7. Positivity: The problem of empty cells

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

13 Sep 2022

Learning goals for today

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

1. Define positivity
2. Understand how positivity relates to the adjustment set
3. Make estimates for the feasible subpopulation
4. Begin translating ideas to actual data

Positivity

Assumption that $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ (i.e., is positive) for all treatment values a in all population strata $\vec{\ell}$ defined by confounders

Positivity

Assumption that $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ (i.e., is positive) for all treatment values a in all population strata $\vec{\ell}$ defined by confounders

Why this matters:

Positivity

Assumption that $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ (i.e., is positive) for all treatment values a in all population strata $\vec{\ell}$ defined by confounders

Why this matters:
If there is a subgroup $\vec{L} = \vec{\ell}$ where the treatment $A = a$ never happens, then we could never learn from data about the outcome in that subgroup under that treatment.

Running example for today's discussion

Government Assistance Protects Low-Income Families from Eviction

Ian Lundberg
Sarah L. Gold
Louis Donnelly
Jeanne Brooks-Gunn
Sara S. McLanahan

Journal of Policy Analysis and Management

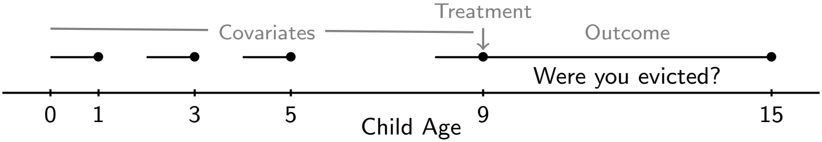
2021

Running example for today's discussion

Does public housing protect families from eviction?

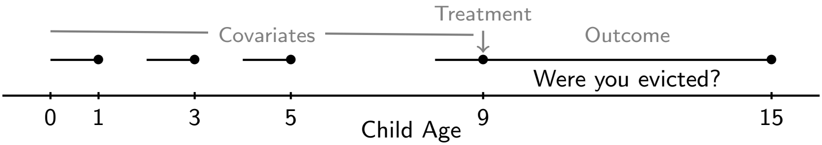
Running example for today's discussion

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1		
	Person 2		
	Person 3		



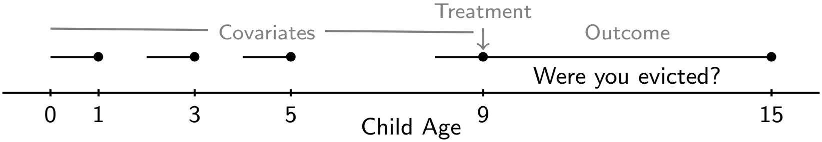
Running example for today's discussion

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	$Y_1(1)$	$Y_1(0)$
	Person 2	$Y_2(1)$	$Y_2(0)$
	Person 3	$Y_3(1)$	$Y_3(0)$



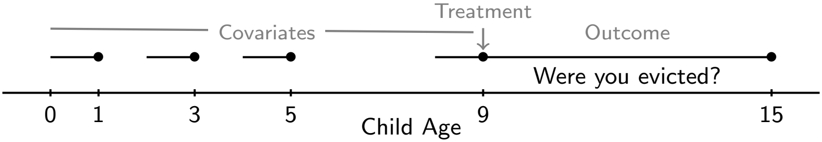
Running example for today's discussion

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	$Y_1(1)$	$Y_1(0)$
	Person 2	$Y_2(1)$	$Y_2(0)$
	Person 3	$Y_3(1)$	$Y_3(0)$



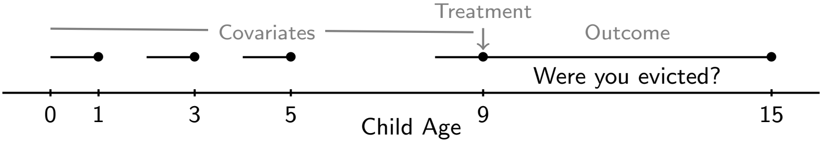
Running example for today's discussion

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	$Y_1(1)$	$Y_1(0)$
	Person 2	$Y_2(1)$	$Y_2(0)$
	Person 3	$Y_3(1)$	$Y_3(0)$



