Parametric g-Formula

Cornell STSCI / INFO / ILRST 3900
Fall 2024
causal3900.github.io

Oct 8 2024

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

Sample vs population

- Conditional Mean: Average outcome for individuals with specific characteristics

Descriptive	Causal
$E(Y \mid A = a, \vec{L} = \vec{\ell})$	$E(Y^a \mid A = a, \vec{L} = \vec{\ell})$

- Population quantities: average outcome for **all units in the population** with specific characteristics

$$E(Y \mid A = a, \vec{L} = \vec{\ell})$$

- Sample conditional mean: average outcome for **units in our sample** with specific characteristics

$$\hat{E}(Y \mid A = a, \vec{L} = \vec{\ell})$$

- Population quantities can be descriptive or causal
- Sample quantities can be descriptive or causal

Sample vs population

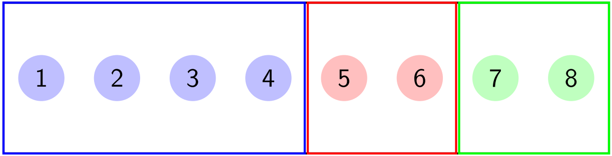
In the experiment from Problem Set 2, what is a population quantity and what is a sample quantity?

- The call back rate for the name Lakisha Washington if we were to send a resume to every single employer in the USA
- The call back rate for the name Lakisha Washington for the 5000 employers which were sent resumes

Standardization

Aggregate the average over sub-groups to get the overall average

$$E(Y^a) = \sum_{\ell} \underbrace{E(Y^a \mid \vec{L} = \vec{\ell})}_{\text{Avg of sub-group}} \quad \times \quad \underbrace{Pr(\vec{L} = \vec{\ell})}_{\text{Prob of sub-group}}$$



$$E(Y^a) = E(Y^a \mid L = {\color{blue}\bullet}) \times \frac{4}{8} + E(Y^a \mid L = {\color{red}\bullet}) \times \frac{2}{8} + E(Y^a \mid L = {\color{green}\bullet}) \times \frac{2}{8}$$

Calculate the same quantity but sum over individuals

$$\begin{aligned} E(Y^a) = (&E(Y^a \mid L = {\color{blue}\text{color}_1}) + E(Y^a \mid L = {\color{blue}\text{color}_2}) + E(Y^a \mid L = {\color{blue}\text{color}_3}) \\ &+ E(Y^a \mid L = {\color{blue}\text{color}_4}) + E(Y^a \mid L = {\color{red}\text{color}_5}) + E(Y^a \mid L = {\color{red}\text{color}_6}) \\ &+ E(Y^a \mid L = {\color{green}\text{color}_7}) + E(Y^a \mid L = {\color{green}\text{color}_8}))\,/8 \end{aligned}$$

Standardization

Aggregate the average over sub-groups to get the overall average

$$\begin{aligned}
\mathsf{E}(Y^a) &= \sum_{\ell} \underbrace{\mathsf{E}(Y^a \mid \vec{L} = \vec{\ell})}_{\text{Avg of sub-group}} \quad \times \quad \underbrace{Pr(\vec{L} = \vec{\ell})}_{\text{Prob of sub-group}} \\
&= \frac{1}{n} \sum_{i} \underbrace{\mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}_i)}_{\text{Avg of sub-group for unit i}}
\end{aligned}$$

Standardization

Aggregate the average over sub-groups to get the overall average

$$\mathrm{E}(Y^a) = \frac{1}{n}\sum_i \underbrace{\mathrm{E}(Y^a \mid L = \ell_i)}_{\text{Avg of sub-group for unit i}}$$

When consistency and conditional exchangeability (given L) hold:

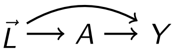
$$\mathrm{E}(Y \mid A = a, L = \ell_i) \overset{\text{consis}}{=} \mathrm{E}(Y^a \mid A = a, L = \ell_i) \quad \overset{\text{exchange}}{=} \mathrm{E}(Y^a \mid L = \ell_i)$$

