# Estimation by weighting

Ian Lundberg
Soc 212b
ilundberg.github.io/soc212b

Winter 2025

# Learning goals for today

At the end of class, you will be able to estimate average causal effects by modeling treatment assignment probabilities.

Optional reading:

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

# Review of what we have learned

Causal assumptions

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

# Review of what we have learned

Causal assumptions

$$\vec{X} \longrightarrow A \longrightarrow Y$$

Nonparametric estimator

- Group by $L$, then mean difference in $Y$ over $A$
- Re-aggregate over subgroups

# Review of what we have learned

Causal assumptions

$$\vec{X} \longrightarrow A \longrightarrow Y$$

Nonparametric estimator

- Group by $L$, then mean difference in $Y$ over $A$
- Re-aggregate over subgroups

Outcome modeling estimator

- Model $Y^1$ given $L$ among the treated
- Model $Y^0$ given $L$ among the untreated
- Predict for everyone and take the difference
- Average over all units

# Inverse probability weighting: Population mean

| **Population Outcomes** | **Randomized Sampling** | **Sampled Outcomes** |
|---|---|---|
| $Y_{\textbf{Maria}}$ | $S_{\textbf{Maria}} = 1$ | $Y_{\textbf{Maria}}$ |
| $Y_{\textbf{William}}$ | $S_{\textbf{William}} = 0$ | |
| $Y_{\textbf{Rich}}$ | $S_{\textbf{Rich}} = 0$ | |
| $Y_{\textbf{Sarah}}$ | $S_{\textbf{Sarah}} = 1$ | $Y_{\textbf{Sarah}}$ |
| $Y_{\textbf{Alondra}}$ | $S_{\textbf{Alondra}} = 0$ | |
| $Y_{\textbf{Jesús}}$ | $S_{\textbf{Jesús}} = 1$ | $Y_{\textbf{Jesús}}$ |

How many people do Maria, Sarah, and Jesús each represent?

# Inverse probability weighting: Population mean

| | Population Outcomes | Randomized Sampling | Sampled Outcomes |
|---|---|---|---|
| No Parent Completed College | $Y_{\textbf{Maria}}$ | $S_{\textbf{Maria}} = 1$ | $Y_{\textbf{Maria}}$ |
| | $Y_{\textbf{William}}$ | $S_{\textbf{William}} = 0$ | |
| | $Y_{\textbf{Rich}}$ | $S_{\textbf{Rich}} = 0$ | |
| A Parent Completed College | $Y_{\textbf{Sarah}}$ | $S_{\textbf{Sarah}} = 1$ | $Y_{\textbf{Sarah}}$ |
| | $Y_{\textbf{Alondra}}$ | $S_{\textbf{Alondra}} = 0$ | |
| | $Y_{\textbf{Jesús}}$ | $S_{\textbf{Jesús}} = 1$ | $Y_{\textbf{Jesús}}$ |

How many people do Maria, Sarah, and Jesús each represent?

# Inverse probability weighting: Population mean

Each unit has a probability of being sampled.

$$P(S = 1 \mid \vec{X})$$

If we believe conditionally exchangeable sampling,

$$S \perp\!\!\!\perp Y \mid \vec{X}$$

weight by the inverse probability of sampling.

$$w = \frac{1}{\mathrm{P}(S = 1 \mid \vec{X})}$$

$$\hat{\mathrm{E}}(Y) = \frac{\sum_i w_i y_i}{\sum_i w_i}$$

# Inverse probability weighting: Non-probability sample

Suppose we have the Xbox sample (Wang et al. 2015)

- Imagine we believe conditional exchangeability
- They have the counts $n_{\vec{x}}$ in each demographic subgroup $\vec{x}$ in the sample
- They estimate the population sizes $N_{\vec{x}}$ from exit polls
- Can we estimate by weighting?
  - Assume for simplicity that each $n_{\vec{x}}$ is much greater than 0

# Inverse probability weighting: Non-probability sample

1. Estimate the probability of sampling

$$\hat{\pi}_i = \hat{\mathrm{P}}(S = 1 \mid \vec{X} = \vec{x}_i) = \frac{n_{\vec{X}=\vec{x}_i}}{N_{\vec{X}=\vec{x}_i}} = \frac{\overbrace{\sum_j S_j \mathbb{I}(\vec{X}_j = \vec{x}_i)}^{\substack{\text{Number of sample}\\\text{members who look}\\\text{like unit } i}}}{\underbrace{\sum_j \mathbb{I}(\vec{X}_j = \vec{x}_i)}_{\substack{\text{Number of population}\\\text{members who look}\\\text{like unit } i}}}$$

