Parametric g-Formula

Cornell STSCI / INFO / ILRST 3900
Fall 2023
causal3900.github.io

28 Sep 2023

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 $\text{P}(A \mid \vec{L})$

After class:

- 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

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

Nonparametric estimation

Causal assumptions

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

Nonparametric estimator

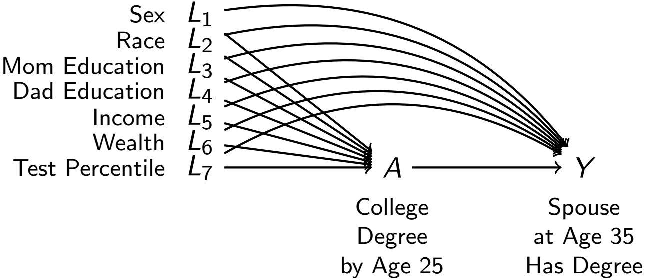
$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n}\sum_{i=1}^{n} \hat{\mathrm{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

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



Sex L_1

Race L_2

Mom Education L_3

Dad Education L_4

Income L_5

Wealth L_6

Test Percentile L_7

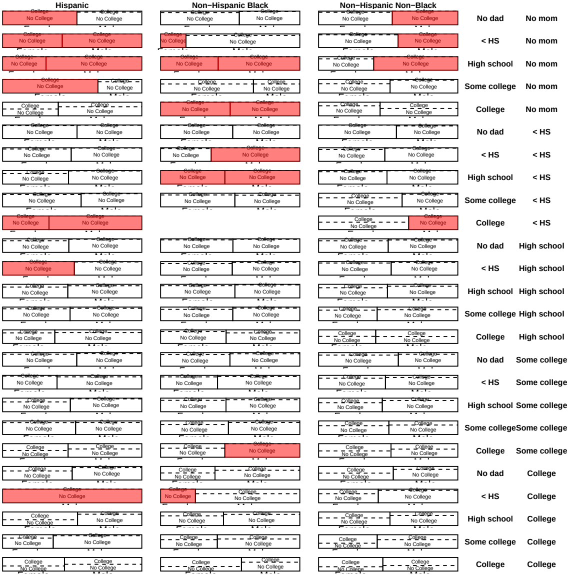
A

Y

College
Degree
by Age 25

Spouse
at Age 35
Has Degree

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

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

First, learn a model to predict Y given $\vec{L}$ and A

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

Parametric estimation: Outcome model

Causal assumptions

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

Parametric estimator

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

First, learn a model to predict Y given $\vec{L}$ and A

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

For every unit i,

- change the treatment value to a
- predict the outcome

Then average over all units

The parametric g-formula: Connection to $\hat{\beta}$

The parametric g-formula: Connection to $\hat{\beta}$

Estimator for the effect $\mathsf{E}(Y^1) - \mathsf{E}(Y^0)$:

The parametric g-formula: Connection to $\hat{\beta}$

Estimator for the effect $\mathrm{E}(Y^1) - \mathrm{E}(Y^0)$:

$$\hat{\mathrm{E}}(Y^1) - \hat{\mathrm{E}}(Y^0) = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 1\right)\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 0\right)\right)$$

The parametric g-formula: Connection to $\hat{\beta}$

Estimator for the effect $\mathrm{E}(Y^1) - \mathrm{E}(Y^0)$:

$$
\begin{aligned}
\hat{\mathrm{E}}(Y^1) - \hat{\mathrm{E}}(Y^0) &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 1\right)\right) \\
&\qquad - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 0\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}
\end{aligned}
$$

The parametric g-formula: Connection to $\hat{\beta}$

Estimator for the effect $\mathsf{E}(Y^1) - \mathsf{E}(Y^0)$:

$$\begin{aligned}
\hat{\mathsf{E}}(Y^1) - \hat{\mathsf{E}}(Y^0) &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 1\right)\right) \\
&\qquad - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 0\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{\beta} \\
&= \hat{\beta}
\end{aligned}$$

The parametric g-formula: Connection to $\hat{\beta}$

Estimator for the effect $\text{E}(Y^1) - \text{E}(Y^0)$:

$$
\begin{aligned}
\hat{\text{E}}(Y^1) - \hat{\text{E}}(Y^0) &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 1\right)\right) \\
&\qquad - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 0\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{\beta} \\
&= \hat{\beta}
\end{aligned}
$$

With OLS, the parametric g-formula collapses on the coefficient.

