8. Parametric g-formula: Categorical treatments

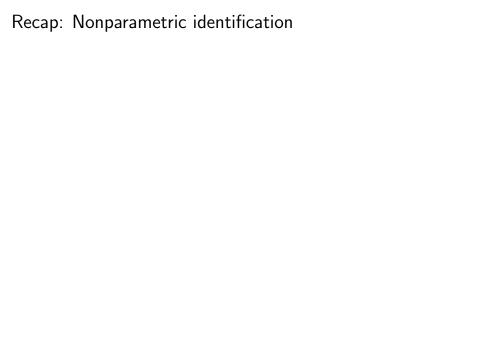
Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

15 Sep 2022

Learning goals for today

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

- 1. Estimate causal effects by outcome modeling with the parametric g-formula
- 2. See how this generalizes a common use of regression



Recap: Nonparametric identification

Three key assumptions:

$$\begin{array}{ll} \text{Consistency} & Y_i = Y_i^{A_i} \\ \text{Exchangeability} & A \perp \!\!\! \perp \{Y^a\} \mid \vec{L} \\ \text{Positivity} & \mathsf{P}(A = a \mid \vec{L} = \vec{\ell}) > 0 \\ \end{array}$$

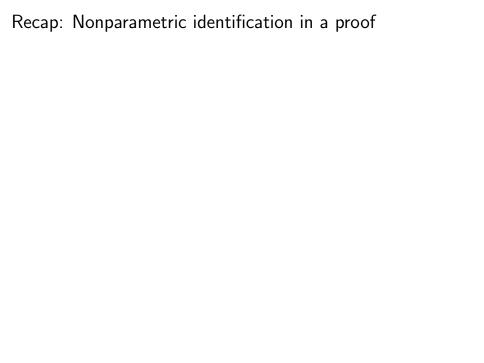
Recap: Nonparametric identification

Three key assumptions:

Consistency
$$Y_i = Y_i^{A_i}$$

Exchangeability $A \perp \{Y^a\} \mid \vec{L}$
Positivity $P(A = a \mid \vec{L} = \vec{\ell}) > 0$

These assumptions yield **nonparametric identification:** a consistent estimator exists using observable sample means



 $E(Y^a)$

$$\mathsf{E}(Y^a) = \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y^a \mid \vec{L} = \vec{\ell})$$

rules of probability

$$\mathsf{E}(Y^a) = \sum_{ec{\ell}} \mathsf{P}(ec{L} = ec{\ell}) \mathsf{E}(Y^a \mid ec{L} = ec{\ell})$$
 rules of probability
$$= \sum_{ec{\ell}} \mathsf{P}(ec{L} = ec{\ell}) \mathsf{E}(Y^a \mid ec{L} = ec{\ell}, A = a)$$
 exchangeability

$$\mathsf{E}(Y^a) = \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}) \qquad \text{rules of probability}$$

$$= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}, A = a) \qquad \text{exchangeability}$$

$$= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y \mid \vec{L} = \vec{\ell}, A = a) \qquad \text{consistency}$$

$$\mathsf{E}(Y^a) = \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}) \qquad \text{rules of probability}$$

$$= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}, A = a) \qquad \text{exchangeability}$$

$$= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y \mid \vec{L} = \vec{\ell}, A = a) \qquad \text{consistency}$$

Positivity ensures $\mathsf{E}(Y\mid \vec{L}=\vec{\ell}, A=a)$ can be estimated from data.

$$\begin{split} \mathsf{E}(Y^a) &= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}) & \text{rules of probability} \\ &= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y^a \mid \vec{L} = \vec{\ell}, A = a) & \text{exchangeability} \\ &= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y \mid \vec{L} = \vec{\ell}, A = a) & \text{consistency} \end{split}$$

Positivity ensures $\mathsf{E}(Y\mid \vec{L}=\vec{\ell}, A=a)$ can be estimated from data.

What we gained: In an infinite sample, we can estimate causal effects by taking means and aggregating!

