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

Recap: Nonparametric identification

Three key assumptions:

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

Recap: Nonparametric identification

Three key assumptions:

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

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

Recap: Nonparametric identification in a proof

Recap: Nonparametric identification in a proof

$\mathrm{E}(Y^a)$

Recap: Nonparametric identification in a proof

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

Recap: Nonparametric identification in a proof

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

Recap: Nonparametric identification in a proof

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

Recap: Nonparametric identification in a proof

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

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

Recap: Nonparametric identification in a proof

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

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

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

Recap: Nonparametric identification with two Es

For simplicity, rewrite the previous result with two expectations

Recap: Nonparametric identification with two Es

For simplicity, rewrite the previous result with two expectations

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

Recap: Nonparametric identification with two Es

For simplicity, rewrite the previous result with two expectations

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

Recap: Nonparametric identification with two Es

For simplicity, rewrite the previous result with two expectations

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

There are two pieces

Recap: Nonparametric identification with two Es

For simplicity, rewrite the previous result with two expectations

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

There are two pieces

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

Recap: Nonparametric identification with two Es

For simplicity, rewrite the previous result with two expectations

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

There are two pieces

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

Recap: Nonparametric identification. An estimator

Nonparametric identification yielded this estimand:

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

Recap: Nonparametric identification. An estimator

Nonparametric identification yielded this estimand:

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

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

Recap: Nonparametric identification. An estimator

Nonparametric identification yielded this estimand:

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

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

- Inner expectation: Sample mean in each subgroup

Recap: Nonparametric identification. An estimator

Nonparametric identification yielded this estimand:

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

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

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

Recap: Nonparametric identification. An estimator

Nonparametric identification yielded this estimand:

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

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

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

Resulting estimator:

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

Recap: Nonparametric identification. An estimator

Putting this in words:

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

- 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.

Why move to **parametric** estimation?

Why move to **parametric** estimation?

Some reasons for a parametric model:

Why move to **parametric** estimation?

Some reasons for a parametric model:

Why move to **parametric** estimation?

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

Why move to **parametric** estimation?

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

Why move to **parametric** estimation?

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
3. **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

The parametric g-formula

Nonparametric identification yielded this estimand:

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

The parametric g-formula

Nonparametric identification yielded this estimand:

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

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

The parametric g-formula: Simple case

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

The parametric g-formula: Simple case

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

Nonparametric identification $\quad \mathrm{E}(Y^a) = \mathrm{E}(\mathrm{E}(Y \mid L, A = a))$

The parametric g-formula: Simple case

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

Nonparametric identification	$\mathrm{E}(Y^a) = \mathrm{E}(\mathrm{E}(Y \mid L, A = a))$
Parametric model	$\mathrm{E}(Y \mid L, A) = \alpha + \gamma L + \beta A$

The parametric g-formula: Simple case

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

Nonparametric identification $\quad \mathrm{E}(Y^a) = \mathrm{E}(\mathrm{E}(Y \mid L, A = a))$

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

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

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

The parametric g-formula: Simple case

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

Nonparametric identification $\quad \mathrm{E}(Y^a) = \mathrm{E}(\mathrm{E}(Y \mid L, A = a))$

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

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

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

Estimator in words:

The parametric g-formula: Simple case

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

Nonparametric identification $\quad \mathsf{E}(Y^a) = \mathsf{E}(\mathsf{E}(Y \mid L, A = a))$

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

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

The parametric g-formula: Simple case

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

Nonparametric identification $\quad \mathsf{E}(Y^a) = \mathsf{E}(\mathsf{E}(Y \mid L, A = a))$

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

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
2. Change all treatment values to 1

The parametric g-formula: Simple case

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

Nonparametric identification	$\mathrm{E}(Y^a) = \mathrm{E}(\mathrm{E}(Y \mid L, A = a))$
Parametric model	$\mathrm{E}(Y \mid L, A) = \alpha + \gamma L + \beta A$

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

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

Estimator in words:

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

The parametric g-formula: Simple case

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

Nonparametric identification $\quad \mathrm{E}(Y^a) = \mathrm{E}(\mathrm{E}(Y \mid L, A = a))$

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

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

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

Estimator in words:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The parametric g-formula is more general

The parametric g-formula is more general

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

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

The parametric g-formula is more general

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

The parametric g-formula is more general

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

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

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

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

The parametric g-formula is more general

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

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

The parametric g-formula is more general

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

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

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

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

The parametric g-formula is more general

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

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

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

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

The parametric g-formula is more general

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

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

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

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

The g-formula no longer collapses to a coefficient!

The parametric g-formula: Word of warning

The parametric g-formula: Word of warning

The parametric g-formula does not replace any assumptions.

The parametric g-formula: Word of warning

The parametric g-formula does not replace any assumptions.

You need:

Identification Assumptions (causal)

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

Estimation Assumptions (statistical)

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

The parametric g-formula: Word of warning

The parametric g-formula does not replace any assumptions.

You need:

Identification Assumptions (causal)

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

Estimation Assumptions (statistical)

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

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

The parametric g-formula: Word of warning

The parametric g-formula does not replace any assumptions.

You need:

Identification Assumptions (causal)

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

Estimation Assumptions (statistical)

Estimation	$\mathrm{E}(Y \mid A, \vec{L})$ 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: Summary

The parametric g-formula: Summary

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

The parametric g-formula: Summary

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

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

The parametric g-formula: Summary

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

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

The parametric g-formula: Summary

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

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

The parametric g-formula: Summary

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

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

The parametric g-formula: Summary

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

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

The parametric g-formula: Summary

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

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

The parametric g-formula: Summary

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

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