2. Weight by inverse probability of sampling

$$\hat{\mathrm{E}}(Y) = \frac{\sum_i \hat{w}_i y_i}{\sum_i \hat{w}_i} \qquad \text{for } \hat{w}_i = \frac{1}{\hat{\pi}_i}$$

# Inverse probability weighting: Non-probability sample

Takeaway: Exactly like a probability sample except

- conditional exchangeability holds only by assumption
- inverse probability of sampling weights must be estimated

# Inverse probability weighting: Mean under treatment

$A = 1$ indicates child completed college

| Population Outcomes | Randomized Sampling | Sampled Treatment |
|---|---|---|
| $Y^1_{\text{Maria}}$ | $A_{\text{Maria}} = 1$ | $Y^1_{\text{Maria}}$ |
| $Y^1_{\text{William}}$ | $A_{\text{William}} = 0$ | |
| $Y^1_{\text{Rich}}$ | $A_{\text{Rich}} = 0$ | |
| $Y^1_{\text{Sarah}}$ | $A_{\text{Sarah}} = 1$ | $Y^1_{\text{Sarah}}$ |
| $Y^1_{\text{Alondra}}$ | $A_{\text{Alondra}} = 0$ | |
| $Y^1_{\text{Jesús}}$ | $A_{\text{Jesús}} = 1$ | $Y^1_{\text{Jesús}}$ |

How many people do Maria, Sarah, and Jesús each represent?

# Inverse probability weighting: Mean under treatment

$A = 1$ indicates child completed college

| | Population Outcomes | Randomized Treatment | Sampled Outcomes |
|---|---|---|---|
| No Parent Completed College | $Y^1_{\text{Maria}}$ | $A_{\text{Maria}} = 1$ | $Y^1_{\text{Maria}}$ |
| | $Y^1_{\text{William}}$ | $A_{\text{William}} = 0$ | |
| | $Y^1_{\text{Rich}}$ | $A_{\text{Rich}} = 0$ | |
| A Parent Completed College | $Y^1_{\text{Sarah}}$ | $A_{\text{Sarah}} = 1$ | $Y^1_{\text{Sarah}}$ |
| | $Y^1_{\text{Alondra}}$ | $A_{\text{Alondra}} = 0$ | |
| | $Y^1_{\text{Jesús}}$ | $A_{\text{Jesús}} = 1$ | $Y^1_{\text{Jesús}}$ |

How many people do Maria, Sarah, and Jesús each represent?

# Inverse probability weighting: Mean under treatment

$A = 1$ indicates child completed college. $\vec{X}$ indicates parent completed college.

When estimating the mean outcome under treatment,

$$\text{E}(Y^1)$$

each unit has a probability of being treated.

$$P(A = 1 \mid \vec{X})$$

Weight treated units by the inverse probability of treatment.

$$w = \frac{A}{\text{P}(A = 1 \mid \vec{X})}$$

# Inverse probability weighting: Mean under control

$A = 1$ indicates child completed college

| | Population Outcomes | Randomized Treatment | Sampled Outcomes |
|---|---|---|---|
| No Parent Completed College | $Y^0_{\text{Maria}}$ | $A_{\text{Maria}} = 1$ | |
| | $Y^0_{\text{William}}$ | $A_{\text{William}} = 0$ | $Y^0_{\text{William}}$ |
| | $Y^0_{\text{Rich}}$ | $A_{\text{Rich}} = 0$ | $Y^0_{\text{Rich}}$ |
| A Parent Completed College | $Y^0_{\text{Sarah}}$ | $A_{\text{Sarah}} = 1$ | |
| | $Y^0_{\text{Alondra}}$ | $A_{\text{Alondra}} = 0$ | $Y^0_{\text{Alondra}}$ |
| | $Y^0_{\text{Jesús}}$ | $A_{\text{Jesús}} = 1$ | |

How many people do William, Rich, and Alondra each represent?

# Inverse probability weighting: Mean under control

$A = 1$ indicates child completed college. $\vec{X}$ indicates parent completed college.