For simplicity, rewrite the previous result with two expectations

For simplicity, rewrite the previous result with two expectations

$$\mathsf{E}(Y^a) = \sum_{ec{\ell}} \mathsf{P}(ec{L} = ec{\ell}) \mathsf{E}(Y \mid ec{L} = ec{\ell}, A = a)$$

For simplicity, rewrite the previous result with two expectations

$$\begin{aligned} \mathsf{E}(Y^{a}) &= \sum_{\vec{\ell}} \mathsf{P}(\vec{L} = \vec{\ell}) \mathsf{E}(Y \mid \vec{L} = \vec{\ell}, A = a) \\ &= \mathsf{E}\Big(\mathsf{E}(Y \mid \vec{L}, A = a)\Big) \end{aligned}$$

For simplicity, rewrite the previous result with two expectations

$$E(Y^{a}) = \sum_{\vec{\ell}} P(\vec{L} = \vec{\ell}) E(Y \mid \vec{L} = \vec{\ell}, A = a)$$
$$= E\left(E(Y \mid \vec{L}, A = a)\right)$$

There are two pieces

For simplicity, rewrite the previous result with two expectations

$$E(Y^{a}) = \sum_{\vec{\ell}} P(\vec{L} = \vec{\ell}) E(Y \mid \vec{L} = \vec{\ell}, A = a)$$
$$= E\left(E(Y \mid \vec{L}, A = a)\right)$$

There are two pieces

- ► Inner expectation: Within groups
 - ► Mean over units within each stratum of confounders

For simplicity, rewrite the previous result with two expectations

$$E(Y^{a}) = \sum_{\vec{\ell}} P(\vec{L} = \vec{\ell}) E(Y \mid \vec{L} = \vec{\ell}, A = a)$$
$$= E\left(E(Y \mid \vec{L}, A = a)\right)$$

There are two pieces

- ► Inner expectation: Within groups
 - ► Mean over units within each stratum of confounders
- ► Outer expectation: Across groups
 - ightharpoonup Mean of that over the population distribution of \vec{L}

Nonparametric identification yielded this estimand:

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

Nonparametric identification yielded this estimand:

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

This points toward a **non-parametric plug-in estimator**:

Nonparametric identification yielded this estimand:

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

This points toward a **non-parametric plug-in estimator**:

► Inner expectation: Sample mean in each subgroup

Nonparametric identification yielded this estimand:

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

This points toward a **non-parametric plug-in estimator**:

- ► Inner expectation: Sample mean in each subgroup
- ► Outer expectation: Mean over full sample

Nonparametric identification yielded this estimand:

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

This points toward a non-parametric plug-in estimator:

- ► Inner expectation: Sample mean in each subgroup
- ► Outer expectation: Mean over full sample

Resulting estimator:

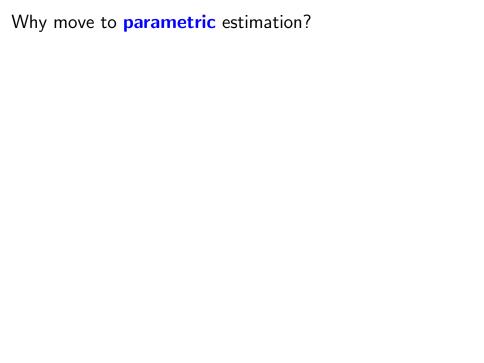
$$\hat{E}(Y^a) = \frac{1}{n} \sum_{i=1}^n \hat{E}(Y \mid \vec{L} = \vec{\ell}_i, A = a)$$

Putting this in words:

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

- ► Sample average over units *i*
- ► For each unit, take the sample average outcome among
 - ▶ Units with their covariate values $\vec{\ell}_i$
 - ▶ But who have the relevant treatment value A = a

It is all just sample means, aggregated a certain way.



Some reasons for a parametric model:

Some reasons for a parametric model:

Some reasons for a parametric model:

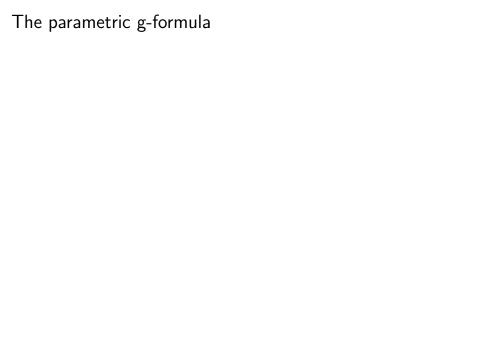
1. **Empty cells:** Despite positivity in the population, some sample strata have no units with A = a. Nonparametric estimation is impossible

Some reasons for a parametric model:

- 1. Empty cells: Despite positivity in the population, some sample strata have no units with A=a. Nonparametric estimation is impossible
- 2. Continuous confounders: If any variables in \vec{L} are continuous, then each stratum of \vec{L} contains only one unit. Empty cells are inevitable

Some reasons for a parametric model:

- 1. Empty cells: Despite positivity in the population, some sample strata have no units with A=a. Nonparametric estimation is impossible
- 2. Continuous confounders: If any variables in \vec{L} are continuous, then each stratum of \vec{L} contains only one unit. Empty cells are inevitable
- Estimation variance: Even if you can do nonparametric estimation, it involves estimating numerous means. This may be a high-variance approach, resulting in extensive statistical uncertainty



The parametric g-formula

Nonparametric identification yielded this estimand:

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

The parametric g-formula

Nonparametric identification yielded this estimand:

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

What if we estimated $\mathsf{E}(Y\mid \vec{L}=\vec{\ell}, A=a)$ with a regression model?

