14. Marginal Structural Models

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

6 Oct 2022

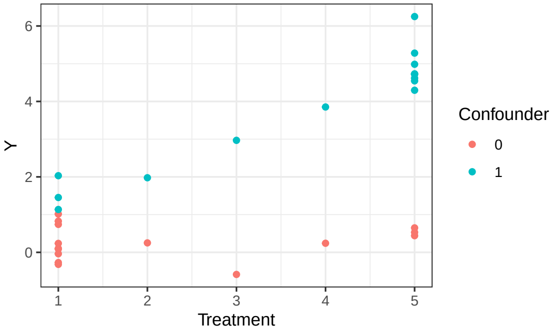
Learning goals for today

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

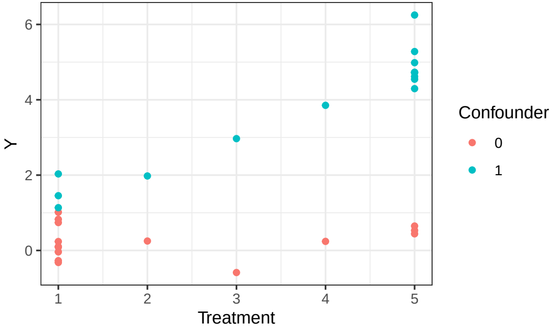
1. Gain efficiency with marginal structural models
2. Recognize how that gain comes through information sharing
3. Understand stabilized weights

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

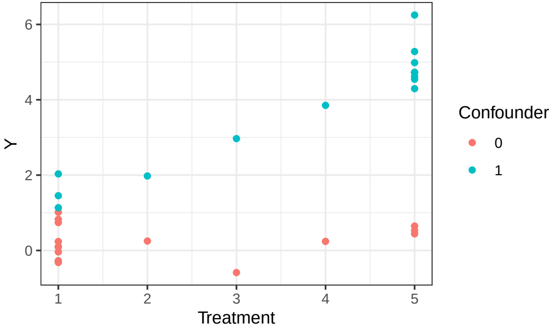


From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\mathrm{E}(Y^3)$?

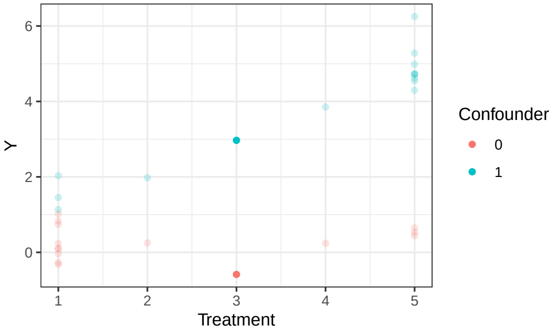
From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\mathrm{E}(Y^3)$?

Inverse probability weighting

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

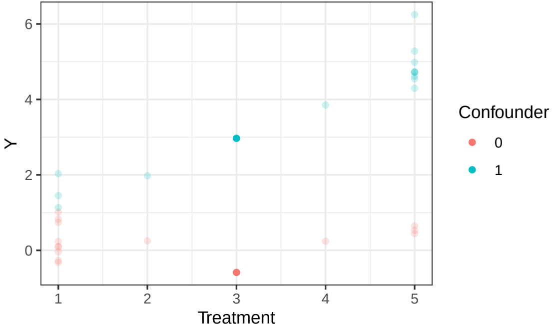


How to estimate $\mathrm{E}(Y^3)$?

Inverse probability weighting

1) Restrict to $A = 3$

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

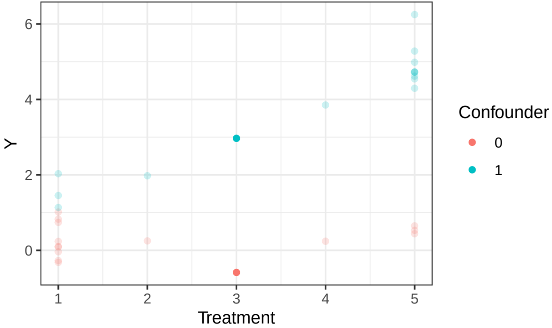


How to estimate $\text{E}(Y^3)$?

Inverse probability weighting

1) Restrict to $A = 3$
2) Take inverse probability weighted average

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\mathrm{E}(Y^3)$?

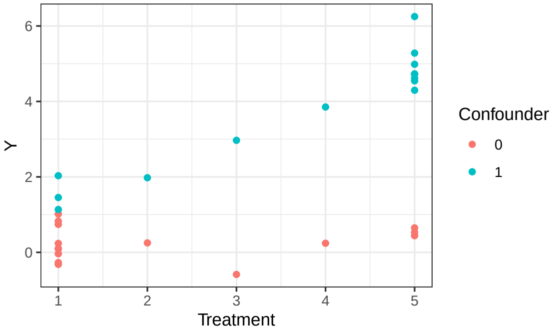
Inverse probability weighting

But only 2 units! High variance!

