

AIPW の発展

機械学習の経済学への応用

川田恵介

Plug-in Principle

Tutorial papers

- Fisher and Kennedy (2021)
- Hines et al. (2022)
- Ichimura and Newey (2022)

Estimand

- Estimand: 推定目標
 - 母集団で定義される要約統計量
 - 任意の分布 P からの写像として定義: $T(P) \subset R_L$
- 予約統計量: 複雑な分布を要約する量 (例: 平均、分散、Best Linear Projection)
 - データ: データ上の分布 \hat{P} (観察される) から計算 $T(P)$
 - 母集団: 観察できない母分布 P から計算 $T(P)$
 - * 推定する必要がある

例

- $p(w)$ 変数群 W の密度関数
- 平均値: $T(P) = \int Y p(Y) dY$
- 歪度もどき: $T(P) = \int [Y - \int Y p(Y) dY]^3 p(Y) dY$
- 加重平均値: $T(P) = \int Y p(Y|D, X) \times p(X) dX dY$

Plugin Estimation

- Estimand: 母分布 P で計算した $T(P)$
- データ上の分布 \hat{P} を代入 (Plugin) $T(\hat{P})$
 - 一般に $T(P) \neq T(\hat{P})$
 - $T(\hat{P})$ の分布が正規分布に収束するとは限らない
- 方針 1: \hat{P} の推定精度改善 (回帰/機械学習による Smoothing など)
- 方針 2: 推定された一次近似線で補正 (OneStep Adjustment)
 - AIPW, PartiallingOut もその一種

OneStep Adjustment

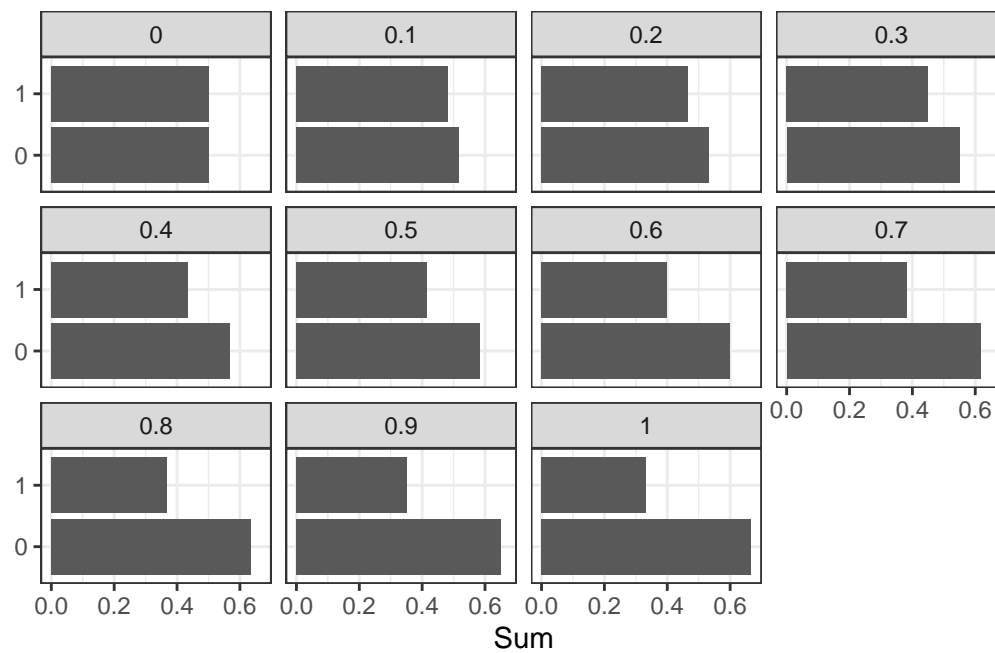
Parametric SubModel

- $p_t = t\hat{p} + (1-t)p$
- 仮想的な推定値: $T(P_t)$
- 補正済み推定量 $= T(P_0) - \underbrace{\partial T(P_t)/\partial t|_{t=1}}_{\text{推定する必要がある}}$

