CS295 Causal Reasoning Paper Presentation

[ACM] Detecting Latent Heterogeneity (Pearl, 2015)

Hao-Che (Howard) Hsu

Department of Economics

University of California-Irvine

2021

Main References

Check for updates

Inference and Causality

8

Estimating Heterogeneous Treatment Effects with Observational Data

Sociological Methodology
Volume 42, 314–347
© American Sociological Association 2012
DOI: 10.1177/0081175012452652
http://sm.sagepub.com
SAGE

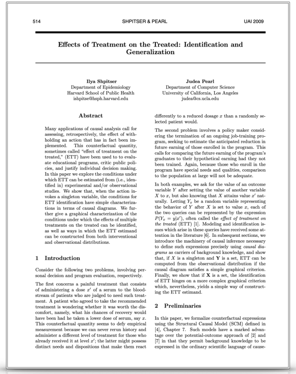
Yu Xie[1], Jennie E. Brand[2], and Ben Jann[3]

Abstract

Individuals differ not only in their background characteristics but also in how they respond to a particular treatment, intervention, or stimulation. In particular, treatment effects may vary systematically by the propensity for treatment. In this paper, we discuss a practical approach to studying heterogeneous treatment effects as a function of the treatment propensity, under the same assumption commonly underlying regression analysis: ignorability. We describe one parametric method and two nonparametric methods for estimating interactions between treatment and the propensity for treatment. For the first method, we begin by estimating propensity scores for the probability of treatment given a set of observed covariates for each unit and construct balanced propensity score strata; we then estimate propensity score stratum-specific average treatment effects and evaluate a trend across them. For the second method, we match control units to treated units based on the propensity score and transform the data into treatment-control comparisons at the most elementary level at which such comparisons can be constructed; we then estimate treatment effects as a function of the propensity score by fitting a nonparametric model as a smoothing device. For the third method, we first estimate nonparametric regressions of the outcome variable as a function of the propensity score separately for treated units and for control units and then take the difference between the two nonparametric regressions. We illustrate the application of these methods with an empirical example of the effects of college attendance on women's fertility.

[1]University of Michigan
[2]University of California–Los Angeles
[3]University of Bern, Switzerland

Corresponding Author:
Yu Xie, Institute for Social Research, University of Michigan, 426 Thompson Street,
Ann Arbor, MI 48104, USA
Email: yuxie@umich.edu



514 SHPITSER & PEARL UAI 2009

Effects of Treatment on the Treated: Identification and Generalization

Ilya Shpitser
Department of Epidemiology
Harvard School of Public Health
ishpitse@hsph.harvard.edu

Judea Pearl
Department of Computer Science
University of California, Los Angeles
judea@cs.ucla.edu

Abstract

Many applications of causal analysis call for assessing, retrospectively, the effect of withholding an action that has in fact been implemented. This counterfactual quantity, sometimes called "effect of treatment on the treated," (ETT) have been used to to evaluate educational programs, critic public policies, and justify individual decision making. In this paper we explore the conditions under which ETT can be estimated from (i.e., identified in) experimental and/or observational studies. We show that, when the action invokes a singleton variable, the conditions for ETT identification have simple characterizations in terms of causal diagrams. We further give a graphical characterization of the conditions under which the effects of multiple treatments on the treated can be identified, as well as ways in which the ETT estimand can be constructed from both interventional and observational distributions.

1 Introduction

Consider the following two problems, involving personal decision and program evaluation, respectively.

The first concerns a painful treatment that consists of administering a dose x' of a serum to the bloodstream of patients who are judged to need such treatment. A patient who agreed to take the recommended treatment is wondering whether it was worth the discomfort, namely, what his chances of recovery would have been had he taken a lower dose of serum, say x. This counterfactual quantity seems to defy empirical measurement because we can never rerun history and administer a different level of treatment for those who already received it at level x'; the latter might possess distinct needs and dispositions that make them react differently to a reduced dosage x than a randomly selected patient would.