The parametric g-formula is more general

Suppose the effect of A depends on L

$$\mathsf{E}(Y \mid A, L) = \alpha + \gamma L + \beta A + \eta A L$$

The parametric g-formula is more general

Suppose the effect of A depends on L

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

Estimator for the effect $\mathrm{E}(Y^1) - \mathrm{E}(Y^0)$:

$$\begin{aligned}\hat{\mathrm{E}}(Y^1) - \hat{\mathrm{E}}(Y^0) = &\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta}\times 1 + \hat{\eta}\times 1\times \ell_i)\right)\right) \\ &- \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta}\times 0 + \hat{\eta}\times 0\times \ell_i\right)\right)\end{aligned}$$

The parametric g-formula is more general

Suppose the effect of A depends on L

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

Estimator for the effect $\mathrm{E}(Y^1) - \mathrm{E}(Y^0)$:

$$\begin{aligned}
\hat{\mathrm{E}}(Y^1) - \hat{\mathrm{E}}(Y^0) &= \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat\alpha + \hat\gamma \ell_i + \hat\beta \times 1 + \hat\eta \times 1 \times \ell_i)\right)\right) \\
&\qquad - \left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat\alpha + \hat\gamma \ell_i + \hat\beta \times 0 + \hat\eta \times 0 \times \ell_i\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\hat\beta + \hat\eta \ell_i\right)
\end{aligned}$$

The g-formula no longer collapses to a coefficient!

Parametric g-formula: Outcome model recap

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

1. Model the outcome mean $\mathrm{E}(Y \mid A, \vec{L})$
2. Change everyone's treatment to the value of interest
3. Predict for everyone
4. Take the average

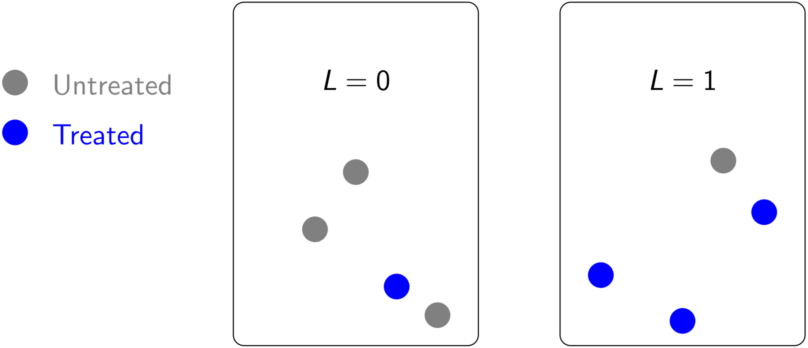
$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n} \sum_{i=1}^{n} \hat{\mathrm{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

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

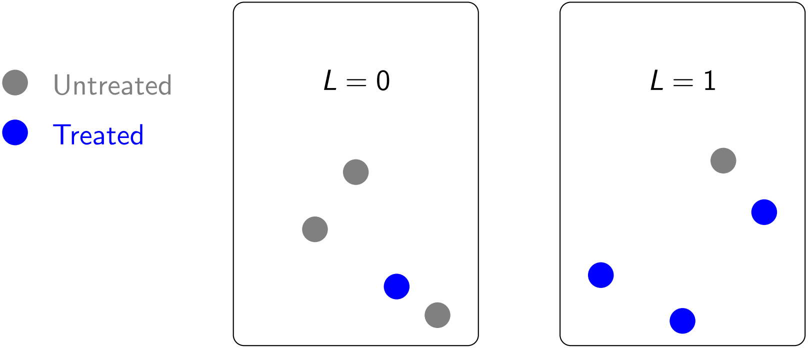
$$\begin{tikzcd}\vec{L} \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y\end{tikzcd}$$