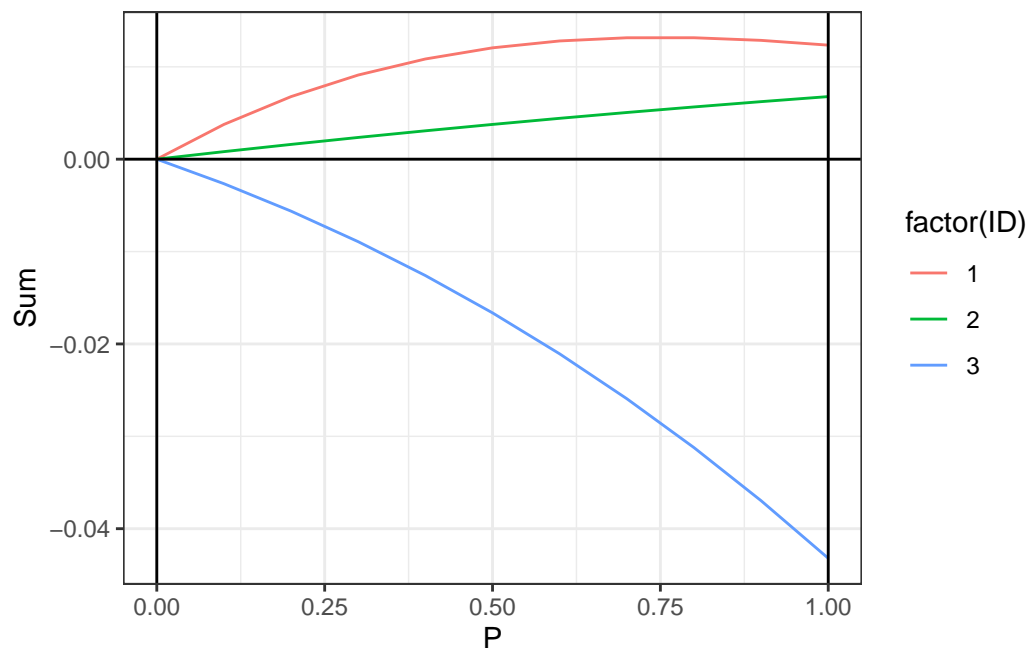
数値例

- 母集団: $p(Y = 1) = p(Y = 0) = 0.5$
- Estimand $\int [Y - \int Y p(Y) dY]^3 p(Y) dY = 0$

