Causal Machine Learning – Fall 2023

Week 6: Deep Nets & Two Step Semiparametrics

Max H. Farrell & Sanjog Misra

Topics to cover

1. Derive the influence function for the ATE
2. Two step semiparametric inference
   - Solving the problem from last time using influence functions and sample splitting
3. More general influence functions

Semiparametric Two Step Estimation & Inference

Remember in week 3 we did **parametric** two step estimation:

$$\widehat{\mathbb{E}[Y(1)]} = \frac{1}{n}\sum_{i=1}^{n} x_i'\hat{\beta}_1$$

The big conclusion was to show that the first stage *estimation* had an impact on the second-stage *inference*.

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\mathbb{E}[Y(1)]} - \mathbb{E}[Y(1)]\right) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\underbrace{x_i'\beta_1 - \mathbb{E}[X\beta_1]}_{\text{Plug in part}} + \underbrace{\mathbb{E}[X']\mathbb{E}[TXX']^{-1}t_i x_i \varepsilon_i}_{\text{First step correction}}\Big\} + o_p(1) \\
&\quad \to_d \mathcal{N}\big(0, \mathbb{V}[\phi(z_i)]\big)
\end{aligned}$$

Semiparametric Two Step Estimation & Inference

Last week we did the same thing, but with ML/nonparametrics in the first stage:

$$\widehat{\mathbb{E}[Y(1)]} = \frac{1}{n}\sum_{i=1}^{n}\hat{\mu}_1(x_i)$$

The big conclusion was that first stage *estimation* had an such a big impact on the second stage that we couldn't get inference

$$\sqrt{n}\left(\widehat{\mathbb{E}[Y(1)]} - \mathbb{E}[Y(1)]\right) \to \infty$$

So we need a two-step estimator that is less sensitive to the first stage.

1. Influence function gives us exactly this
2. Sample splitting helps even more

Semiparametric Two Step Estimation & Inference

Prove that the influence function for $\mathbb{E}[Y(1)]$ is

$$\psi(z_i) = \mu_1(x_i) - \mathbb{E}[Y(1)] + \frac{t_i(y_i - \mu_1(x_i))}{p(x_i)}$$

Remember that we got influence functions by taking the derivative of the parameter with respect to a perturbation in the DGP.

$$\mu(F) = \mathbb{E}[Y(1)] = \mathbb{E}\big[\mathbb{E}[Y \mid T = 1, X]\big] = \mathbb{E}\big[\theta(X)\big],\ \theta(X) = \mathbb{E}[Y \mid T = 1, X].$$

So we need to find

$$\frac{d}{d\varepsilon}\mu(F_\varepsilon)\bigg|_{\varepsilon=0} = \frac{d}{d\varepsilon}\mathbb{E}_\varepsilon\big[\mathbb{E}_\varepsilon[Y \mid T = 1, X]\big]\bigg|_{\varepsilon=0} = \frac{d}{d\varepsilon}\int \theta_\varepsilon(x) f_\varepsilon(x) dx\bigg|_{\varepsilon=0}$$

Semiparametric Two Step Estimation & Inference

Instead of

$$\widehat{\mathbb{E}[Y(1)]} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_1(x_i)$$

We will use

$$\widehat{\mathbb{E}[Y(1)]} = \frac{1}{n}\sum_{i=1}^{n} \hat{\mu}_1(x_i) + \frac{t_i(y_i - \hat{\mu}_1(x_i))}{\hat{p}(x_i)}$$

for nonparametric $\hat{\mu}_1(x_i)$ and $\hat{p}(x_i)$.
Show that this works...

Semiparametric Two Step Estimation & Inference

More generally:

$$\mu(F) = \mu = \mathbb{E}[H(X, \theta(X))], \qquad \theta_0(x) = \arg\max \sum_{i=1}^{n} \ell(y_i, t_i, \theta(x))$$

Has influence function

$$H(x_i, \theta(x_i)) - \mu + (\nabla_\theta H)\, \mathbb{E}[\ell_{\theta\theta} \mid x_i]^{-1} \ell_\theta(y_i, t_i, \theta(x_i))$$