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

Inverse probability of treatment weighting

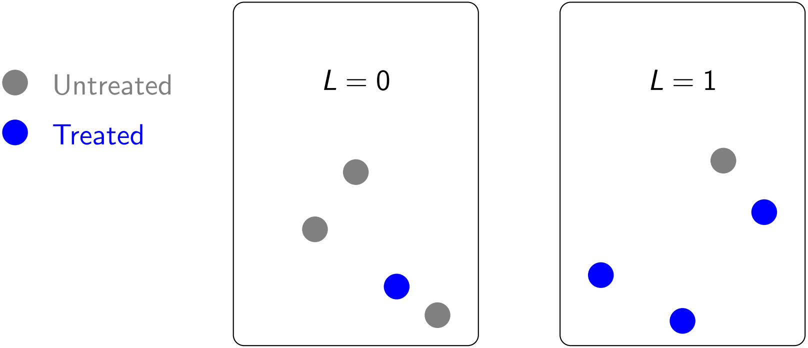


Inverse probability of treatment weighting



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

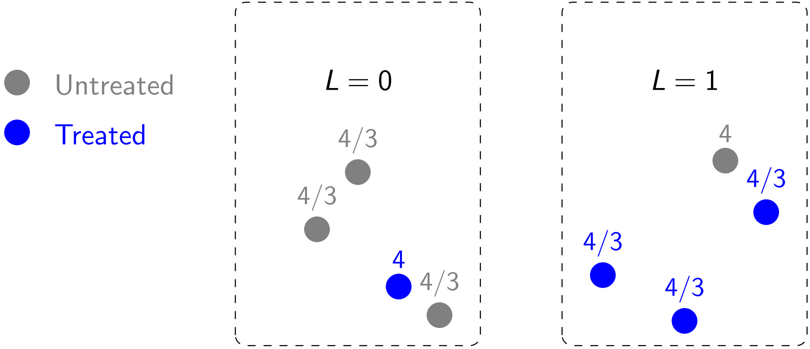
Inverse probability of treatment weighting



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

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

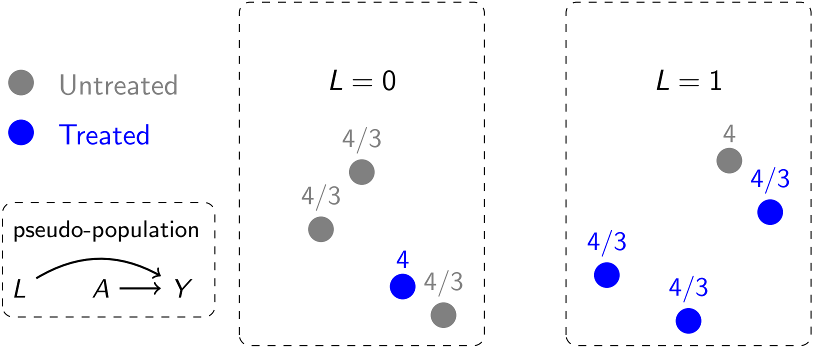
Inverse probability of treatment weighting



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

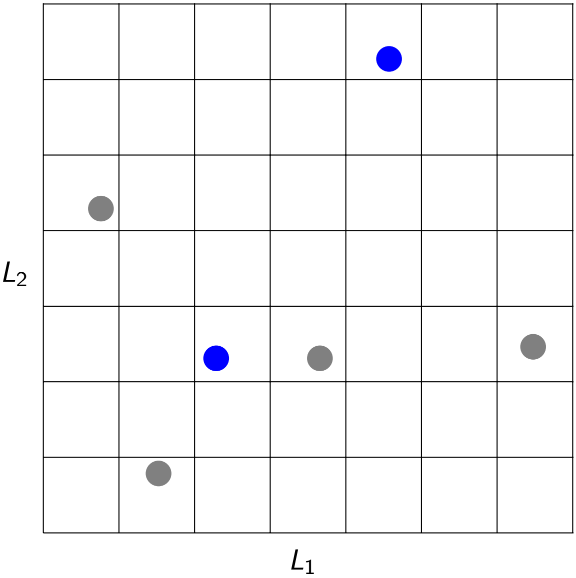
Inverse probability weight: $w_i = \frac{1}{\pi_i}$

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

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

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

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

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 $\text{P}(A \mid \vec{L})$

After class:

- 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 → 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.