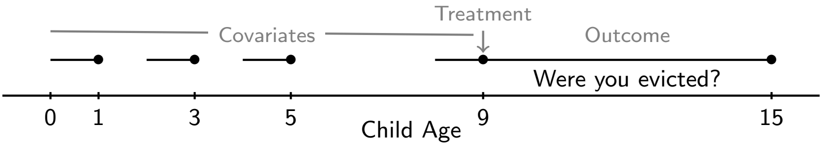
Running example for today's discussion

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	$Y_1(1)$	$Y_1(0)$
	Person 2	$Y_2(1)$	$Y_2(0)$
	Person 3	$Y_3(1)$	$Y_3(0)$



Running example for today's discussion

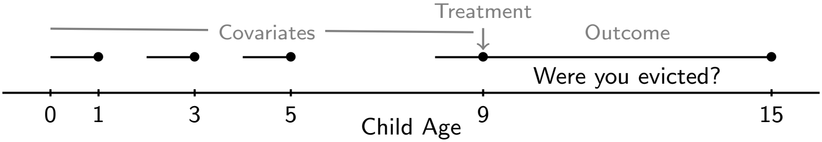
		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	?	$Y_1(0)$
	Person 2	$Y_2(1)$	?
	Person 3	?	$Y_3(0)$



Running example for today's discussion

Learn a prediction function

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	?	$Y_1(0)$
	Person 2	$Y_2(1)$	?
	Person 3	?	$Y_3(0)$



Running example for today's discussion

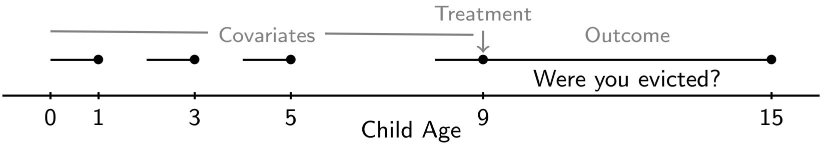
Learn a prediction function

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	?	$Y_1(0)$
	Person 2	$Y_2(1)$	?
	Person 3	?	$Y_3(0)$

Predict the whole table

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	$\hat{Y}_1(1)$	$\hat{Y}_1(0)$
	Person 2	$\hat{Y}_2(1)$	$\hat{Y}_2(0)$
	Person 3	$\hat{Y}_3(1)$	$\hat{Y}_3(0)$

Robins 1986
Hahn 1998



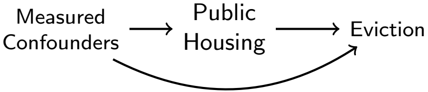
Running example for today's discussion

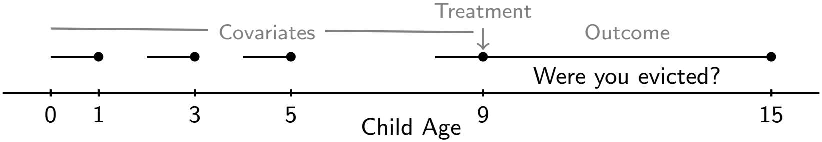
Learn a prediction function

Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	?	$Y_1(0)$
Person 2	$Y_2(1)$	?
Person 3	?	$Y_3(0)$

Predict the whole table

Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	$\hat{Y}_1(1)$	$\hat{Y}_1(0)$
Person 2	$\hat{Y}_2(1)$	$\hat{Y}_2(0)$
Person 3	$\hat{Y}_3(1)$	$\hat{Y}_3(0)$





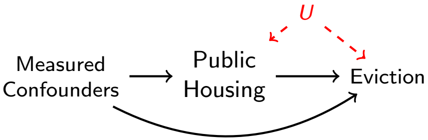
Running example for today's discussion

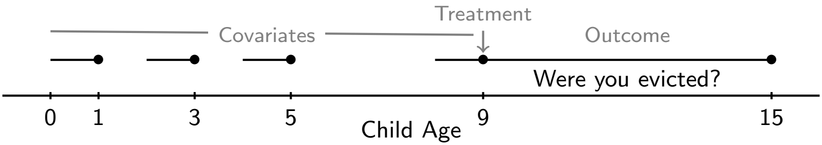
Learn a prediction function

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	?	$Y_1(0)$
	Person 2	$Y_2(1)$	?
	Person 3	?	$Y_3(0)$