The second problem involves a policy maker considering the termination of an ongoing job-training program, seeking to estimate the anticipated reduction in future earning of those enrolled in the program. This calls for comparing the future earning of the program's graduates to their hypothetical earning had they not been trained. Again, because those who enroll in the program have special needs and qualities, comparison to the population at large will not be adequate.

In both examples, we ask for the value of an outcome variable Y after setting the value of another variable X to x, but also knowing that X attains value x' naturally. Letting Y_x be a random variable representing the behavior of Y after X is set to value x, each of the two queries can be represented by the expression $P(Y_x = y|x')$, often called the *effect of treatment on the treated* (ETT) [1]. Modeling and identification issues which arise in these queries have received some attention in the literature [6]. In subsequent sections, we introduce the machinery of causal inference necessary to define such expressions precisely using *causal diagrams* as carriers of background knowledge, and show that, if X is a singleton and $\mathbf{Y}$ is a set, ETT can be computed from the observational distribution if the causal diagram satisfies a simple graphical criterion. Finally, we show that if $\mathbf{X}$ is a set, the identification of ETT hinges on a more complex graphical criterion which, nevertheless, yields a simple way of constructing the ETT estimand.

2 Preliminaries

In this paper, we formalize counterfactual expressions using the Structural Causal Model (SCM) defined in [4], Chapter 7. Such models have a marked advantage over the potential-outcome approach of [2] and [7] in that they permit background knowledge to be expressed in the ordinary scientific language of cause-

1. Estimating Heterogeneous Treatment Effects with Observational Data (Yu Xie, Jennie E. Brand, Ben Jann, 2012)
2. Effects of Treatment on the Treated: Identification and Generalization (Ilya Shpitser and Judea Pearl, 2009)

Heterogeneity

- Treatment might affect different experimental subjects in different ways.
  - Individuals respond differently to treatment or intervention.
  - Idiosyncratic groups in the population.
- Why do we care?
  - vaccine is uniformly beneficial
  - program evaluations
  - bias causal estiamtes

Heterogeneity

How do we detect heterogeneity?

Program evaluation

- job training program
- randomized experiment: training $\rightarrow$ get a job
- Study (a year later): hiring rate among the trained is even higher
- eligible and enrolled: smarter, more resourceful, more socially connected
  - would have found a job regardless of training
- population is not homogeneous
- informed$\rightarrow$ enroll (little benefit) / uninformed(weak)$\rightarrow$ not aggressively recruited

Assess the Degree of Heterogeneity

1. Covariate-specific methods
2. Compare the treated and the untreated

Covariate-induced Heterogeneity

- Does having this characteristic respond differently from those not having it?
- The unbiased estimator for the counterfactuals: $\frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$
- Comparing the covariates (C):

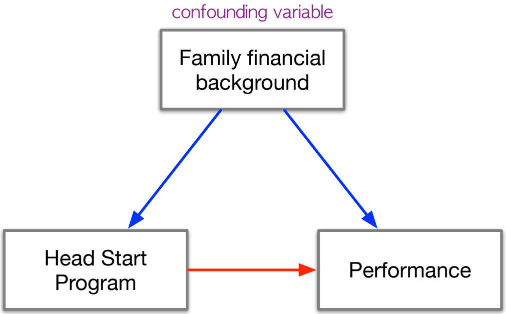
$$D(c_i, c_j) = \text{abs}\Big[E(Y_1 - Y_0 | C = c_i) - E(Y_1 - Y_0 | C = c_j)\Big]$$

  - $E(Y_1 - Y_0 | C = c_0) = E(Y | X = 1, C = c_0) - E(Y | X = 0, C = c_0)$
- Heterogeneity lowerbound: $\text{LB}_{\text{heterogeneity}} = \max_{\{c_i, c_j\} \in C} D(c_i, c_j)$
- Does set C satisfy the backdoor criterion?