When estimating the mean outcome under treatment,

$$\text{E}(Y^0)$$

each unit has a probability of being untreated.

$$P(A = 0 \mid \vec{X})$$

Weight treated units by the inverse probability of treatment.

$$w = \frac{1 - A}{\text{P}(A = 0 \mid \vec{X})}$$

# Inverse probability weighting: Average causal effect

Define inverse probability of treatment weights

$$w_i = \begin{cases} \frac{1}{\mathrm{P}(A=1|\vec{X}=\vec{x}_i)} & \text{if treated} \\ \frac{1}{\mathrm{P}(A=0|\vec{X}=\vec{x}_i)} & \text{if untreated} \end{cases}$$

Estimate each mean potential outcome by a weighted mean

$$\hat{\mathrm{E}}(Y^1) = \sum_{i:A_i=1} w_i Y_i \quad / \quad \sum_{i:A_i=1} w_i$$

$$\hat{\mathrm{E}}(Y^0) = \sum_{i:A_i=0} w_i Y_i \quad / \quad \sum_{i:A_i=0} w_i$$

Take the difference between $\hat{\mathrm{E}}(Y^1)$ and $\hat{\mathrm{E}}(Y^0)$

# Exercise: Weight for ATT

Goal: Average treatment effect on the treated

When $X = 1$,

- 7 treated units
- 3 untreated units
- $\mathrm{P}(A = 1 \mid X = 1) = 0.7$

When $X = 0$,

- 4 treated units
- 6 untreated units
- $\mathrm{P}(A = 1 \mid X = 0) = 0.4$

Each treated unit weighted by 1. Total untreated weight at each $x$ should equal total treated weight.

# Inverse probability weighting: Experiment

Takeaway:

- weight $=$ inverse probability of observed treatment condition
- estimate by weighted means

# Inverse probability weighting: Observational study

Now treatment is not randomly assigned. How do we use weighting?

# Inverse probability weighting: Observational study

Now treatment is not randomly assigned. How do we use weighting?

- assume conditionally exchangeable treatment assignment
- estimate inverse probability of treatment weights

# Inverse probability weighting: Observational study

Model probability of treatment

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

Estimate inverse probability of treatment weights

$$\hat{w}_i = \begin{cases} \frac{1}{\hat{\mathrm{P}}(A=1 \mid \vec{X}=\vec{x}_i)} & \text{if treated} \\ \frac{1}{\hat{\mathrm{P}}(A=0 \mid \vec{X}=\vec{x}_i)} & \text{if untreated} \end{cases}$$

Estimate each mean potential outcome by a weighted mean

$$\hat{\mathrm{E}}(Y^1) = \sum_{i:A_i=1} \hat{w}_i Y_i \quad / \quad \sum_{i:A_i=1} w_i$$

$$\hat{\mathrm{E}}(Y^0) = \sum_{i:A_i=0} \hat{w}_i Y_i \quad / \quad \sum_{i:A_i=0} w_i$$

# Unequal sampling and unequal treatment assignment

# Unequal sampling and unequal treatment assignment

Unit $i$ was sampled with probability 0.25.

$$\mathrm{P}(S = 1 \mid \vec{X} = \vec{x}_i) = \frac{1}{4} = 0.25$$
$$w_i^{\text{Sampling}} = 4$$

# Unequal sampling and unequal treatment assignment

Unit $i$ was sampled with probability 0.25.

$$\mathrm{P}(S = 1 \mid \vec{X} = \vec{x}_i) = \frac{1}{4} = 0.25$$
$$w_i^{\text{Sampling}} = 4$$

Given sampling, received treatment with probability 0.33.

$$\mathrm{P}(A = 1 \mid \vec{X} = \vec{x}_i, S = 1) = \frac{1}{3} = 0.33$$
$$w_i^{\text{Treatment}} = 3$$

# Unequal sampling and unequal treatment assignment

Unit $i$ was sampled with probability 0.25.

$$\text{P}(S = 1 \mid \vec{X} = \vec{x}_i) = \frac{1}{4} = 0.25$$
$$w_i^{\text{Sampling}} = 4$$

Given sampling, received treatment with probability 0.33.

$$\text{P}(A = 1 \mid \vec{X} = \vec{x}_i, S = 1) = \frac{1}{3} = 0.33$$
$$w_i^{\text{Treatment}} = 3$$

How many population $Y^1$ values does unit $i$ represent?