Predict the whole table

		Treatment Condition	
		1: Public Housing	0: No Rental Assistance
Population	Person 1	$\hat{Y}_1(1)$	$\hat{Y}_1(0)$
	Person 2	$\hat{Y}_2(1)$	$\hat{Y}_2(0)$
	Person 3	$\hat{Y}_3(1)$	$\hat{Y}_3(0)$





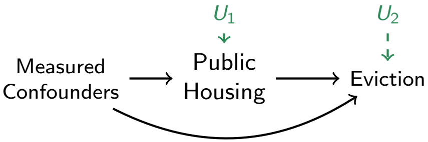
Running example for today's discussion

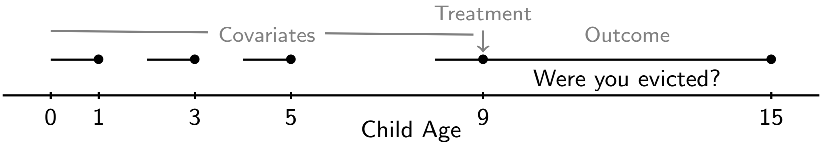
Learn a prediction function

Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	?	$Y_1(0)$
Person 2	$Y_2(1)$	?
Person 3	?	$Y_3(0)$

Predict the whole table

Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	$\hat{Y}_1(1)$	$\hat{Y}_1(0)$
Person 2	$\hat{Y}_2(1)$	$\hat{Y}_2(0)$
Person 3	$\hat{Y}_3(1)$	$\hat{Y}_3(0)$





Running example for today's discussion

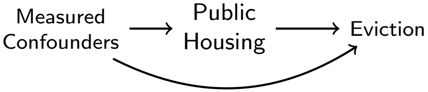
Learn a prediction function

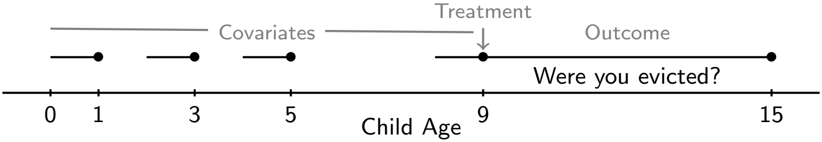
	Treatment Condition	
Population	1: Public Housing	0: No Rental Assistance
Person 1	?	$Y_1(0)$
Person 2	$Y_2(1)$	?
Person 3	?	$Y_3(0)$

Those factually in public housing

Predict the whole table

	Treatment Condition	
Population	1: Public Housing	0: No Rental Assistance
Person 1	$\hat{Y}_1(1)$	$\hat{Y}_1(0)$
Person 2	$\hat{Y}_2(1)$	$\hat{Y}_2(0)$
Person 3	$\hat{Y}_3(1)$	$\hat{Y}_3(0)$





Running example for today's discussion

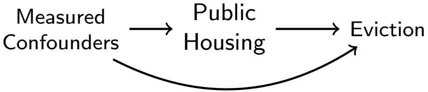
Learn a prediction function

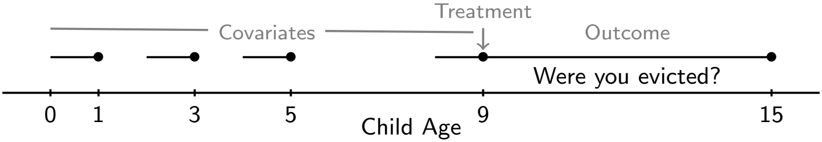
Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	?	$Y_1(0)$
Person 2	$Y_2(1)$	?
Person 3	?	$Y_3(0)$

Those factually in public housing

Predict the whole table

Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	$\hat{Y}_1(1)$	$\hat{Y}_1(0)$
Person 2	$\hat{Y}_2(1)$	$\hat{Y}_2(0)$
Person 3	$\hat{Y}_3(1)$	$\hat{Y}_3(0)$
Average	**3%**	





Running example for today's discussion

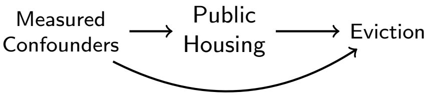
Learn a prediction function

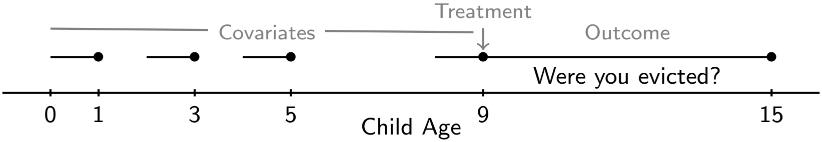
Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	?	$Y_1(0)$
Person 2	$Y_2(1)$	?
Person 3	?	$Y_3(0)$