Unconfoundedness

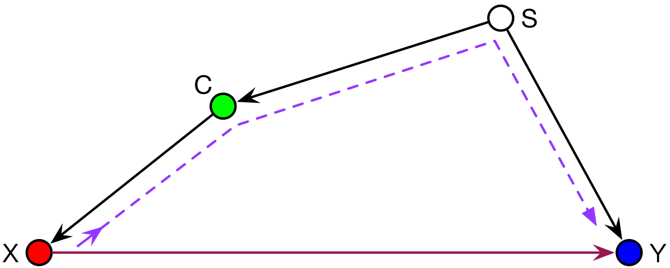
Conditional independence assumption

- Also known as
  - Statistics: ignorability
  - Parametric: selection on observables (Heckman and Robb, 1984)
  - Missing data: missing at random
  - Causal DAG: backdoor criterion
- $(Y_1, Y_0) \perp X | C$
- $Y = \beta_0 + \alpha X + C\beta + \epsilon$, where $X \perp \epsilon | C$ (X is exogenous)
- Children from poor families are selected into the program $\Longrightarrow$ not able to compete



Identify c-specific effect (of X on Y)

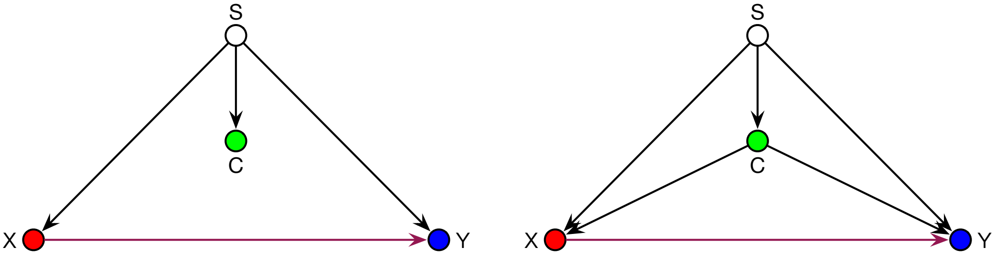
C is admissible



- Identification: $E(Y_1 - Y_0|C = c) = E(Y|X = 1, C = c) - E(Y|X = 0, C = c)$

Identify c-specific effect

$(C \cup S)$ is admissible

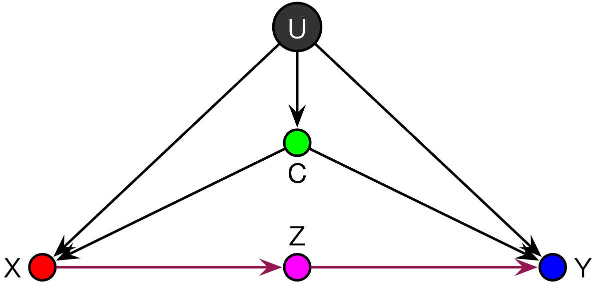


- Adjust S:

$$E(Y_1 - Y_0 | C = c)$$
$$= \sum_s \Big[E(Y | X = 1, S = s, C = c) - E(Y | X = 0, S = s, C = c) \Big] \cdot P(s|c)$$

Identify c-specific effect

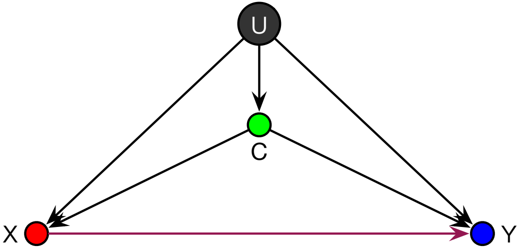
C is not admissible (identifiable)

