AIPW 推定

Semiparametric 推定への応用

川田恵介

Table of contents

AIPW

- Partialling out は、variance weight を用いた estimand を推定
  - しばしば解釈が困難
- D がカテゴリカル (density が推定可能) であれば、frequency weight を使った estimand も推定可能
  - 有力な代替案: AIPW (Robins and Rotnitzky 1995)

Get Start: AIPW

1. $E[Y|D=d,X]=g_{Y(d)}(X), E[D|X]=g_D(X)$ を教師付き学習などを用いて推定
2. S_i^{AIPW} の母平均を推定、ただし

$$S_i^{AIPW}=g_{Y(1)}(X_i)-g_{Y(0)}(X_i)$$

$$+\frac{D_i(Y_i-g_{Y(1)}(X_i))}{g_D(X_i)}-\frac{(1-D_i)(Y_i-g_{Y(0)}(X_i))}{1-g_D(X_i)}$$

Summary

- AIPW と Partialling out では、Estimand が異なることに注意

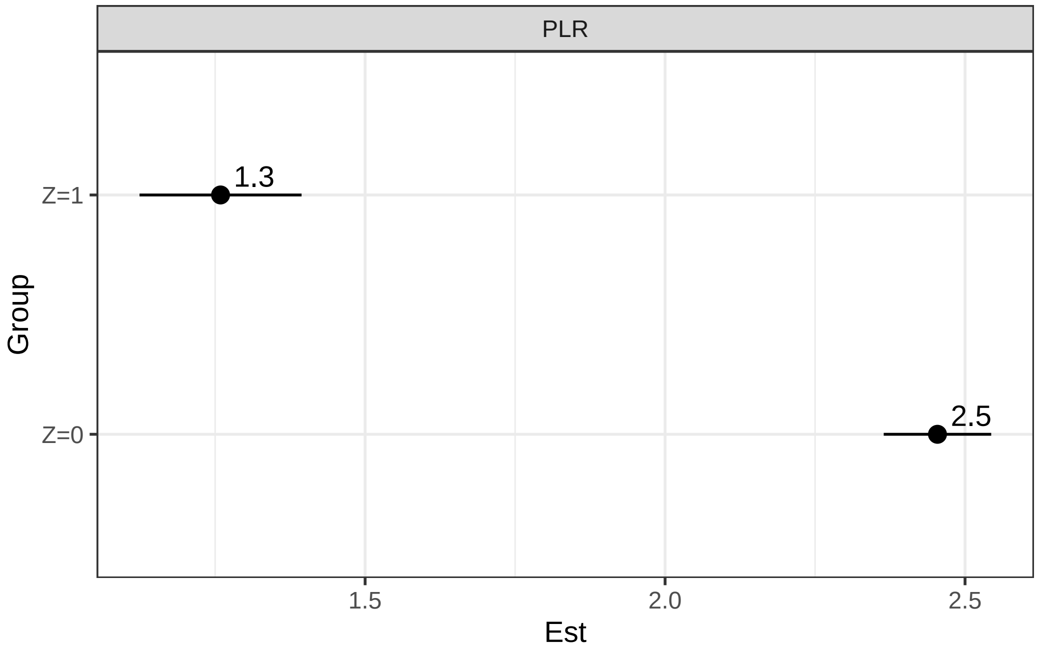
$$\tau=\int\omega_P(x)\times\tau_P(x)dx$$

- AIPW: $\omega_P(x)=f_P(x)$ (frequency weight)
- Partialling out: $\omega_P(x)=f_P(x)\times var_P(D_i|x)$ (variance weight)

frequency weight への動機づけ

- しばしば決定的に異なる: 例えば、Sub group 分析
- $\int E_P[Y|D=1,X] - E_P[Y|D=0,X,Z] \times \omega_P(X)dX$ を、$Z = \{0,1\}$ ごとに比較したい
  - Y : 賃金、D : 高卒/大卒、X : 東京出身 $=1$ /その他 $=0$、Z : 1984 年生まれ $=1$ /1960 年生まれ $=0$
  - 1984 年と 1960 年で、出身地をコントロールした上での、大卒・高卒間賃金格差を比較したい

結果: N = 10000



- $Z = 1$ (1984 年生まれ) の方が学歴間賃金格差が小さい

なぜグループ間で差が異なるのか?