Those factually in public housing

Predict the whole table

Population	Treatment Condition: 1: Public Housing	Treatment Condition: 0: No Rental Assistance
Person 1	$\hat{Y}_1(1)$	$\hat{Y}_1(0)$
Person 2	$\hat{Y}_2(1)$	$\hat{Y}_2(0)$
Person 3	$\hat{Y}_3(1)$	$\hat{Y}_3(0)$
Average	**3%**	**11%**





Setting 1: A highly simplified DAG

Setting 1: A highly simplified DAG

Setting 1: A highly simplified DAG

Question: How would you estimate in this setting?

— Caveat: Your method must be nonparametric (no regression)

Setting 1: A highly simplified DAG

$$\begin{tikzcd} \text{Lagged Eviction} \arrow[r] \arrow[rr, bend right] & \text{Public Housing} \arrow[r] & \text{Eviction} \end{tikzcd}$$

Question: How would you estimate in this setting?

— Caveat: Your method must be nonparametric (no regression)

1. Among those with (lagged eviction = TRUE), estimate the effect
2. Among those with (lagged eviction = FALSE), estimate the effect
3. Take a weighted average of the two estimates, weighted by the number of cases in each group

Nonparametric estimation: In R code

Nonparametric estimation: In R code

```
# Load packages
library(tidyverse)

# Simulate some data
sim_data <- data.frame(x1 = rbinom(n,1,.5),
                       x2 = rbinom(n,1,.5),
                       x3 = rbinom(n,1,.5)) %>%
  # Generate the treatment
  mutate(a = rbinom(n,1,plogis(x1 + x2 + x3))) %>%
  # Generate the outcome
  mutate(y = rnorm(n, x1 + x2 + x3 + a))
```

Nonparametric estimation: In R code

```
# Define the confounders
confounders <- c("x1","x2","x3")

# Count cases in each stratum
strata_counts <- sim_data %>%
  # Group by the confounders
  group_by(across(all_of(confounders))) %>%
  # Count the number of cases
  summarize(cases = n(),
            .groups = "drop")
```

Nonparametric estimation: In R code

```
# Estimate effect in each stratum
strata_effects <- sim_data %>%
  # Group by the confounders and treatment
  group_by(a, across(all_of(confounders))) %>%
  # Estimate the mean outcome
  summarize(ybar = mean(y),
            .groups = "drop") %>%
  # Prepare to make the data wider by re-valuing the treatment
  mutate(a = paste0("ybar_",a)) %>%
  # Make the data wide
  pivot_wider(names_from = "a", values_from = "ybar") %>%
  # Estimate the effect
  mutate(conditional_effect = ybar_1 - ybar_0)
```

The `pivot_wider` step makes a conversion like this:

Stratum	a	ybar
1	ybar_1	3.6
1	ybar_0	3.2
2	ybar_1	3
2	ybar_0	2.8
⋮	⋮	⋮

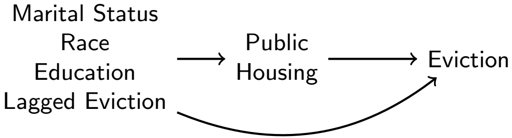
$\rightarrow$

Stratum	ybar_1	ybar_0
1	3.6	3.2
2	3	2.8
⋮	⋮	⋮

Nonparametric estimation: In R code

```
# Aggregate over strata
strata_counts %>%
  # Merge the effects into the counts data frame
  full_join(strata_effects, by = confounders) %>%
  # Stop working within strata. Average the effect
  ungroup() %>%
  summarize(average_effect = weighted.mean(conditional_effect,
                                           w = cases))
```

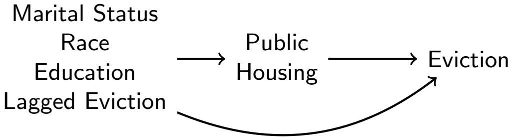
Setting 2: A slightly more complex DAG



Question: How would you estimate in this setting?