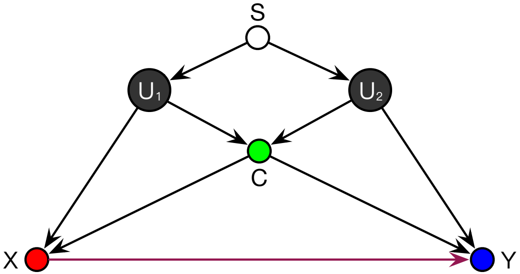


- Frontdoor adjustment

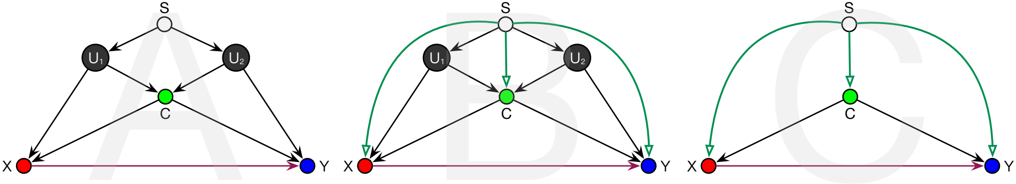
Identify c-specific effect

C is not admissible (not identifiable)

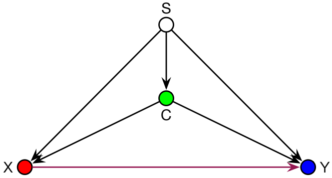




Statistically Indistinguishable



not identifiable



identifiable

Latent Heterogeneity

- What if when heterogeneity is not presented in any covariates?
- If there is heterogeneity, would the causal effect on the treated group vs the control/untreated group the same?
- Noticed that in RCT:
  - The treated is just as bad as the control if the were not assigned treatment.
  - If the control were given treatment they will be as good as the treated.
- There must be bias in the ATE.

ATE Decomposition

$p(X=1)$: **proportion treated,** $p(X=0)$: **proportion untreated (control)**

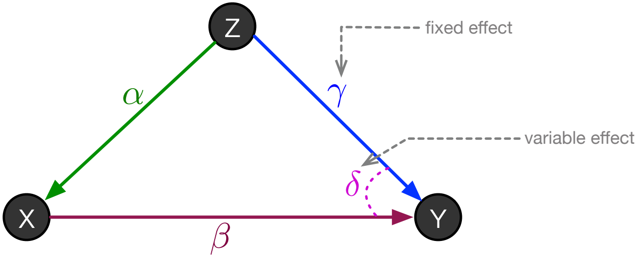
$$
\begin{aligned}
\text{ATE} &= E(Y_1 - Y_0) \\
&= E(Y_1|X=1)p(X=1) + E(Y_1|X=0)[1-p(X=1)] - E(Y_0|X=0)p(X=0) - E(Y_0|X=1)[1-p(X=0)] \\
&= \Big[\underbrace{E(Y_1|X=1) - E(Y_0|X=0)}_{\text{ATE from observed RCT data}}\Big] \\
&\quad \left.\begin{aligned}
&- \Big[\underbrace{E(Y_0|X=1) - E(Y_0|X=0)}_{\text{pretreatment heterogeneity bias (Type-I bias, baseline bias, selection bias)}}\Big] \\
&- \underbrace{\Big\{\Big[\underbrace{E(Y_1|X=1) - E(Y_0|X=1)}_{\text{ATT}}\Big] - \Big[\underbrace{E(Y_1|X=0) - E(Y_0|X=0)}_{\text{ATU}}\Big]\Big\}}_{\text{Unweighted treatment-effect heterogeneity bias (Type-II bias, variable-effect bias)}} \cdot \underbrace{p(X=0)}_{\text{proportion untreated}}
\end{aligned}\right\} \text{bias}
\end{aligned}
$$

- There is no C (covariates) involved in the decomposition.