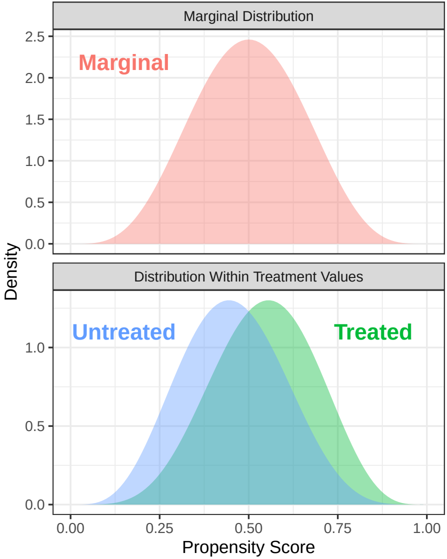
The estimate depends heavily on which treated unit happens to be included in the sample → high-variance estimator

Two solutions

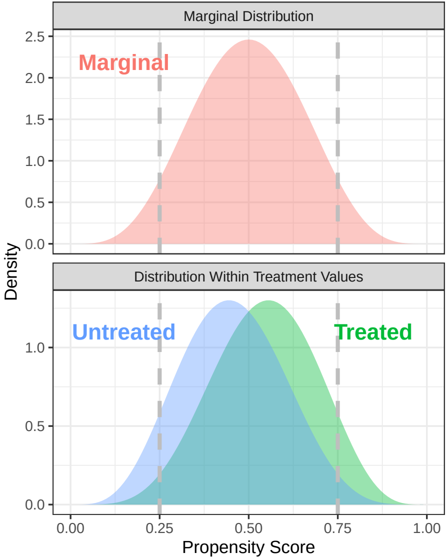
1. Trim the weights
2. Truncate the weights

Both solutions accept bias in order to reduce variance

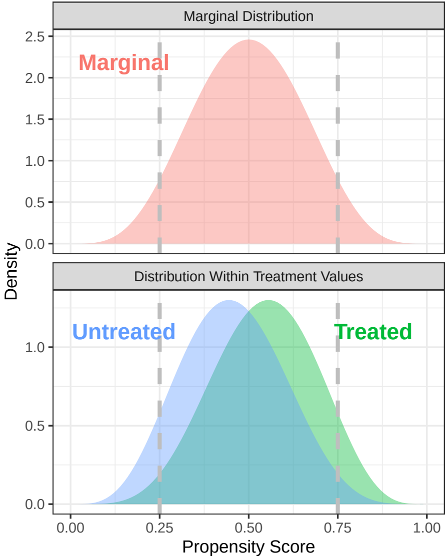
Accepting bias to reduce variance: Trimming



