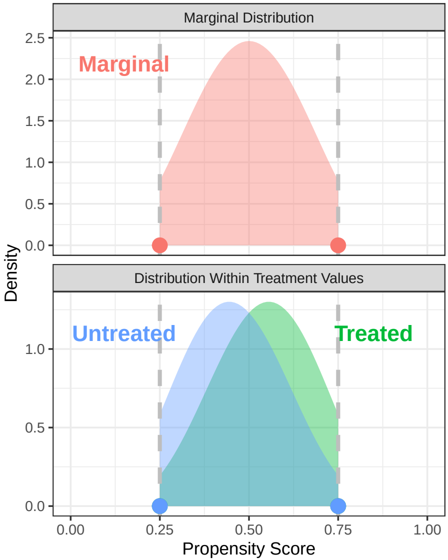


Accepting bias to reduce variance: Weight truncation



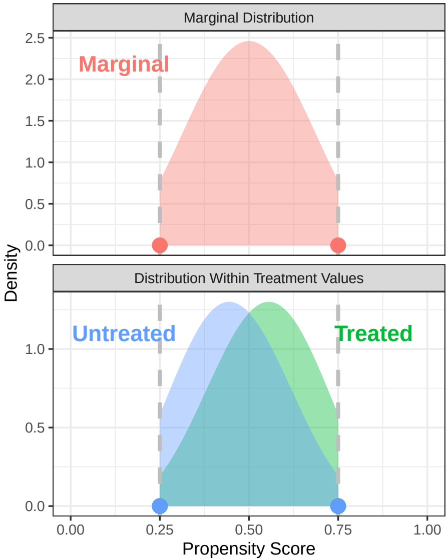
Truncate values of extreme weights

Accepting bias to reduce variance: Weight truncation



Truncate values of extreme weights

Accepting bias to reduce variance: Weight truncation



Truncate values of extreme weights

Biased: Ignores some confounding

Learning goals for today

At the end of class, you will be able to

- estimate average causal effects with a parametric model
  - for the outcome $\text{E}(Y \mid A, \vec{L})$
  - for the treatment $\text{P}(A \mid \vec{L})$

Optional reading:

- Hernán and Robins 2020 Chapter 12.1–12.5, 13, 15.1