The Bias

- **Pretreatment heterogeneity bias (selection bias)**
  - Output difference of two groups if neither receives treatment
  - If the control were given treatment they will be as good as the treated
  - Source: fixed effect and covariates
  - Fixed by controlling covariates
  - e.g. Head Start program select poor family children
- **Treatment-effect heterogeneity bias (ATT − ATU)·p(X=0)**
  - Difference in the average treatment effect of two groups
  - If there's no heterogeneity, this term vanishes
  - Source: variable effect
  - Can't be controlled for by covariates
  - e.g. attending college is selective

Fixed Effect & Variable Effect



- Model structural equations:

$$y = \beta x + \gamma z + \delta xz + \epsilon_y$$

$$x = \alpha z + \epsilon_x$$

$$z = \epsilon_z$$

Detecting Latent Heterogeneity

Goal: Identify **ATT** and **ATU**.

Three ways of detection

1. Randomized trails with binary treatments
2. Covariate adjustment
3. Instrumental variables

1. Randomized Trails with Binary Treatments

Before randomization...

- Pretreatment heterogeneity bias (selection bias)
  - $$E(Y_0|X=1) - E(Y_0|X=0) = \frac{[E(Y_0) - E(Y_0|X=0)]}{p(X=1)}$$
- **Unweighted Treatment-effect heterogeneity bias**
  - ATT − ATU
    $$= E(Y_1|X=1) - \frac{[E(Y_0) - E(Y_1|X=0)[1-p(X=1)]]}{p} - \frac{[E(Y_1) - E(Y_1|X=1)p(X=1)]}{1-p} - E(Y_0|X=0)$$
    $$= \frac{[E(Y_1|X=1) - E(Y_1)]}{1-p} + \frac{[E(Y_0|X=0) - E(Y_0)]}{p}$$
  - not population heterogeneity, but the het. that has preferential selection to treatment
  - None zero?
- Estimate $E(Y_0) = E(Y|X=0)$ and $E(Y_1) = E(Y|X=1)$ empirically in RCT.

2. Detecting Heterogeneity Through Adjustments

If backdoor criterion holds for some set Z

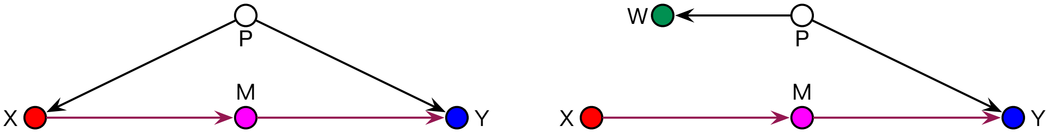
Adjustment formula and variation:

- $E(Y_0) = \sum_z E(Y|X=0,z) \cdot p(z)$
- $E(Y_0|X=1) = \sum_z E(Y|X=0,z) \cdot p(z|X=1)$

2. Detecting Heterogeneity Through Adjustments

Theorem

$p(Y_x = y|X = x')$ is identifiable in G iff $p(y|w, do(x))$ is identifiable in G' which from G, adds a new node W with the same set of parents as X and no children.



- $p(y|w, do(x)) = \dfrac{p(y, w|do(x))}{p(w)} = \dfrac{\sum_z p(y|z, x)p(w, z)}{p(w)} = \sum_z p(y|z, x)p(z|w)\Big|_{w=x'}$

2. Detecting Heterogeneity Through Adjustments

Z of covariates is an admissible set

- $E(Y_a|X=b) = p(y|b, do(a)) = \sum_z E(Y|X=a,z) \cdot p(z|X=b)$

- ATT-ATU

$$
\begin{aligned}
&= E(Y_1 - Y_0|X=1) - E(Y_1 - Y_0|X=0) \\
&= E(Y_1|X=1) - E(Y_0|X=1) - E(Y_1|X=0) + E(Y_0|X=0) \\
&= \sum_z [E(Y|X=1,z) - E(Y|X=0,z)]\,[p(z|X=1) - p(z|X=0)]
\end{aligned}
$$

3. Detecting Heterogeneity Through Mediating Instruments Z

The frontdoor criterion

- set Z intercept all directed paths from X to Y

- $ATT - ATU = \sum_z [E(Y|X=1,z) - E(Y|X=0,z)]\,[p(z|X=1) - p(z|X=0)]$