# Unequal sampling and unequal treatment assignment

Unit $i$ was sampled with probability 0.25.

$$\mathrm{P}(S = 1 \mid \vec{X} = \vec{x}_i) = \frac{1}{4} = 0.25$$
$$w_i^{\text{Sampling}} = 4$$

Given sampling, received treatment with probability 0.33.

$$\mathrm{P}(A = 1 \mid \vec{X} = \vec{x}_i, S = 1) = \frac{1}{3} = 0.33$$
$$w_i^{\text{Treatment}} = 3$$

How many population $Y^1$ values does unit $i$ represent?

$$w_i^{\text{Sampling}} \times w_i^{\text{Treatment}} = 4 \times 3 = 12$$

# Unequal sampling and unequal treatment assignment

In math: To observe $Y^1$, a unit must be sampled and treated.

$$\begin{aligned}\text{P}(\text{Observe } Y^1 \mid \vec{X}) &= \text{P}(S = 1, A = 1 \mid \vec{X}) \\ &= \text{P}(A = 1 \mid S = 1, \vec{X})\text{P}(S = 1 \mid \vec{X})\end{aligned}$$

# Unequal sampling and unequal treatment assignment

In math: To observe $Y^1$, a unit must be sampled and treated.

$$\begin{aligned}\text{P}(\text{Observe } Y^1 \mid \vec{X}) &= \text{P}(S = 1, A = 1 \mid \vec{X}) \\ &= \text{P}(A = 1 \mid S = 1, \vec{X})\text{P}(S = 1 \mid \vec{X})\end{aligned}$$

The inverse probability weight is thus the product of sampling and treatment weights.

$$\frac{1}{\text{P}(\text{Observe } Y^1 \mid \vec{X})} = \underbrace{\frac{1}{\text{P}(A = 1 \mid S = 1, \vec{X})}}_{\substack{\text{inverse probability} \\ \text{of treatment weight}}} \times \underbrace{\frac{1}{\text{P}(A = 1 \mid S = 1, \vec{X})}}_{\substack{\text{inverse probability} \\ \text{of sampling weight}}}$$

# Outcome and treatment modeling: A visual summary

Outcome modeling: Model $Y^0$ and $Y^1$ given $\vec{X}$

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

Treatment modeling: Model $A$ given $\vec{X}$. Reweight.

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

Original population

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

Reweighted population

# What are the advantages of each strategy? How to choose?

1. Outcome modeling
   - Model $Y^1$ and $Y^0$ given $\vec{X}$
   - Predict for everyone
   - Unweighted average
2. Treatment modeling
   - Model $A$ given $X$
   - Create weights: how many units each case represents
   - Weighted average