1) Restrict to $A = 3$

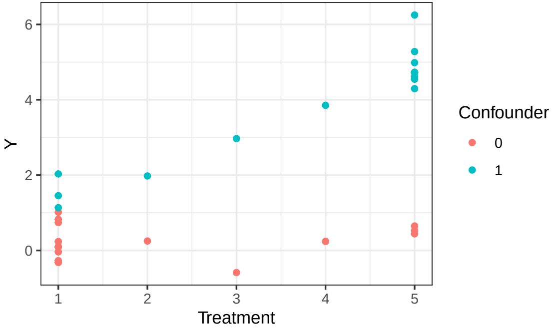
2) Take inverse probability weighted average

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\mathrm{E}(Y^3)$?

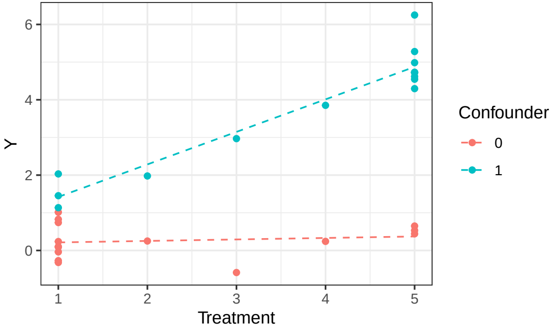
From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\text{E}(Y^3)$?

Outcome modeling

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

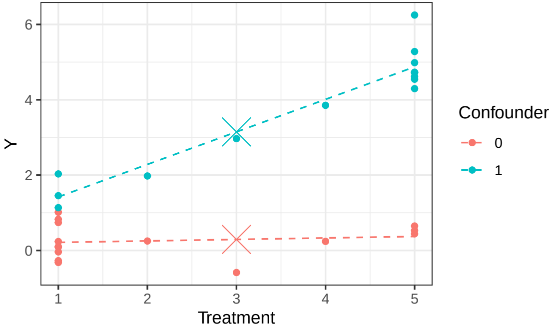


How to estimate $E(Y^3)$?

Outcome modeling

1) Fit a model for $E(Y \mid A, L)$

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

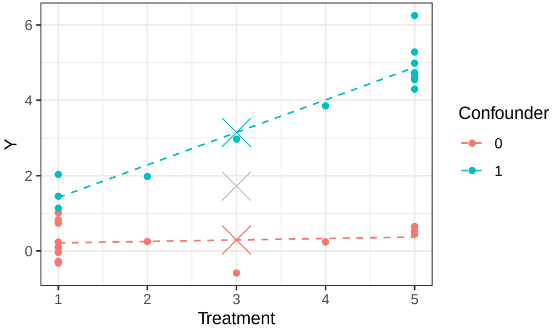


How to estimate $E(Y^3)$?

Outcome modeling

1) Fit a model for $E(Y \mid A, L)$
2) Predict at $A = 3$ in each group

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

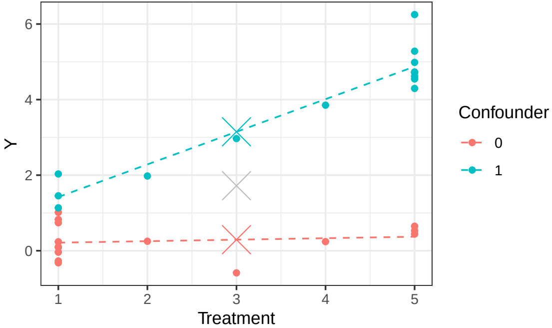


How to estimate $\mathrm{E}(Y^3)$?

Outcome modeling

1) Fit a model for $\mathrm{E}(Y \mid A, L)$
2) Predict at $A = 3$ in each group
3) Average

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



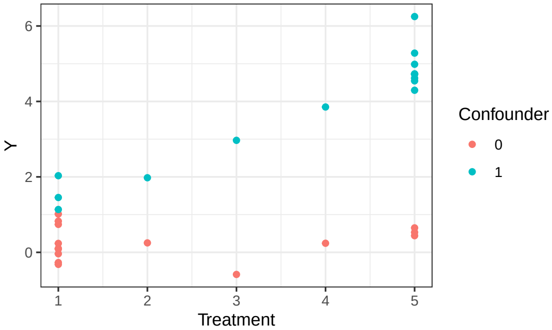
How to estimate $E(Y^3)$?