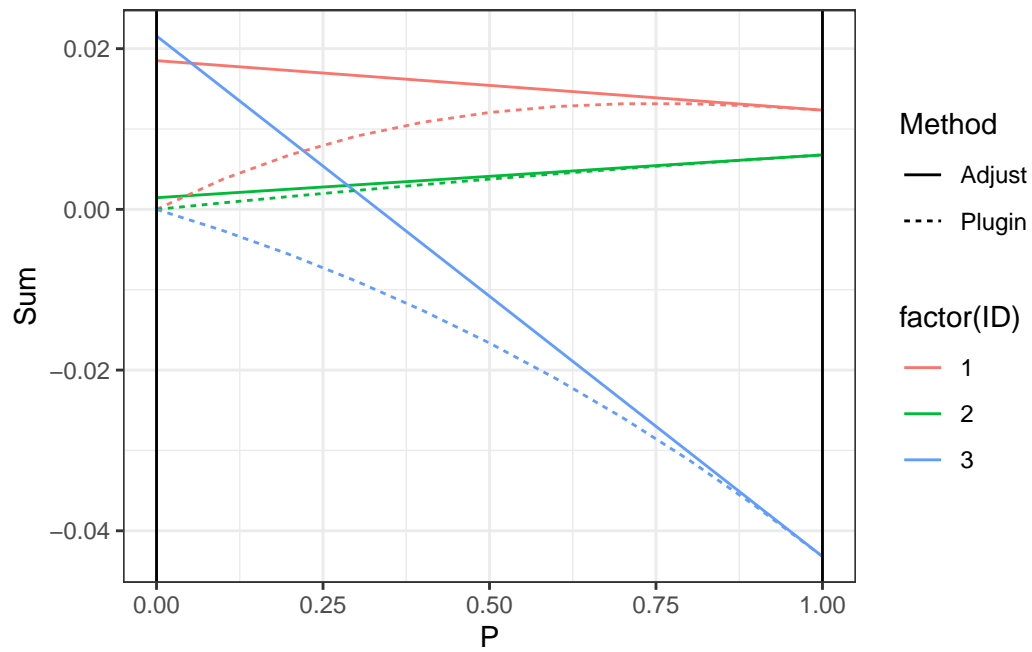
イメージ: N = 30



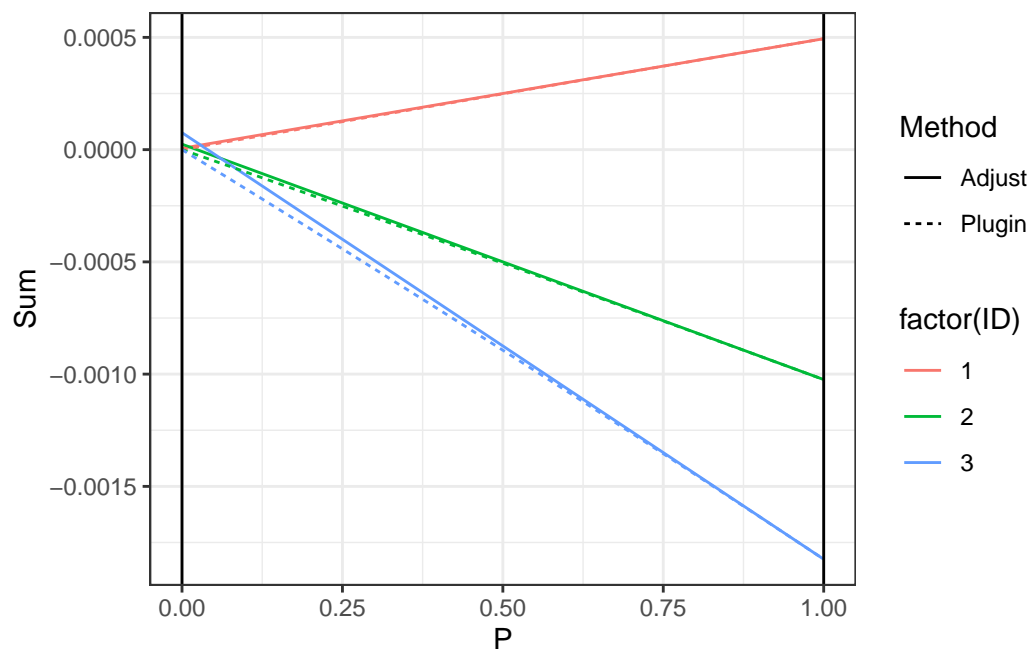
イメージ: Parametric SubModel



イメージ: OneStep Adjustment



イメージ: N = 1000



- 傾き”さえ”推計できれば OK

OneStep Adjustment の推定

$$T(P) = \underbrace{T(\hat{P}) - T'(P_t)|_{t=1}}_{OneStepEstimator} + \underbrace{Residual}_{\rightarrow 0}$$

- 一次微分なので、平均値の推定問題となる!!!

例: AIPW

- Estimand: $\Phi_1(P_t) - \Phi_0(P_t)$

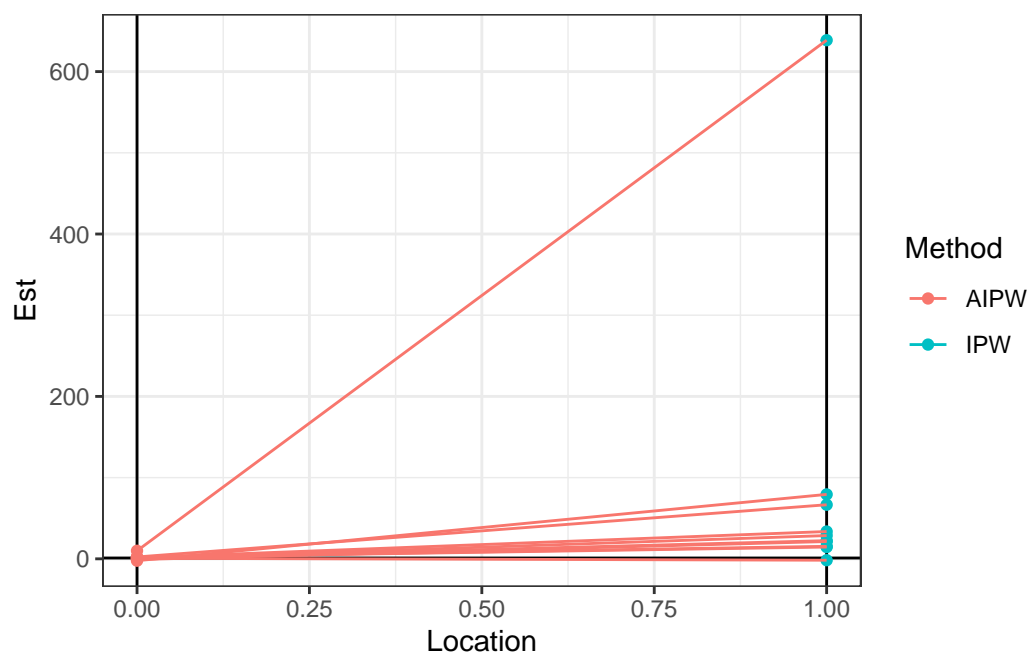
$$\begin{aligned}\Phi_d(P_t) &= \int Y p_t(Y|1, X) p_t(X) dY dX \\ &= \int Y p_t(Y, 1, X) p_t(X) p_t(1, X)^{-1} dY dX\end{aligned}$$

