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

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

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

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

E(

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

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

$$1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8$$

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

Calculate the same quantity but sum over individuals $E(Y^{a}) = (E(Y^{a} | L = color_{1}) + E(Y^{a} | L = color_{2}) + E(Y^{a} | L = color_{3}) + E(Y^{a} | L = color_{4}) + E(Y^{a} | L = color_{5}) + E(Y^{a} | L = color_{6}) + E(Y^{a} | L = color_{7}) + E(Y^{a} | L = color_{8})) / 8$

Standardization

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

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

Standardization

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

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

When consistency and conditional exchangeability (given L) hold:

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

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

$$\mathsf{E}(Y^{a}) = \frac{1}{n} \sum_{i} \mathsf{E}(Y \mid L = \ell_{i}, A = a)$$

Nonparametric estimation

Causal assumptions

$$\vec{L} \xrightarrow{A \to Y} Y$$

Population quantity:

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

Nonparametric estimator using sample:

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

Nonparametric estimation breaks down



Nonparametric estimation breaks down

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

Hispanic		Non-Hispanic Black		Non-Hispanic Non-Black			
No College	No College	No College	No College	No College	No College	No dad	No mom
6.000					11.1.		
No College	No College	to College N	i College	Ni Colege	No College	< HS	No mom
No Callege	No College	No College	No College	No College	No College	High school	No mom
		College	Callede	Califor			
No College	No College	No College	No College	No College	No College	Some college	No mom
Ni Calege	No College	No College	No College	No College	College No College	College	No mom
No Callege	No College	No College	No Callege	No-College	No College	No dad	< HS
no compe	No Calify	Per contra	na comp	no comp	na camp	< no	< no
No College	No College	No College	No College	No College	Nu Cullege	High school	< HS
	the Galage					C	
No Callege	No College	No College	No College	No College	No College	Some conege	< HS
No College	No College			No College	No College	College	< HS
No College	No College	No College	No College	No College	No College	No dad	High school
No College	No Catego	No College	No Callege	No College	No Callege	< HS	High school
No College	No Catege	No College	No College	No College	No College	High school	High school
						C	Web estant
No College	No College	No College	No College	No College	No College	Some conege	High school
No College	No College	No College	No College	- Callege No College		College	High school
NuCations	No College	NoColaca	No.College	No Colores	NuCations	No dad	Some college
- Provide		Posseda	11.1.	P			
No College	No College	No Cullege	No College	No College	No College	< HS	Some college
No College	No College	No College	No College	No College	No Callege	High school	Some college
No College	No College	No College	No College	No College	No Callege	Some college	Some college
College No College	College No College	No College	No College	n the second	No College	College	Some college
Na Callege	No College	No College	No College	No College	No College	No dad	College
Ns C	ikepe	No Calege	No College	No College	No College	< HS	College
College Turbular	No College	No College	No College	College No College	No College	High school	College
No College	No College	No College	No College	College	No College	Some college	College
	11.1.5	Ti	11.1.	P.0007.		-	-
- Norodene	Cutege No College	College C.College	No College	College GCGaGe		College	College

Parametric estimation: Outcome model

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

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

 $\hat{\mathsf{E}}(\mathsf{Y}_i \mid \mathsf{Test} \; \mathsf{Score}_i, \mathsf{A}_i) = .2 + .003 \times \mathsf{Test} \; \mathsf{Score}_i + .2 \times \mathsf{A}_i$

Parametric estimation: Outcome model

Causal assumptions

$$\vec{L} \xrightarrow{A \to Y} Y$$

Parametric estimator

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

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

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

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 $E(Y^1) - E(Y^0)$:

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

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 | 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

$$\vec{L} \xrightarrow{A \to Y} Y$$

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

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





Inverse probability of treatment weighting



Propensity score: Inverse probability weight:

$$\pi_i = \mathsf{P}(A = A_i \mid L = L_i)$$
$$w_i = \frac{1}{\pi_i}$$

Model the treatment assignment

$$\hat{\mathsf{P}}(\mathsf{A}=1\mid ec{\mathcal{L}}) = \mathsf{logit}^{-1}\left(\hat{lpha} + \hat{ec{\gamma}}ec{\mathcal{L}}
ight)$$

Predict the propensity score for each unit

$$\hat{\pi}_{i} = \begin{cases} \mathsf{logit}^{-1} \left(\hat{\alpha} + \hat{\vec{\gamma}} \vec{L} \right) & \text{if } A_{i} = 1 \\ 1 - \mathsf{logit}^{-1} \left(\hat{\alpha} + \hat{\vec{\gamma}} \vec{L} \right) & \text{if } A_{i} = 0 \end{cases}$$

Estimate by inverse probability weighting

$$\hat{\mathsf{E}}(Y^{a}) = \frac{1}{N} \sum_{i:A_{i}=a} \frac{Y_{i}}{\hat{\pi}_{i}}$$

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

After class:

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