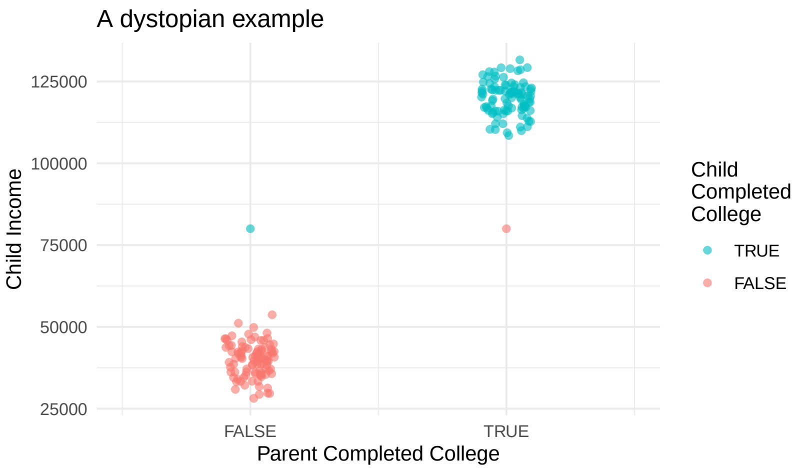
# An advantage of treatment modeling

how most social scientists think about research:
model the outcome

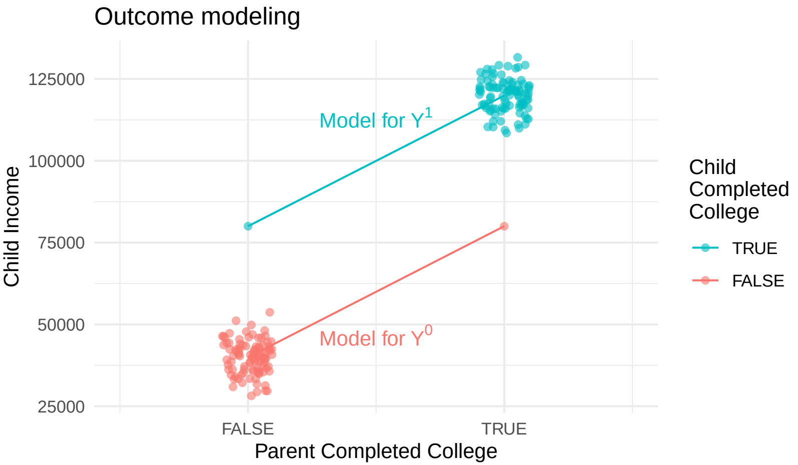
# Advantages of each strategy: Treatment modeling

- how we already think about population sampling: reweight observed cases to learn about all cases
- transparency about influential observations

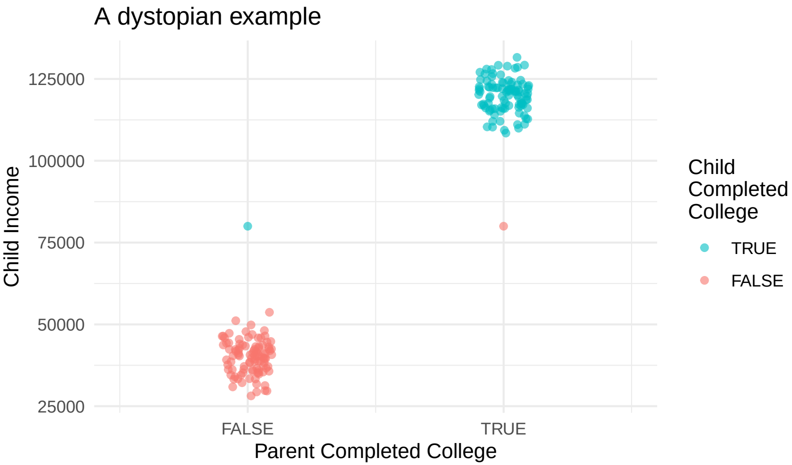
# Transparency about influential observations



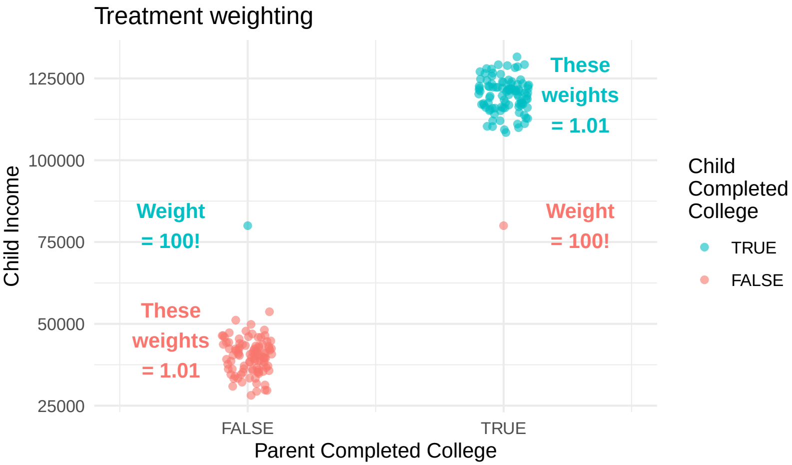
# Transparency about influential observations



