

Conditional independence

労働経済学

川田恵介

2025-06-01

1 Selection on observable

1.1 基本アイデア

- 統計的因果推論の基本的方針は、Chance (偶然) を利用した因果効果の推論 (Imbens, 2022)
 - ▶ RCT においては、 D は偶然(ランダム)決まる
- 母集団全体で（自然）実験が発生している応用はまれ
 - ▶ “局所的”に発生した自然実験を活用する

1.2 局所的な実験

- 多くの手法が発展
- 本スライド:、conditional independence を仮定し、因果効果を識別し、慎重に推定する
 - ▶ 推定については、Balanced comparison の手法が活用できる
 - Sensitivity 分析との併用を推奨
 - ▶ 代替的な識別方法としては、操作変数、Regression Discontinuity 等

2 識別

2.1 推定対象

- Conditional Average Treatment Effect
- X 内での D の平均効果

$$\tau(x) = E[y_i(1) - y_i(0) \mid X_i = x]$$

2.2 潜在結果モデル

$$\begin{aligned} & \bullet E[Y_i | D_i = 1, x] - E[Y_i | D_i = 0, x] \\ & \bullet = E[y_i(1) | D_i = 1, x] - E[y_i(0) | D_i = 0, x] \\ & \bullet = \underbrace{E[y_i(1) | D_i = 0, x] - E[y_i(0) | D_i = 0, x]}_{=E[\tau_i | D_i=0, x]} \\ & \quad + \underbrace{E[y_i(1) | D_i = 1, x] - E[y_i(1) | D_i = 0, x]}_{=Selection} \end{aligned}$$

2.3 識別の仮定

- Conditional independence:

$$E[y_i(1) | D_i = 1, x] = E[y_i(1) | D_i = 0, x]$$

- ▶ X が同じであれば、 D はランダムに決まっている

2.4 識別の仮定

- Conditional independence に加え、
 - ▶ すべての x, d について、 $0 < \Pr[d | X = x] < 1$ (Positivity)
 - ▶ 他者の d に影響を受けない (No interference)

2.5 違反例

- $D =$ 「労働経済学」の講義への参加 $\rightarrow Y =$ 30 歳時点での所得
 - ▶ 受講できない研究科が存在: Positivity 違反
 - ▶ 勉強会などで非受講者にも講義内容を共有: No interference 違反
 - ▶ そもそも興味関心など、データから観察しにくい要因に、受講するかどうか依存: Conditional independence 違反

3 推定

3.1 課題

- 因果効果が識別できたとして、限られた事例数から、どのように推定するか?
 - ▶ X のバランスを達成する必要がある
- Balanced comparison (Slide07)の方法が使用可能

3.2 サブサンプル内での平均差

- Conditional average treatment effect $\tau(X)$ が識別できたとしても、推定は容易ではない

- 最もシンプルな推定方法は、 X 内での Y の D 間での平均差

3.3 Subsample size 問題

- X の数が多いと、サブサンプルサイズが小さくなり、推定精度が悪化する
- 条件付き平均効果の”集計値”を推定する必要がある
 - ▶ どのように集計するか、という問題が発生する

3.4 Average treatment effect

- $\tau(X)$ の平均値を推定対象とする

$$\tau = \sum_{X=x} \omega(x) \times \tau(x)$$

- $\omega(x) = \text{Weight}$

3.5 τ の推定

- 識別

$$\tau(x) = E[Y \mid D = 1, x] - E[Y \mid D = 0, x]$$

より

$$\tau = \sum_{X=x} \omega(x) \times [E[Y \mid D = 1, x] - E[Y \mid D = 0, x]]$$

- バランス後の比較!!!
 - ▶ 前回までの議論が活用できる

3.6 まとめ

- コントロール変数で因果効果を識別するのであれば、 X 内で RCT が行われている必要がある
- 因果効果の異質性を想定する場合、OLS による推定は、平均効果を推定できない
 - ▶ X をサンプル平均とバランスさせる方法を併用する必要がある

4 X の選択

4.1 変数選択

- データに含まれる変数から、研究課題に必要な変数を選ぶ必要がある
- 研究課題から、 Y/D は比較的容易に決まる
 - ▶ X の選択はより困難
 - ▶ 様々な補助ツール(概念)が存在する

