Identification Analysis

Kosuke Imai

Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2018

Difficulties of Observational Studies

- Observational studies $\rightsquigarrow$ No randomization of treatment assignment

$$\{Y_i(1), Y_i(0)\} \not\perp\!\!\!\perp T_i$$

- Treatment assignment mechanism is often unknown
- Possible existence of observed and unobserved confounders

- Credible causal inference in observational studies
- Key questions:
  1. What is your identification assumption?
  2. What is your identification strategy?

Identification vs. Statistical Inference

- Identification: How much can you learn about the estimand if you had an infinite amount of data?
- Statistical Inference: How much can you learn about the estimand from a finite sample?
- Identification precedes statistical inference
- Key questions for identification analysis:
  1. What can be learned without making any assumption other than the ones which we know are satisfied by the design?
  2. What is a minimum set of assumptions required for the point identification of an estimand?
  3. Can we characterize the identification region if we relax some or all of these assumptions?
- Law of Decreasing Credibility: The credibility of inference decreases with the strength of the assumptions maintained

  (Manski. 2007. *Identification for Prediction and Decision*. Harvard UP)

Identification of the Average Treatment Effect

- Identification assumptions:
  1. Overlap (i.e., no extrapolation):

$$0 < \Pr(T_i = 1 \mid \mathbf{X}_i = \mathbf{x}) < 1 \text{ for any } \mathbf{x}$$

  2. Ignorability (exogeneity, unconfoundedness, no omitted variable, selection on observables, etc.)

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i \mid \mathbf{X}_i = \mathbf{x} \text{ for any } \mathbf{x}$$

- Under these assumptions:

$$\tau \; = \; \mathbb{E}(Y_i(1) - Y_i(0)) \; = \; \mathbb{E}\{\mu(1, \mathbf{X}_i) - \mu(0, \mathbf{X}_i)\}$$

- Regression-based Estimator:

$$\hat{\tau}_{\text{reg}} \; = \; \frac{1}{n} \sum_{i=1}^{n} \{\hat{\mu}(1, \mathbf{X}_i) - \hat{\mu}(0, \mathbf{X}_i)\}$$

Regression-based Estimation

- Example: Logistic regression

$$\mu(t, \mathbf{x}) \;=\; \frac{\exp(\alpha + \beta \cdot t + \mathbf{x}^\top \gamma)}{1 + \exp(\alpha + \beta \cdot t + \mathbf{x}^\top \gamma)}$$

- The Estimator:

$$\hat{\tau}_{\text{reg}} \;=\; \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\exp(\hat{\alpha} + \hat{\beta} + \mathbf{X}_i^\top \hat{\gamma})}{1 + \exp(\hat{\alpha} + \hat{\beta} + \mathbf{X}_i^\top \hat{\gamma})} - \frac{\exp(\hat{\alpha} + \mathbf{X}_i^\top \hat{\gamma})}{1 + \exp(\hat{\alpha} + \mathbf{X}_i^\top \hat{\gamma})} \right\}$$

- Variance for the Conditional Average Treatment Effect:

$$\mathbb{V}(\hat{\tau}_{\text{reg}} \mid \mathbf{X}) \;=\; \frac{1}{n^2} \mathbb{V}\left(\sum_{i=1}^{n} \hat{\mu}(1, \mathbf{X}_i) - \hat{\mu}(0, \mathbf{X}_i) \mid \mathbf{X} \right)$$

Asymptotic Variance Calculation

- Delta method:

$$\frac{1}{n^2}\left\{\sum_{i=1}^{n}\mathbb{V}(\hat{\mu}(1,\mathbf{X}_i)\mid\mathbf{X})+\mathbb{V}(\hat{\mu}(0,\mathbf{X}_i)\mid\mathbf{X})-2\mathrm{Cov}(\hat{\mu}(1,\mathbf{X}_i),\hat{\mu}(0,\mathbf{X}_i)\mid\mathbf{X})\right.$$
$$\left.+\sum_{i=1}^{n}\sum_{i'\neq i}\mathrm{Cov}(\hat{\mu}(1,\mathbf{X}_i)-\hat{\mu}(0,\mathbf{X}_i),\hat{\mu}(1,\mathbf{X}_{i'})-\hat{\mu}(0,\mathbf{X}_{i'})\mid\mathbf{X})\right\}$$

- Bootstrap (Unconditional inference for the ATE):
  1. Independently sample n observations with replacement
  2. Fit the logistic regression and compute $\hat{\tau}_{\text{reg}}$
  3. Repeat

- Quasi-Bayesian Monte Carlo (Zelig; King et al. 2000. *Amer. J. Political Sci*):
  1. Sample (α,β,γ) from $\mathcal{N}((\hat{\alpha},\hat{\beta},\hat{\gamma}),\mathbb{V}(\widehat{(\hat{\alpha},\hat{\beta},\hat{\gamma})}))$
  2. Compute τ_{reg}
  3. Repeat