So we can replace $\mathrm{E}(Y^a \mid L = \ell_i)$ with $\mathrm{E}(Y \mid A = a, L = \ell_i)$ to get

$$\mathrm{E}(Y^a) = \frac{1}{n}\sum_i \mathrm{E}(Y \mid L = \ell_i, A = a)$$

Nonparametric estimation

Causal assumptions



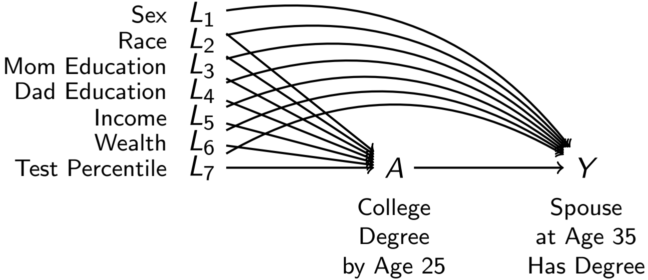
Population quantity:

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

Nonparametric estimator using sample:

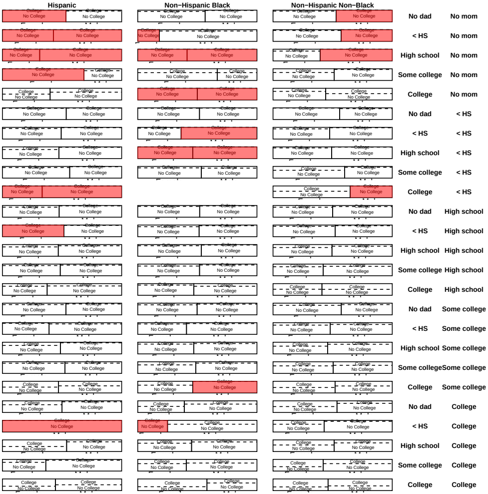
$$\hat{\mathrm{E}}(Y^a) = \frac{1}{n} \sum_i \hat{\mathrm{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

Nonparametric estimation breaks down



Nonparametric estimation breaks down

Can't estimate $\hat{\text{E}}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$ for every sub-group



Parametric estimation: Outcome model

Model the conditional expectation of Y given $\vec{L}$ and A

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

$$\hat{\text{E}}(Y_i \mid \text{Test Score}_i, A_i) = .2 + .003 \times \text{Test Score}_i + .2 \times A_i$$

- If Test Score$_i = 80$ and $A_i = 1$ then

$$\hat{\text{E}}(Y \mid \text{Test Score}, A) = .2 + .003 \times (80) + .2(1) = .64$$

- If Test Score$_i = 62$ and $A_i = 0$ then

$$\hat{\text{E}}(Y \mid \text{Test Score}, A) = .2 + .003 \times (62) + .2(0) = 0.386$$

Parametric estimation: Outcome model

Causal assumptions

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

Parametric estimator

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

Learn a model to predict Y given $\vec{L}$ and A

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

For every unit i,

- Set the treatment value to a
- Predict the outcome

Then average over all units

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

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

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

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

The parametric g-formula allows for more complex models

As long as we have a model for predicting $E(Y \mid L, A)$, we can apply the g-formula

- Linear models with interaction terms
- Other types of regression
- Machine learning models

Parametric g-formula: Outcome model recap

$$\vec{L} \longrightarrow A \longrightarrow Y \quad (\vec{L} \to Y)$$

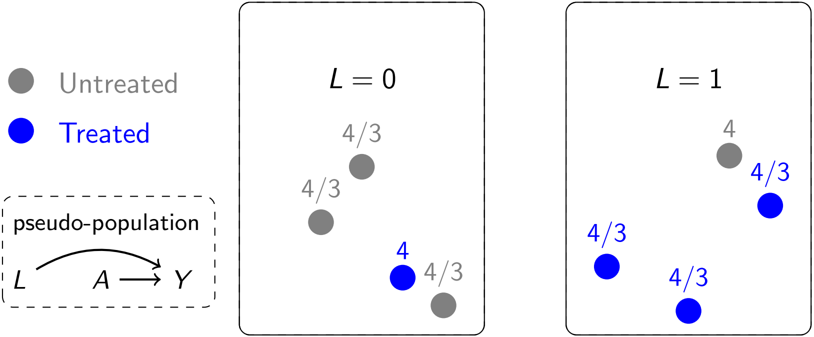
1. Estimate the outcome mean $\mathrm{E}(Y \mid A, \vec{L})$ with some model
2. Change everyone's treatment to the value of interest
3. Predict for everyone
4. Take the average

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


$\vec{L}$
A
Y

$\vec{L}$
A
Y

Inverse probability of treatment weighting



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

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

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 $\mathrm{E}(Y \mid A, \vec{L})$
  - for the treatment $\mathrm{P}(A \mid \vec{L})$

After class:

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