The parametric g-formula: Simple case

$$L \longrightarrow A \longrightarrow$$

The parametric g-formula: Simple case

$$L \longrightarrow A \rightrightarrows Y$$

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

The parametric g-formula: Simple case

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

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

Parametric model $E(Y \mid L, A) = \alpha + \gamma L + \beta A$

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

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

Parametric model
$$E(Y \mid L, A) = \alpha + \gamma L + \beta A$$

Estimator for the potential outcome under treatment (A = 1):

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

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

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

Parametric model
$$E(Y \mid L, A) = \alpha + \gamma L + \beta A$$

Estimator for the potential outcome under treatment (A = 1):

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

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

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

Parametric model $E(Y \mid L, A) = \alpha + \gamma L + \beta A$

Estimator for the potential outcome under treatment (A = 1):

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

Estimator in words:

1. Estimate the regression model

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

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

Parametric model

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

Estimator for the potential outcome under treatment (A = 1):

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

- 1. Estimate the regression model
- 2. Change all treatment values to 1

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

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

Parametric model

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

Estimator for the potential outcome under treatment (A = 1):

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

- 1. Estimate the regression model
- 2. Change all treatment values to 1
- 3. Predict for everyone

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

Nonparametric identification $E(Y^a) = E(E(Y \mid L, A = a))$

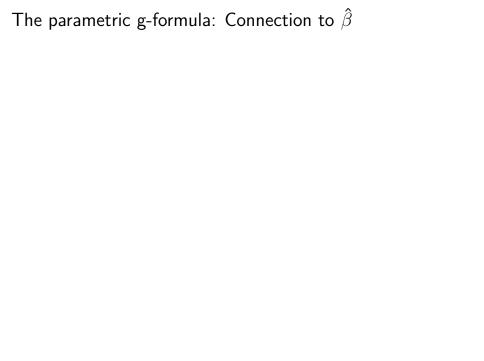
Parametric model

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

Estimator for the potential outcome under treatment (A = 1):

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

- 1. Estimate the regression model
- 2. Change all treatment values to 1
- 3. Predict for everyone
- 4. Take the sample mean



$$\hat{\mathsf{E}}(Y^1) - \hat{\mathsf{E}}(Y^0)$$

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

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

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

Estimator for the effect $E(Y^1) - E(Y^0)$:

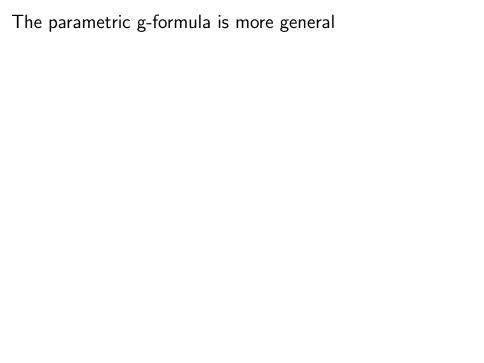
$$\hat{\mathsf{E}}(Y^1) - \hat{\mathsf{E}}(Y^0) = \left(\frac{1}{n} \sum_{i=1}^n \left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 1\right)\right)$$
$$-\left(\frac{1}{n} \sum_{i=1}^n \left(\hat{\alpha} + \hat{\gamma}\ell_i + \hat{\beta} \times 0\right)\right)$$
$$= \frac{1}{n} \sum_{i=1}^n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

 $=\hat{\beta}$

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.