Analysis of Bounds

- What can we learn about the ATE if we make no assumption?
- No-assumption bounds as the starting point of analysis:

$$[-\Pr(Y_i = 0 \mid T_i = 1)\Pr(T_i = 1) - \Pr(Y_i = 1 \mid T_i = 0)\Pr(T_i = 0), \\ \Pr(Y_i = 1 \mid T_i = 1)\Pr(T_i = 1) + \Pr(Y_i = 0 \mid T_i = 0)\Pr(T_i = 0)\}]$$

- The width of the bounds is 1: "A glass is half empty/full"

- Monotone treatment selection (Manski and Pepper. 2000. *Econometrica*):

$$[\mathbb{E}(Y_i \mid T_i = 1)\Pr(T_i = 1) + \ell \cdot \Pr(T_i = 0) - \mathbb{E}(Y_i \mid X_i), \\ \mathbb{E}(Y_i) - \mathbb{E}(Y_i \mid T_i = 0)\Pr(T_i = 0) - \ell \cdot \Pr(T_i = 1)].$$

where $\ell \leq Y_i$

Perry Preschool Project

- Randomized evaluation of HighScope early childhood curriculum
- A couple of hundred at-risk children in Ypsilanti (Michigan)
- Follow-ups at older ages
- High school graduation by age 19: 67% (treated) vs. 49% (control)

- What would the graduation rate be if only some children receive the intervention? (Manski. 1997. *Rev. Econ. Stud.*)
  - no-assumption bounds: $[16\%, 100\%]$
  - intervention can never hurt: $[49\%, 67\%]$
  - independent potential outcomes: $[33\%, 83\%]$
  - 10% receives the intervention: $[39\%, 59\%]$
  - 90% receives the intervention: $[57\%, 77\%]$

Statistical Inference for Bounds

1. Confidence intervals for the true bounds:

$$\left[\hat{\tau}_{\text{lower}} - z_{1-\alpha/2} \cdot \hat{\sigma}_{\text{lower}}, \ \hat{\tau}_{\text{upper}} + z_{1-\alpha/2} \cdot \hat{\sigma}_{\text{upper}}\right]$$

   where $\sigma^2_{\text{lower}} = \mathbb{V}(\hat{\tau}_{\text{lower}})$ and $\sigma^2_{\text{upper}} = \mathbb{V}(\hat{\tau}_{\text{upper}})$

2. Confidence intervals for the true value:

$$[\hat{\tau}_{\text{lower}} - z_{1-\alpha} \cdot \hat{\sigma}_{\text{lower}}, \ \hat{\tau}_{\text{upper}} + z_{1-\alpha} \cdot \hat{\sigma}_{\text{upper}}]$$

   - If $\tau = \tau_{\text{lower}}$ or $\tau = \tau_{\text{upper}}$, the coverage prob. converges to $1 - \alpha$
   - If $\tau_{\text{lower}} < \tau < \tau_{\text{upper}}$, the coverage prob. converges to 1

More Efficient Interval (Imbens and Manski. 2004. *Econometrica*)

- Consider the estimation of $\theta = \mathbb{E}(Y_i(1))$ with $0 \leq Y_i(1) \leq 1$
- No assumption bounds:

$$[\theta_{\text{lower}}, \theta_{\text{upper}}] \;=\; [\mu_1 \pi,\ \mu_1 \pi + (1-\pi)] \text{ where } \mu_1 = E(Y_i \mid T_i = 1)$$

- Symmetric confidence interval:

$$[\hat{\theta}_{\text{lower}} - D,\ \hat{\theta}_{\text{upper}} + D]$$

where we choose D such that

$$\begin{aligned}
& \Pr(\hat{\theta}_{\text{lower}} - D \leq \theta \leq \hat{\theta}_{\text{upper}} + D \mid \hat{\pi}) \\
= \;& 1 - \Pr(\theta < \hat{\theta}_{\text{lower}} - D \mid \hat{\pi}) - \Pr(\theta > \hat{\theta}_{\text{upper}} + D \mid \hat{\pi}) \\
\geq \;& \Phi\left(\sqrt{n\hat{\pi}} \cdot \frac{D + 1 - \pi}{\sigma_1 \pi}\right) - \Phi\left(-\sqrt{n\hat{\pi}} \cdot \frac{D}{\sigma_1 \pi}\right) \;=\; 1 - \alpha
\end{aligned}$$

Concluding Remarks

- Distinction between associational and causal relationships
- Do not just report coefficients: e.g., calculate ATE/ATT

- Causal inference in observational studies requires additional assumptions
  1. Overlap
  2. Ignorability
- These assumptions may be difficult to defend

- Identification analysis $\rightsquigarrow$ partial identification
  1. What can data alone tell us about causal effects?
  2. How can some identification assumptions narrow bounds?
- Credibility of untestable assumptions