Outcome modeling

1) Fit a model for $E(Y \mid A, L)$
2) Predict at $A = 3$ in each group
3) Average

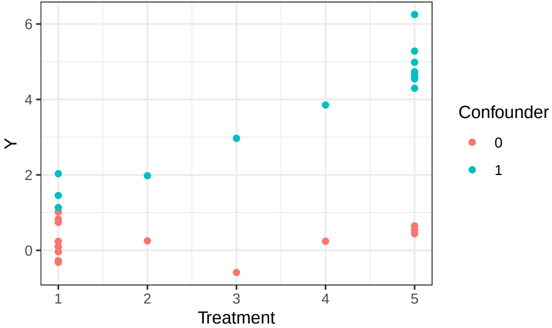
But so much modeling!

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\mathrm{E}(Y^3)$?

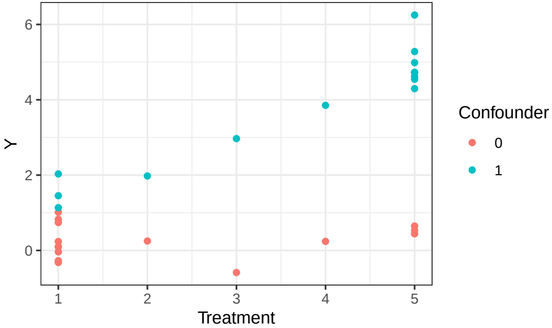
From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\mathrm{E}(Y^3)$?

Marginal structural modeling

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

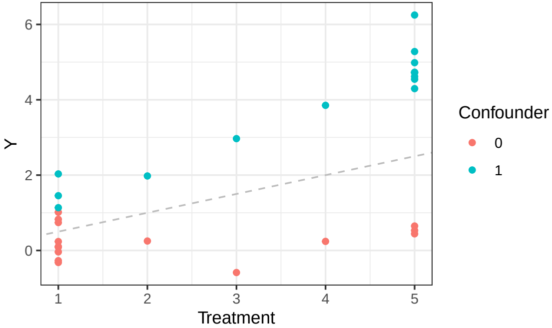


How to estimate $\mathrm{E}(Y^3)$?

Marginal structural modeling

1) Reweight to a pseudo-population (inverse probability weights)

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$

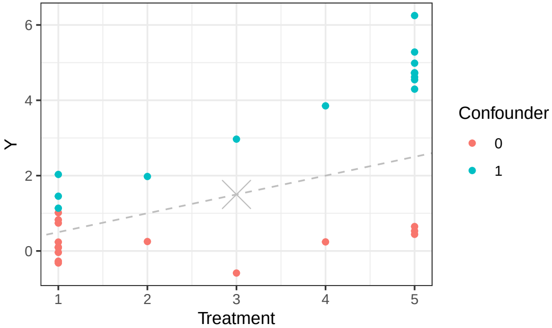


How to estimate $\mathrm{E}(Y^3)$?

Marginal structural modeling

1) Reweight to a pseudo-population (inverse probability weights)
2) Model $\mathrm{E}(Y^a)$ directly

From $A \in \{0, 1\}$ to $A \in \{1, 2, 3, 4, 5\}$



How to estimate $\text{E}(Y^3)$?

Marginal structural modeling

1) Reweight to a pseudo-population (inverse probability weights)
2) Model $\text{E}(Y^a)$ directly
3) Predict at $A = 3$

Reweight to a pseudo-population

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

Reweight to a pseudo-population

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

Within each $\vec{L}$, reweight units
so that every value of A is equally prevalent.

Reweight to a pseudo-population

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

Within each $\vec{L}$, reweight units
so that every value of A is equally prevalent.

- Effectively: Remove the dashed edge

Reweight to a pseudo-population

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

Within each $\vec{L}$, reweight units
so that every value of A is equally prevalent.

- Effectively: Remove the dashed edge

In our pseudo-population, the mean given $A = a$ equals the expected outcome under an intervention to set $A = a$

$$\mathrm{E}_{\text{PseudoPopulation}}(Y \mid A = a) = \mathrm{E}(Y^a)$$

Marginal structural models

Example:

$$\mathsf{E}(Y^a) = \alpha + \beta a$$

Marginal structural models

Example:

$$\mathsf{E}(Y^a) = \alpha + \beta a$$

- Not OLS: Modeling potential rather than factual outcomes

Marginal structural models

Example:

$$\mathrm{E}(Y^a) = \alpha + \beta a$$

- Not OLS: Modeling potential rather than factual outcomes
- Assume a functional form on the thing you will report

Marginal structural models

Example:

$$\text{E}(Y^a) = \alpha + \beta a$$

- Not OLS: Modeling potential rather than factual outcomes
- Assume a functional form on the thing you will report
- Deal with confounding by inverse probability weights