例: AIPW

$$\begin{aligned}T(\hat{P}) - \Phi'(P_t)|_{t=1} &= \\ &= \underbrace{\int Y \hat{p}(1|X)^{-1} p(Y, 1, X) dY dX - \int Y \hat{p}(Y|1, X) \hat{p}(1|X)^{-1} p(1, X) dY dX}_{\simeq \sum_i D \times (Y - f_Y(X)) / f_D(X)} \\ &\quad + \underbrace{\int Y \hat{p}(Y|1, X) p(X) dY dX}_{\simeq \sum_i f_Y(X)}\end{aligned}$$

数値例: AIPW: n = 500

- $X \sim U(-2, 2), \Pr[D = 1] = \Pr[D = 0] = 0.5$
- $Y = D + 10 * X + 10 * I(X^2) + \underset{N(0,1)}{u}$



Partiallingout 推定

- $T(P) = E\left[\frac{\text{Var}(D|X)}{E[\text{Var}(D|X)]} \times \beta(X)\right]$ の OneStep Adjustment 推定量
 - $\beta(x) : X = x$ グループ内の Y と D についての Best Linear Projection
- Vansteelandt and Dukes (2022)

補正の精度

- 一般に $T(P) = \text{OneStepAdjustment} + \text{Residual}$
- OneStepAdjustment の推定 = 平均値の推定
 - 漸近正規性を見たしながら、効率的な推定が可能
- Residual を本当に無視して良い？
 - 元々の Plugin 推定の精度
 - T の Smoothness
- AIPW, Partiallingout について、NuisanceFunction の推定値が n^{-4} 以上の速度で真の値に収束すれば OK
 - 一致推定量でなければ（誤定式化があれば）、満たし得ない (機械学習の比較優位)

AIPW の問題点

使用上の注意

- ほとんどの応用例において、PartiallingOut に比べて、AIPW の方が解釈が容易
 - 加重平均 VS 単純平均
- AIPW は、推定精度が極めて悪化する場合がある
 - 格差研究や” 荒い ” 自然実験データにおいて深刻

AIPW 推定量

$$\sum_i f_Y(D = 1, X_i) - f_Y(D = 0, X_i) \\ + \frac{D_i(Y_i - f_Y(D = 1, X_i))}{f_D(X_i)} - \frac{(1 - D_i)(Y_i - f_Y(D = 0, X_i))}{1 - f_D(X_i)}$$

- 傾向スコア $f_D(X)$ が 0,1 に近い値をとる事例があれば、推定誤差がインフレする

直感

- やりたいことは $E[E[Y|D = 1, X] - E[Y|D = 0, X]]$
- $f_D(X) \simeq 0$: $D = 1$ が極めて少ないサブグループ → 比較困難
 - $f_D(X) \simeq 1$ についても同様
- 理想的な RCT では問題ない
 - X の分布が偏っているかもしれない、格差・自然実験では大問題

解決: Moving Goal Post

- そもそも推定困難な Estimand をターゲットにしている
- Estimand を以下に変更
 - 加重平均差
 - $\Pr[D|X]$ が 0 や 1 に近いグループの排除 (Trimming) Crump et al. (2009)
 - $D = 1$ 内での比較 (AverageTreatmentEffect on Treated)

Reference

- Crump, Richard K, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik. 2009. “Dealing with Limited Overlap in Estimation of Average Treatment Effects.” *Biometrika* 96 (1): 187–99.
- Fisher, Aaron, and Edward H Kennedy. 2021. “Visually Communicating and Teaching Intuition for Influence Functions.” *The American Statistician* 75 (2): 162–72.
- Hines, Oliver, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. 2022. “Demystifying Statistical Learning Based on Efficient Influence Functions.” *The American Statistician*, 1–13.
- Ichimura, Hidehiko, and Whitney K Newey. 2022. “The Influence Function of Semiparametric Estimators.” *Quantitative Economics* 13 (1): 29–61.
- Khan, Samir, and Johan Ugander. 2022. “Heteroscedasticity-Aware Sample Trimming for Causal Inference.” *arXiv Preprint arXiv:2210.10171*.
- Vansteelandt, Stijn, and Oliver Dukes. 2022. “Assumption-Lean Inference for Generalised Linear Model Parameters.” *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. <https://doi.org/10.48550/arXiv.2006.08402>.