Accepting bias to reduce variance: Trimming

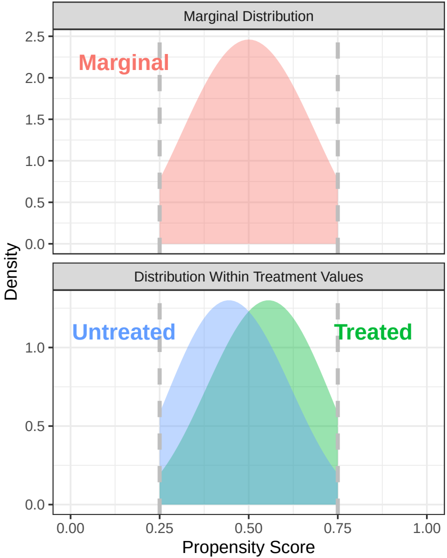


Accepting bias to reduce variance: Trimming



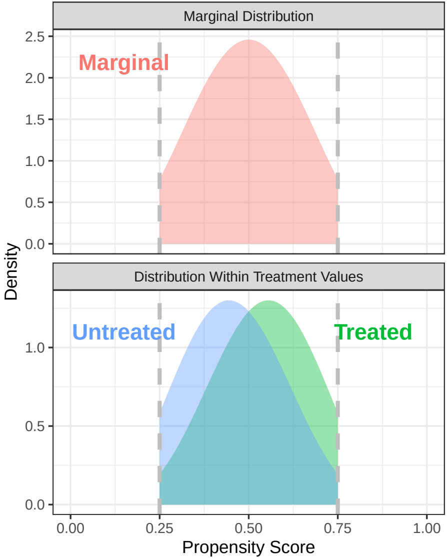
Drop units with extreme weights

Accepting bias to reduce variance: Trimming



Drop units with extreme weights

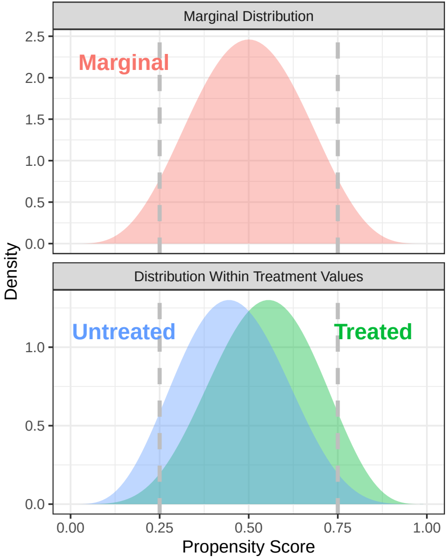
Accepting bias to reduce variance: Trimming



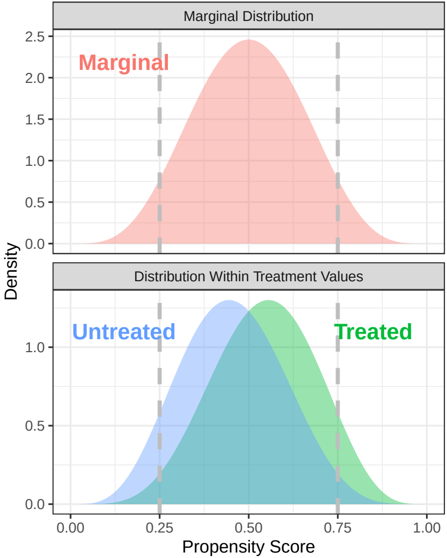
Drop units with
extreme weights

Changes target population
— Biased for full population

Accepting bias to reduce variance: Weight truncation

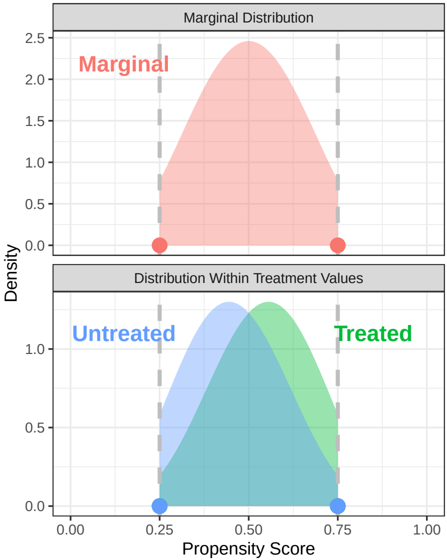


Accepting bias to reduce variance: Weight truncation



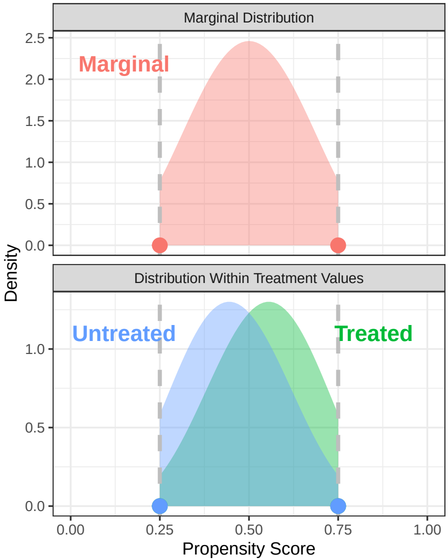
Truncate values of extreme weights

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