Suppose we add the $A \times L$ interaction to the model.

$$E(Y \mid A, L) = \alpha + \gamma L + \beta A + \eta A L$$

Suppose we add the $A \times L$ interaction to the model.

$$E(Y \mid A, L) = \alpha + \gamma L + \beta A + \eta A L$$

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

Suppose we add the $A \times L$ interaction to the model.

$$E(Y \mid A, L) = \alpha + \gamma L + \beta A + \eta A L$$

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

$$\hat{\mathsf{E}}(Y^1) - \hat{\mathsf{E}}(Y^0)$$

Suppose we add the $A \times L$ interaction to the model.

$$E(Y \mid A, L) = \alpha + \gamma L + \beta A + nAL$$

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

Suppose we add the $A \times L$ interaction to the model.

$$E(Y \mid A, L) = \alpha + \gamma L + \beta A + \eta A L$$

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

Suppose we add the $A \times L$ interaction to the model.

$$E(Y \mid A, L) = \alpha + \gamma L + \beta A + \eta A L$$

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

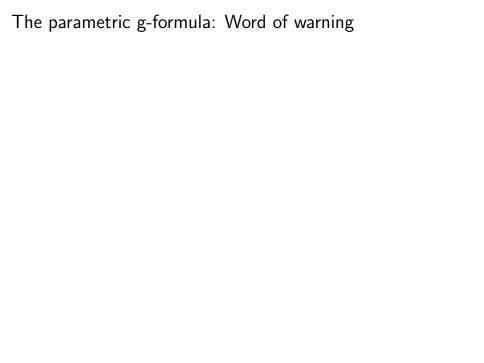
Suppose we add the $A \times L$ interaction to the model.

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

Estimator for the effect $E(Y^1) - E(Y^0)$:

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

The g-formula no longer collapses to a coefficient!



The parametric g-formula does not replace any assumptions.

The parametric g-formula does not replace any assumptions.

You need:

```
Identification Assumptions (causal)
```

Consistency
$$Y_i = Y_i^{A_i}$$

Exchangeability
$$A \perp \{Y^a\} \mid \vec{L}$$

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

Estimation
$$E(Y \mid A, \vec{L})$$
 follows the assumed model

The parametric g-formula does not replace any assumptions.

You need:

```
Identification Assumptions (causal)
```

Consistency
$$Y_i = Y_i^{A_i}$$

Exchangeability
$$A \perp \{Y^a\} \mid \vec{L}$$

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

Estimation
$$E(Y \mid A, \vec{L})$$
 follows the assumed model

To use a parametric model, we added the bottom assumption.

The parametric g-formula does not replace any assumptions.

You need:

```
Identification Assumptions (causal)
```

Consistency
$$Y_i = Y_i^{A_i}$$

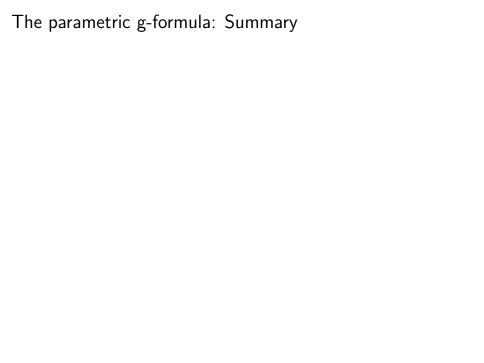
Exchangeability
$$A \perp \{Y^a\} \mid \vec{L}$$

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

Estimation
$$E(Y \mid A, \vec{L})$$
 follows the assumed model

To use a parametric model, we added the bottom assumption.

▶ But testable! Diagnostics and out-of-sample performance



The parametric g-formula is how we answer nonparametric causal questions with parametric outcome modeling tools

► Allows you to use a model for $E(Y \mid A, \vec{L})$

- ► Allows you to use a model for $E(Y \mid A, \vec{L})$
- ► The model is exactly as in standard non-causal statistics

- ► Allows you to use a model for $E(Y \mid A, \vec{L})$
- ► The model is exactly as in standard non-causal statistics
- ► The estimator is operationalized by prediction:

- ► Allows you to use a model for $E(Y \mid A, \vec{L})$
- ► The model is exactly as in standard non-causal statistics
- ► The estimator is operationalized by prediction:
 - ► Estimate the regression

- ► Allows you to use a model for $E(Y \mid A, \vec{L})$
- ► The model is exactly as in standard non-causal statistics
- ► The estimator is operationalized by prediction:
 - ► Estimate the regression
 - Modify the treatment

- ► Allows you to use a model for $E(Y \mid A, \vec{L})$
- ► The model is exactly as in standard non-causal statistics
- ► The estimator is operationalized by prediction:
 - ► Estimate the regression
 - Modify the treatment
 - Make new predictions

- ► Allows you to use a model for $E(Y \mid A, \vec{L})$
- ► The model is exactly as in standard non-causal statistics
- ► The estimator is operationalized by prediction:
 - ► Estimate the regression
 - Modify the treatment
 - ▶ Make new predictions
 - Average over the sample

Learning goals for today

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

- 1. Estimate causal effects by outcome modeling with the parametric g-formula
- 2. See how this generalizes a common use of regression

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!