Causal Machine Learning

Week 1

Booth School Of Business
University of Chicago

Max Farrell & Sanjog Misra

Causality

> *The paradigmatic assertion in causal relationships is that manipulation of a cause will result in the manipulation of an effect*. . . . Causation implies that by varying one factor I can make another vary.
>
> (Cook & Campbell 1979: 36, emphasis in original)

Causal Constructs

- What types of outcomes do social scientists care about?
- What types of treatments?
- What objects do they wish to estimate?
- *Why* do they wish to estimate these?

Causal Constructs

- Consider the following:

$$w = \alpha + \beta \text{ Gender } + \delta' x + \epsilon$$

- What is the interpretation of β?
- What would you include in x? Does this change the interpretation of β?
- How should we think of Bertrand & Mullainathan (AER 2004)?

Why estimate causal effects?

- **Hypotesis testing**
- **Policy descriptors**
- **Counterfactual Policy evaluation**
- **Policy design**

Models and Causality

- **Q: Do we need models to get causal effects?**
- **Q: What exactly is a model?**
- **Q: Do we need a structural model? What's that?**

Models: Some Examples

- The simple linear model
$$Y_i = \alpha + \beta T_i + \varepsilon_i$$
- What assumptions do we need here to complete this model and interpret β as a casual effect?
- What other assumptions would you make? Why?
- Let's talk about some other "models"

Heterogeneity

- **Q: What exactly is heterogeneity?**
  - Observed and Unobserved heterogeneity
  - Fixed and Random Effects
  - Random Coefficients

Parameteric Models

- Consider the standard difference in means estimator
- One can equivalently write this as

$$Y_i^{\mathrm{obs}} \mid \mathbf{W}, \tilde{\theta} \sim \mathcal{N}\left(\mu_{\mathrm{c}} + W_i \cdot \tau, \sigma^2\right)$$

- In this case (as in the difference in means) the **ATE** is simple τ

Parameteric Models: Extended Example

- Now consider the following:

$$\begin{pmatrix} \ln\left(Y_i(0)\right) \\ \ln\left(Y_i(1)\right) \end{pmatrix} \mid \theta \sim \mathcal{N}\left(\begin{pmatrix} \mu_{\mathrm{c}} \\ \mu_{\mathrm{t}} \end{pmatrix}, \begin{pmatrix} \sigma_{\mathrm{c}}^2 & 0 \\ 0 & \sigma_{\mathrm{t}}^2 \end{pmatrix}\right)$$

- What is the ATE here?

Parameteric Models: Extended Example

- Now consider the following:

$$\begin{pmatrix} \ln(Y_i(0)) \\ \ln(Y_i(1)) \end{pmatrix} \mid \theta \sim \mathcal{N}\left(\begin{pmatrix} \mu_{\mathrm{c}} \\ \mu_{\mathrm{t}} \end{pmatrix}, \begin{pmatrix} \sigma_{\mathrm{c}}^2 & 0 \\ 0 & \sigma_{\mathrm{t}}^2 \end{pmatrix}\right)$$

- The ATE is:

$$\tau = \tau(\theta) = \exp\left(\mu_{\mathrm{t}} + \frac{1}{2} \cdot \sigma_{\mathrm{t}}^2\right) - \exp\left(\mu_{\mathrm{c}} + \frac{1}{2} \cdot \sigma_{\mathrm{c}}^2\right)$$

Parameteric Models: Inference

- Define $\theta = \{\mu_c, \mu_t, \sigma_c, \sigma_t\}$
- Set up likelihood $\ell\left(\mathbb{D}|\theta\right)$ and obtain $\widehat{\theta}$
- Compute the Hessian at $\widehat{\theta}$: $\widehat{\mathbf{H}} = \left\{ \frac{\partial^2 \ell}{\partial \theta_{jk}} \right\}$
- Use Delta method

$$\text{se}\left(\tau\left(\widehat{\theta}\right)\right) = \left[\frac{\partial \tau\left(\theta\right)}{\partial \theta}\vert_{\theta=\widehat{\theta}}\right]' \widehat{\mathbf{H}}^{-\mathbf{1}} \left[\frac{\partial \tau\left(\theta\right)}{\partial \theta}\vert_{\theta=\widehat{\theta}}\right]$$

Discussion Problem I

- Imagine a firm that sends customers catalogs
- They randomize the treatment so that 90% get the catalog and 10% are held out
- A firm wishes to figure out the causal effect of the catalog (x) on buying behavior (y)
- **What should the model be?**
- **What decisions do we make?**

Discussion Problem II

- Imagine a firm that charges a price for a product. Consumer buys a single unit or not at all.
- A firm wishes to set optimal prices. To do so they need to figure out the causal effect of prices (x) on purchase decision (y)
- **What should the model be?**
- **How should the firm set prices?**

Discussion Problem III

- Detailing refers to the act of pharma reps calling on physicians
- A firm wishes to figure out the causal effect of detailing (x) on prescribing behavior (y)
- Currently the avearge number of calls is 10.
- **What should the model be?**
- **Should the firm increase the number of calls?**