Marginal structural models

Example:

$$\mathsf{E}(Y^a) = \alpha + \beta a$$

- Not OLS: Modeling potential rather than factual outcomes
- Assume a functional form on the thing you will report
- Deal with confounding by inverse probability weights
- Gains efficiency by pooling information across treatments

Marginal structural models

Example:

$$\mathrm{E}(Y^a) = \alpha + \beta a$$

- Not OLS: Modeling potential rather than factual outcomes
- Assume a functional form on the thing you will report
- Deal with confounding by inverse probability weights
- Gains efficiency by pooling information across treatments
  - Very useful when treatment has sparse values

Marginal structural models

Example:

$$E(Y^a) = \alpha + \beta a$$

- Not OLS: Modeling potential rather than factual outcomes
- Assume a functional form on the thing you will report
- Deal with confounding by inverse probability weights
- Gains efficiency by pooling information across treatments
    - Very useful when treatment has sparse values

To estimate:

$$E(Y^a) = E_{\text{PsuedoPopulation}}(Y \mid A = a) = \alpha + \beta a$$

This is OLS weighted to the pseudo-population

Marginal structural models: Concrete steps

Marginal structural models: Concrete steps

1. Assume a DAG where $\vec{L}$ blocks backdoor paths

Marginal structural models: Concrete steps

1. Assume a DAG where $\vec{L}$ blocks backdoor paths
2. Estimate inverse probability weights

$$\hat{w}_i = \frac{1}{\hat{\mathrm{P}}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

Marginal structural models: Concrete steps

1. Assume a DAG where $\vec{L}$ blocks backdoor paths
2. Estimate inverse probability weights

$$\hat{w}_i = \frac{1}{\hat{\text{P}}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

3. Assume a functional form

$$\text{E}(Y^a) = f(a) \quad \text{for some simple function } f()$$

- Example: $\text{E}(Y^a) = \alpha + \beta a$
- "marginal": only modeling as a function of a, not $\vec{L}$
- "structural": causal response to an intervention on A

Marginal structural models: Concrete steps

1. Assume a DAG where $\vec{L}$ blocks backdoor paths
2. Estimate inverse probability weights

$$\hat{w}_i = \frac{1}{\hat{\mathrm{P}}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

3. Assume a functional form

$$\mathrm{E}(Y^a) = f(a) \quad \text{for some simple function } f()$$

   - Example: $\mathrm{E}(Y^a) = \alpha + \beta a$
   - "marginal": only modeling as a function of a, not $\vec{L}$
   - "structural": causal response to an intervention on A

4. Estimate $\hat{\mathrm{E}}(Y^a)$: Weighted regression of Y on A, using $\hat{w}$

Stabilized weights

Standard weights can be high-variance: denominator is small

$$w_i = \frac{1}{\mathrm{P}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

Stabilized weights

Standard weights can be high-variance: denominator is small

$$w_i = \frac{1}{\mathrm{P}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

Stabilized weights can have lower variance

$$w_i = \frac{P(A = a_i)}{\mathrm{P}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

Stabilized weights

Standard weights can be high-variance: denominator is small

$$w_i = \frac{1}{\mathrm{P}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

Stabilized weights can have lower variance

$$w_i = \frac{P(A = a_i)}{\mathrm{P}(A = a_i \mid \vec{L} = \vec{\ell}_i)}$$

This yields efficiency gains only for when the model for $\mathrm{E}(Y^a)$ is not saturated (Hernán & Robins p. 158)

Word of warning: Continuous treatments

When A is continuous, pooling information over A is appealing.

Word of warning: Continuous treatments

When A is continuous, pooling information over A is appealing.

A marginal structural model is possible

$$w_i = \frac{1}{f_{A|\vec{L}_i}(a_i)}$$

where $f_{A|\vec{L}_i}(a_i)$ is the conditional *density* of A given $\vec{L}$.

Word of warning: Continuous treatments

When A is continuous, pooling information over A is appealing.

A marginal structural model is possible

$$w_i = \frac{1}{f_{A|\vec{L}_i}(a_i)}$$

where $f_{A|\vec{L}_i}(a_i)$ is the conditional *density* of A given $\vec{L}$.

But densities are hard to estimate

- Requires not just the mean—the whole distribution
- Can be very sensitive

Word of warning: Continuous treatments

When A is continuous, pooling information over A is appealing.

A marginal structural model is possible

$$w_i = \frac{1}{f_{A|\vec{L}_i}(a_i)}$$

where $f_{A|\vec{L}_i}(a_i)$ is the conditional *density* of A given $\vec{L}$.

But densities are hard to estimate

- Requires not just the mean—the whole distribution
- Can be very sensitive

See Hernán & Robins 12.4.

Learning goals for today

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

1. Gain efficiency with marginal structural models
2. Recognize how that gain comes through information sharing
3. Understand stabilized weights

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!