- 可能性 1: X が同じであったとしても、$Z = 1$ の方が平均差が小さい (学歴間賃金格差が減った)
- 可能性 2: 差が小さい X の割合が多い (賃金格差が大きい地域で人口が減った)
- 可能性 3: 差が大きい X について、D の分散が大きい (?)
  - Variance weight を使って集計しているため

可能性 1

```
# A tibble: 4 x 5
  `Tau(X,Z)`     Z     X `f(X,Z)` `E[D|X,Z]`
       <dbl> <dbl> <dbl>    <dbl>      <dbl>
1          5     1     1     0.25        0.5
2          5     1     0     0.25        0.5
3         10     0     1     0.25        0.5
4         10     0     0     0.25        0.5
```

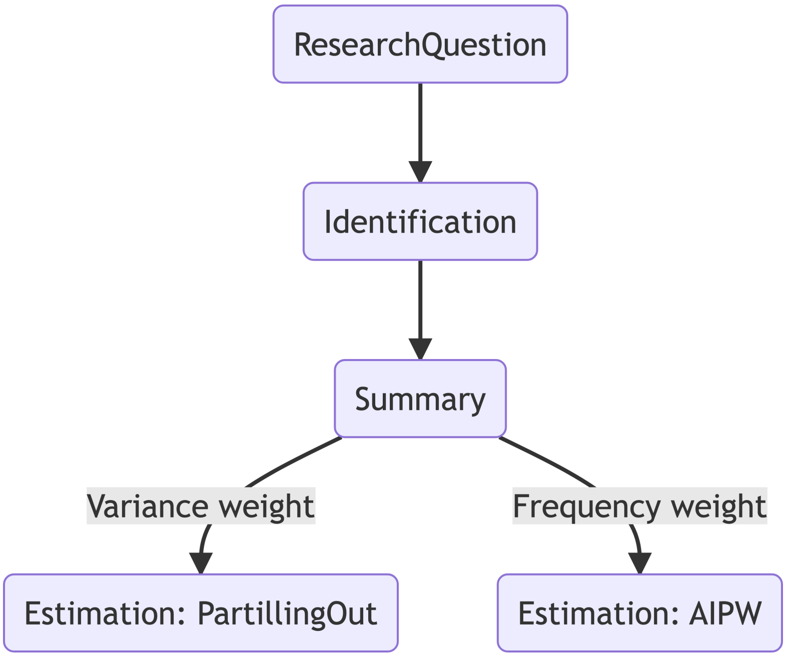
可能性 2

```
# A tibble: 4 x 5
  `Tau(X,Z)`     Z     X `f(X,Z)` `E[D|X,Z]`
       <dbl> <dbl> <dbl>    <dbl>      <dbl>
1         10     1     1     0.15        0.5
2          5     1     0     0.35        0.5
3         10     0     1     0.25        0.5
4          5     0     0     0.25        0.5
```

可能性 3

```
# A tibble: 4 x 5
  `Tau(X,Z)`     Z     X `f(X,Z)` `E[D|X,Z]`
       <dbl> <dbl> <dbl>    <dbl>      <dbl>
1         10     1     1     0.25        0.9
2          5     1     0     0.25        0.5
3         10     0     1     0.25        0.5
4          5     0     0     0.25        0.5
```

整理



Estimation

- frequency weight を用いた平均差推定について、いくつか (無数?) の方法が考えられる
- (本講義における) 非推奨: (Naive) Plugin, **I**nverse **P**ropensity **W**eight (IPW)
- 推定: **A**ugmented **I**nverse **P**robability **W**eighted (AIPW)

Plugin

1. $E[Y|D=d,X]=g_{Y(d)}(X)$ を教師付き学習などを用いて推定
2. S_i^{Plugin} の母平均を推定、ただし

$$S_i^{Plugin}=g_{Y(1)}(X_i)-g_{Y(0)}(X_i)$$

- $g_{Y(d)}$ の推定誤差に敏感: 一般に $\sqrt{N}$ CAN estimator にならない

IPW

1. $E[D|X]=g_D(X)$ を教師付き学習などを用いて推定

2. S_i^{IPW} の母平均を推定、ただし

$$S_i^{IPW} = \frac{D_i Y_i}{g_D(X_i)} - \frac{(1-D_i)Y_i}{1-g_D(X_i)}$$

- g_D の推定誤差に敏感: 一般に $\sqrt{N}$ CAN estimator にならない

AIPW

1. $E[Y|D=d,X] = g_{Y(d)}(X), E[D|X] = g_D(X)$ を教師付き学習などを用いて推定

2. S_i^{AIPW} の母平均を推定、ただし

$$S_i^{AIPW} = S_i^{Plugin}$$

$$+\frac{D_i(Y_i - g_{Y(1)}(X_i))}{g_D(X_i)} - \frac{(1-D_i)(Y_i - g_{Y(0)}(X_i))}{1-g_D(X_i)}$$