# Transparency about influential observations

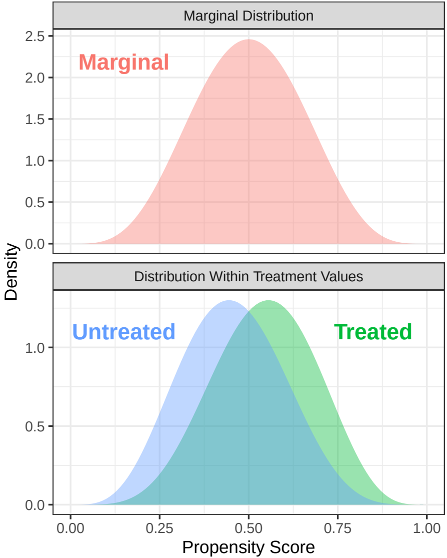


# Transparency about influential observations



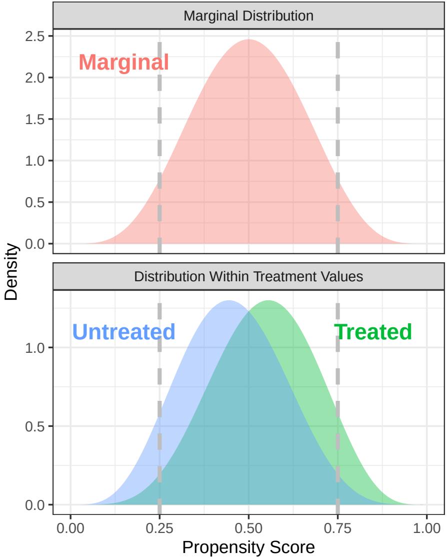
# What to do when some weights are big?

## Focus on a feasible subpopulation: Region of common support



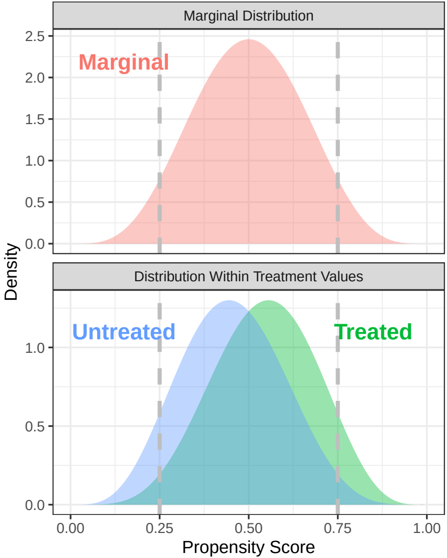
# What to do when some weights are big?

## Focus on a feasible subpopulation: Region of common support



# What to do when some weights are big?

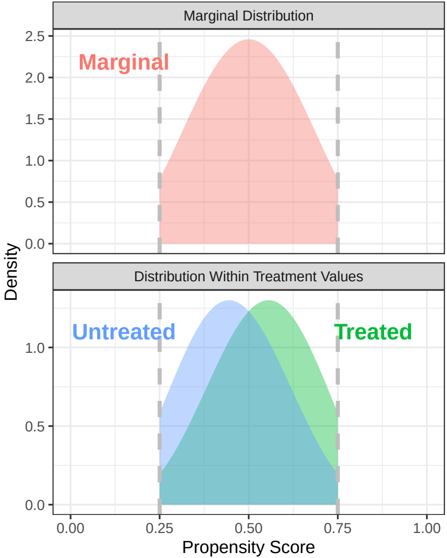
Focus on a feasible subpopulation: Region of common support




Restrict to a subgroup

# What to do when some weights are big?

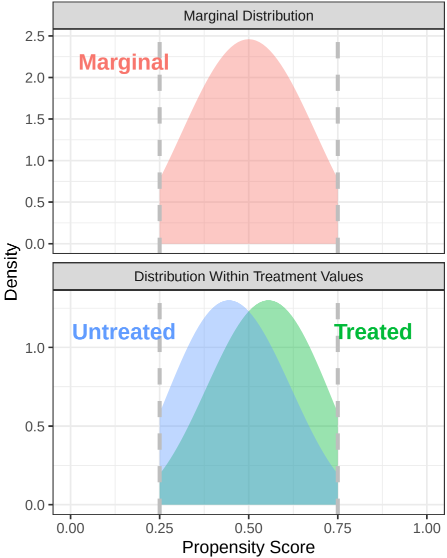
Focus on a feasible subpopulation: Region of common support




Restrict to a subgroup

# What to do when some weights are big?

Focus on a feasible subpopulation: Region of common support




Restrict to a subgroup

Estimate in the subgroup

$$\text{E}\left(Y^1 - Y^0 \mid k_1 < \text{P}(A = 1 \mid \vec{X}) < k_2\right)$$

# Learning goals for today

At the end of class, you will be able to estimate average causal effects by modeling treatment assignment probabilities.

Optional reading:

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