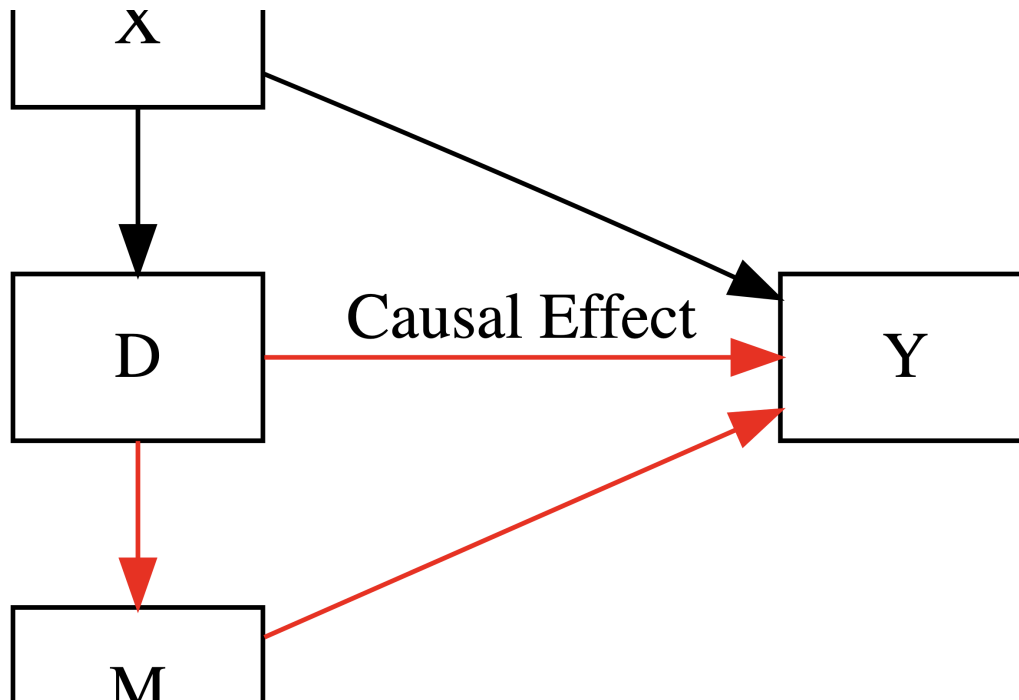
4.2 Confounders

- Conditional independence について、より踏み込んだ議論のために、概念 Confounders (交絡因子)を導入
 - ▶ D と Y に影響を与える変数
 - D についての理想的な RCT で差が生じない変数
- 例: 経済学研究科の院生の方が参加しやすい
 - ▶ 経済学研究科と他研究科の間で 30 歳時点での所得にも差異がある

4.3 注意点: Bad Control

- データの中には、バランスすべきではない変数も通常含まれている
 - ▶ 理想的な RCT においても、 D 間で差異が生まれる変数 (Post-treatment M)
- 例: M = 修士論文の内容
 - ▶ RCT においては、実験前に収集した変数は Bad Control ではない
 - ▶ 非 RCT データでは、背景情報を用いて判断するしかない

4.4 イメージ



4.5 まとめ

- より発展的な議論は、VanderWeele (2019)などを参照
- 問題点: すべての Confounders をデータから観察できるとは限らない

- ▶ 観察できない(Unobservable) confounders が存在
 - X 内の比較は、因果効果と一致しない
- Conditional independence は、Unobservable confounders が存在しないことを仮定
 - ▶ しばしば強すぎる仮定であり、代替案が提案されている

5 応用例

5.1 Anger et al. (2017)

- 研究目標: 非自発的失業 (D)が、労働者の性格 (*BigFiveY*) に与える因果効果
 - ▶ 性格特性が人生の出来事によって変わる (心理学) VS 石膏のように固まる (経済学)
 - 私見: 本当?

5.2 識別

- 推定対象: 初期時点において失業していない労働者: $D = 1$ 事業所閉鎖により失職, $D = 0$ 就業継続
 - ▶ 自発的失職を除外することで、背景属性をある程度の”バランス”している
 - ▶ X = 初期時点での社会経済変数/就業状態/企業属性
- 仮定: X 内で、事業所が閉鎖されるかどうかは、(労働者にとっては)ランダムに決まっている

5.3 推定

- 重回帰 + Entropy Weighting
- 推定値: Big Five の大部分は、失職について安定的
 - ▶ 開放性は、失職により、上昇する

5.4 Reference

Bibliography

- Anger, S., Camehl, G., & Peter, F. (2017). Involuntary job loss and changes in personality traits. *Journal of Economic Psychology*, 60, 71–91.
- Imbens, G. W. (2022). Causality in econometrics: Choice vs chance. *Econometrica*, 90(6), 2541–2566.
- VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology*, 34, 211–219.