AIPW の性質

- Partialling out と本質的に同じ性質
  - $g_{Y(d)}, g_D$ が少なくとも $n^{-1/4}$ よりも早い速度で $E_P[Y|D,X], E_P[D|X]$ に収束するのであれば、推定値は $\sqrt{N}$ CAN estimator になる
  - 注意: 一致推定量であることは大前提
- 理由も同じ
  - Oracle 推定量に $\sqrt{N}$ 以上の速度で収束する
  - 推定誤差同士の掛け算になるため

Overlap 問題

- 因果推論/比較研究における重要な **Identification の仮定**
  - Sumamry/Estimation においても厳重な注意が必要
- 注意が払われていない応用研究も散見される

Overlap の仮定: Identification

- すべての x について、$1 > E_P[D|X = x] > 0$
  - 満たされないと、原理的に比較できない (母集団において存在しない) グループが存在
  - Well-specifeid model であれば、外挿によって" 可能" になる (後述)

Overlap の重要性: Estimation

- $1 > E_P[D|X = x] > 0$ が成り立っていたとしても、$E_P[D|X = x] \simeq 0$または1 であれば識別できても、推定困難
- AIPW の推定量も、漸近的に Oracle 推定量と同じ挙動をする
  - Oracle 推定量の性質が悪いと、実際の統計量も性質が悪くなる

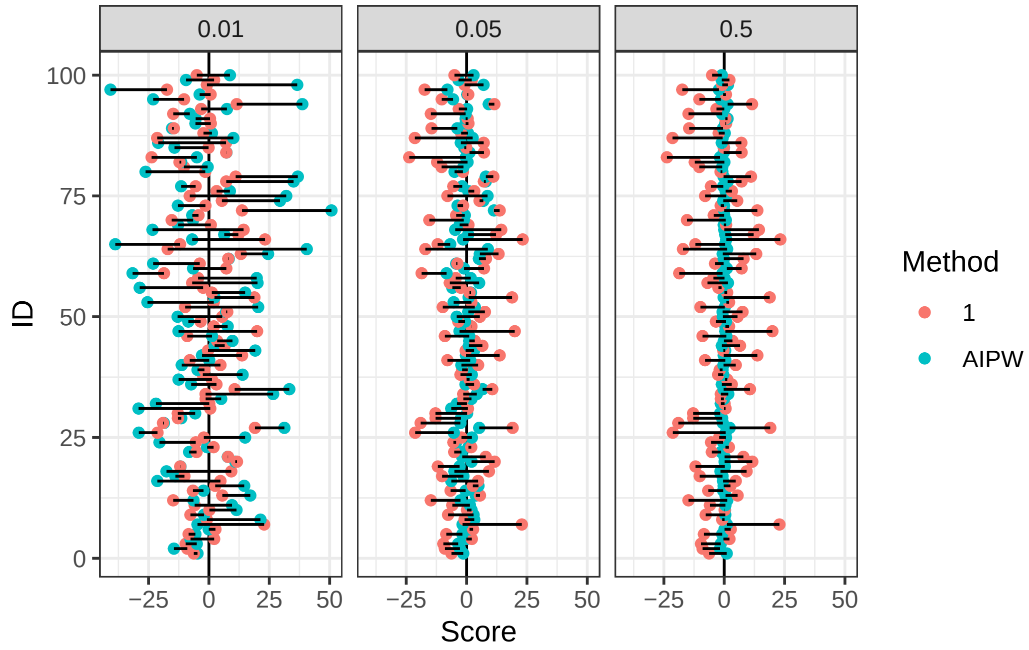
Oracle 推定への影響

$$S_i^{AIPW} = g_{Y(1)}^P(X_i) - g_{Y(1)}^P(X_i)$$

$$+ \underbrace{\frac{D_i(Y_i - g_{Y(1)}^P(X_i))}{g_D^P(X_i)} - \frac{(1 - D_i)(Y_i - g_{Y(1)}^P(X_i))}{1 - g_D^P(X_i)}}_{\text{"不安定"}}$$