— Caveat: Your method must be nonparametric (no regression)

Setting 2: A slightly more complex DAG

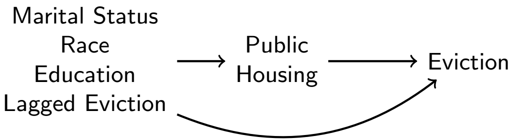


Question: How would you estimate in this setting?

— Caveat: Your method must be nonparametric (no regression)

1. In each subgroup defined by the covariates, estimate the effect
2. Take a weighted average of the two estimates, weighted by the number of cases in each group

Setting 2: A slightly more complex DAG



Question: How would you estimate in this setting?

— Caveat: Your method must be nonparametric (no regression)

1. In each subgroup defined by the covariates, estimate the effect
2. Take a weighted average of the two estimates, weighted by the number of cases in each group

PROBLEM: In some subgroups, the treatment does not vary

To discuss

Situation: In some subgroups, the treatment does not vary

Example: (oversimplified for concreteness)

When a mother has a college degree,
her family is never seen in public housing

Discuss: Does causal inference make sense for this subgroup

1. if this happens in the sample but not the population?
2. if this happens in the sample and in the population?

If it makes sense, how might you go about it?

Positivity assumption (in the population)

$$\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$$

for all treatment values a in the causal estimand
and
for every covariate stratum $\vec{\ell}$ in the population of interest

Positivity assumption (in the population)

$$\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$$

for all treatment values a in the causal estimand
and
for every covariate stratum $\vec{\ell}$ in the population of interest

Guarantees that in an infinite sample you will eventually see the needed treatment values

Positivity assumption (in the population)

A few examples where positivity would not hold:

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?
  - Nonsensical: No 65-year-old can be a pitcher

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?
  - Nonsensical: No 65-year-old can be a pitcher
- Among bulbs planted in Ithaca, would the spring blooms change if it never froze all winter?

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?
  - Nonsensical: No 65-year-old can be a pitcher
- Among bulbs planted in Ithaca, would the spring blooms change if it never froze all winter?
  - Nonsensical: It always freezes in Ithaca

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?
  - Nonsensical: No 65-year-old can be a pitcher
- Among bulbs planted in Ithaca, would the spring blooms change if it never froze all winter?
  - Nonsensical: It always freezes in Ithaca
- Among employees at Microsoft, what is the effect on productivity of searching using Bing vs. Google?

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?
  - Nonsensical: No 65-year-old can be a pitcher
- Among bulbs planted in Ithaca, would the spring blooms change if it never froze all winter?
  - Nonsensical: It always freezes in Ithaca
- Among employees at Microsoft, what is the effect on productivity of searching using Bing vs. Google?
  - Nonsensical: At Microsoft, they all use Bing

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?
  - Nonsensical: No 65-year-old can be a pitcher
- Among bulbs planted in Ithaca, would the spring blooms change if it never froze all winter?
  - Nonsensical: It always freezes in Ithaca
- Among employees at Microsoft, what is the effect on productivity of searching using Bing vs. Google?
  - Nonsensical: At Microsoft, they all use Bing

You can theorize about these questions.

Positivity assumption (in the population)

A few examples where positivity would not hold:

- Among 65-year-old teachers, how much would the earnings go up if they became a pitcher in Major League Baseball?
  - Nonsensical: No 65-year-old can be a pitcher
- Among bulbs planted in Ithaca, would the spring blooms change if it never froze all winter?
  - Nonsensical: It always freezes in Ithaca
- Among employees at Microsoft, what is the effect on productivity of searching using Bing vs. Google?
  - Nonsensical: At Microsoft, they all use Bing

You can theorize about these questions.

But they will never happen—in an infinite sample, you'd never learn the answer.

Empirical positivity (in the sample)

Empirical positivity (in the sample)

An empirical version of positivity:

- Within our sample, we observe every treatment value *a* within every covariate stratum $\vec{\ell}$

Empirical positivity (in the sample)

An empirical version of positivity:

- Within our sample, we observe every treatment value *a* within every covariate stratum $\vec{\ell}$

Note:

- Empirical positivity implies theoretical positivity
  - If we saw all treatments in this subgroup in our sample, they must exist in the population

Empirical positivity (in the sample)

An empirical version of positivity:

- Within our sample, we observe every treatment value *a* within every covariate stratum $\vec{\ell}$

Note:

- Empirical positivity implies theoretical positivity
  - If we saw all treatments in this subgroup in our sample, they must exist in the population
- A lack of empirical positivity does not imply a lack of theoretical positivity
  - All the treatment values may exist in this subgroup, and our sample just happened to miss them

Feasible subsample (and subpopulation)

(No empirical positivity) $\rightarrow$ (No nonparametric SATE estimator)

where SATE is the Sample Average Treatment Effect.

Feasible subsample (and subpopulation)

(No empirical positivity) $\rightarrow$ (No nonparametric SATE estimator)

where SATE is the Sample Average Treatment Effect.

But we can get FSATE: Feasible Sample Average Treatment Effect

- The effect averaged over strata where all treatment values are observed

Feasible subsample (and subpopulation)

(No empirical positivity) $\rightarrow$ (No nonparametric SATE estimator)

where SATE is the Sample Average Treatment Effect.

But we can get FSATE: Feasible Sample Average Treatment Effect

- The effect averaged over strata where all treatment values are observed

We will go do this in R.

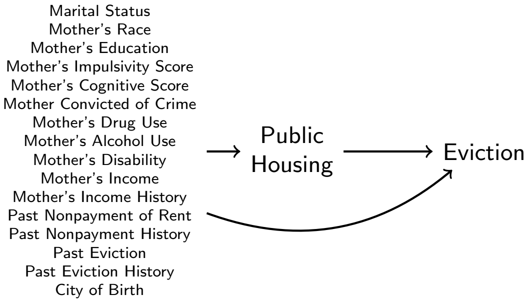
Lots of confounders: Empirical positivity dies

Lots of confounders: Empirical positivity dies

Suppose we have the full adjustment set:

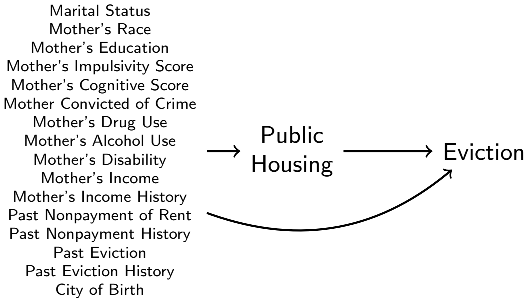
Lots of confounders: Empirical positivity dies

Suppose we have the full adjustment set:



Lots of confounders: Empirical positivity dies

Suppose we have the full adjustment set:



Let's see how many strata have both

- Cases in public housing and
- Cases with no assistance

Conclusion: Positivity is deceptively hard to satisfy

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Conclusion: Positivity is deceptively hard to satisfy

Positivity $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ may seem straightforward.

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Conclusion: Positivity is deceptively hard to satisfy

Positivity $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ may seem straightforward.
It quickly becomes difficult

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Conclusion: Positivity is deceptively hard to satisfy

Positivity $\text{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ may seem straightforward.
It quickly becomes difficult

- Confounding may seem to require many covariates

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Conclusion: Positivity is deceptively hard to satisfy

Positivity $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ may seem straightforward.
It quickly becomes difficult

- Confounding may seem to require many covariates
- But the more confounders, the harder positivity becomes[1]

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Conclusion: Positivity is deceptively hard to satisfy

Positivity $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ may seem straightforward.
It quickly becomes difficult

- Confounding may seem to require many covariates
- But the more confounders, the harder positivity becomes[1]
  - Tons of strata

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Conclusion: Positivity is deceptively hard to satisfy

Positivity $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ may seem straightforward.
It quickly becomes difficult

- Confounding may seem to require many covariates
- But the more confounders, the harder positivity becomes[1]
    - Tons of strata
    - Hard to populate all treatments in all of them

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Conclusion: Positivity is deceptively hard to satisfy

Positivity $\mathrm{P}(A = a \mid \vec{L} = \vec{\ell}) > 0$ may seem straightforward.
It quickly becomes difficult

- Confounding may seem to require many covariates
- But the more confounders, the harder positivity becomes[1]
  - Tons of strata
  - Hard to populate all treatments in all of them
- You end up leaning on a model to extrapolate (next class)

[1] D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. Journal of Econometrics, 221(2), 644-654.

Learning goals for today

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

1. Define positivity
2. Understand how positivity relates to the adjustment set
3. Make estimates for the feasible subpopulation
4. Begin translating ideas to actual data

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!