- g^P : 母平均
- g_D^P が 0 に近いグループがいれば、実質的にそのグループの実現値が推定値を決める
  - S_i^{AIPW} の母分散が爆発

数値例



まとめ

- $E_P[D_i|X=x]$ が 0 ないし 1 になるグループがあれば、“識別できない”
- ゼロに近いグループがいれば、**推定**困難
  - frequeny weight を用いた summary は、“実装困難”

注意: Overlap 問題の普遍性

- かつての実証研究ではしばしば、Overlap 問題に大きな関心が払われなかった
  - 計算上は、推定誤差の爆発が起きない (ただし解釈困難な) 方法で推定してきた
  - 教科書はしばしば well specified model を前提としてきた

注意: Overlap 問題の普遍性

- Partialling out であれば、Overlap に問題があっても、推定値の” 信頼区間” は爆発しない
  - Overlap の問題を抱えているサブグループを” 無視している” だけ

- 直感的な解釈からの乖離拡大

- 理想的な実験データであれば、Overlap 問題は生じない
- Well-specified model があれば、外挿によって解決できる
  - “東京とそれ以外で賃金格差が変わらない” のを知っているので、overlap の問題が生じないサブグループで推定すれば良い

対応

Research question の修正

- 原理的に比較できないサブグループ $E[D|X] = 0, 1$ は、分析対象から外す
  - 直接比較を行う” 実証研究” としては不適切
  - 他のアプローチ (理論?) で頑張る

Moving goal posts

- 識別可能かつ推定可能な Estimand に目標変更する
  - Summary を変える
- 有名なものとして
  - Propensity score weight
  - Triming
  - Variance weight (既習)

Variance weight

- Variance weight であれば、推定困難なサブグループへの Weight が自動的に下がる
- Overlap が微妙でも、推定値は” 安定的” だが、
  - Overlap の度合いは、Estimand の解釈に決定的に重要
  - Variance が少ないグループは、勝手に無視されている

Propensity score weight

- $\omega_P(X) = E_P[D|X] \times f_P(X)$ ないし $= (1 - E_P[D|X]) \times f_P(X)$

- $D=1$ (あるいは $D=0$) の比率が大きいサブグループを重点評価
  - Variance よりは解釈しやすい？
- ある種の Overlap 問題を解消する

グループサイズの不均衡への対応

- しばしば 特定の $D=d$ の人数が極めて少ない場合がある
  - 金銭的制約等により、東大生の 5% のみを強制的に留学させる実験
  - $E_P[D|X]\sim 0$ が発生しがち
- $E_P[D|X]$ を Weight として使用すれば、そのようなグループは無視される

ATET

- 因果推論の文脈で、Propensity score weight は、Average treatment effect in treated (ないし controlled) とも呼ばれる
  - 解釈も比較的容易
  - 識別の家庭のもとで、介入を受けたグループ ($D=1$) 内での平均効果
    * ないし、受けなかったグループ内での平均効果

まとめ

- AIPW はカテゴリカルな D についての、"Default standard"
  - DoubleML vignett
  - ただし overlap は深刻な問題
- 他にも選択肢
  - overlap weight (Crump et al. 2006)
  - TMLE (Van Der Laan and Rubin 2006)
  - "soft intervention" (Kennedy 2019)
- 推定の容易さ VS 解釈
  - 盛んに研究されている

Reference

Crump, Richard K, V Joseph Hotz, Guido Imbens, and Oscar Mitnik. 2006. "Moving the Goalposts: Addressing Limited Overlap in the Estimation of Average Treatment Effects by Changing the Estimand." National Bureau of Economic Research Cambridge, Mass., USA.

Kennedy, Edward H. 2019. "Nonparametric Causal Effects Based on Incremental Propensity Score Interventions." *Journal of the American Statistical Association* 114 (526): 645–56.

Robins, James M., and Andrea Rotnitzky. 1995. "Semiparametric Efficiency in Multivariate Regression Models with Missing Data." *Journal of the American Statistical Association* 90 (429): 122129. https://doi.org/10.1080/01621459.1995.10476494.

Van Der Laan, Mark J, and Daniel Rubin. 2006. "Targeted Maximum Likelihood Learning." *The International Journal of Biostatistics* 2 (1).