SemiParametric 推定

異質性分析

川田恵介

Table of contents

Semiparametric 推定への応用 1
「異質性を調べよ」 1
ここまでの手法 2
Treatment effect risk 2
Get Start 2
性質 3
実例 3
実例 4

Neyman の直行条件 4
Estimand の一般的定義 4
Estimand の一般的定義 4
仮定の区分 4
Parametric Assumption の有無 5
練習問題 5
モーメント法: 単純なケース 5
モーメント法: 複雑なケース 5
Neyman's ohthogonal condition 5
まとめ 6
Reference 6

Semiparametric 推定への応用

「異質性を調べよ」

- 具体的には？
    - $\tau_P(X)$ に特に大きな影響を与える X を知りたい

        * BLP の活用

    - $\tau_P(X)$ がどの程度異なるのか知りたい

        * ?

ここまでの手法

- 条件付き平均差 $\tau^P(X)$ を
    - BLP で近似: 定式化依存し、ApproximationError が大きい可能性
    - NonParametric に近似: 信頼区間計算に不安, Too much な推定かも

Treatment effect risk

- Kallus (2022) により提案
- $$E_P[\tau_P(X)|\tau_P(X) \leq Q_P(q)]$$
    - $$Q_P(q) =$$

      qth quantile of $\tau_P(X)$
    - $\tau_P(X)$ のモデル $g_\tau(X)$ を Nuisance として活用する
- CHERNOZHUKOV et al. (2022) と類似したアイディア
    - Estimand が完全に母集団上で定義できており、解釈がより明確

Get Start

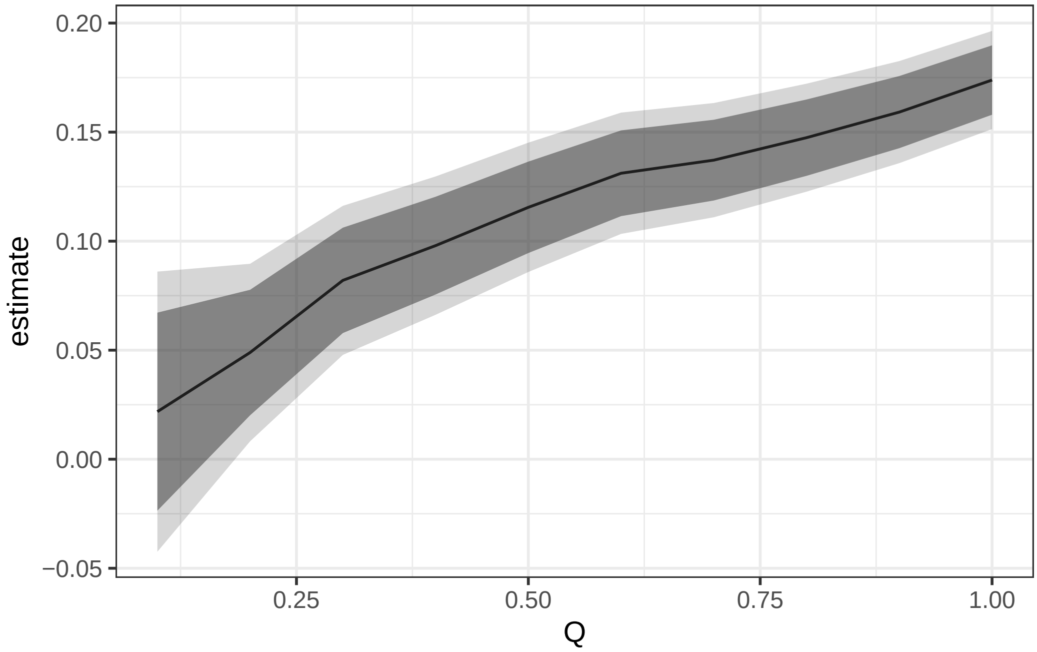
1. データを 3 分割 (交差推定可能)
2. $g_y(D,X), g_D(X)$ を (教師付き学習などで) 推定
3. $m^{AIPW} \sim g_\tau(X)$ を (教師付き学習などで) 推定
4. $E[m_i^{Risk}]$ を推定

- $$m_a^{Risk} = Q_P(q)$$
- $$+\frac{I(g_\tau(X) \leq Q_P(q))}{a}[m^{AIPW} - Q_P(q)]$$

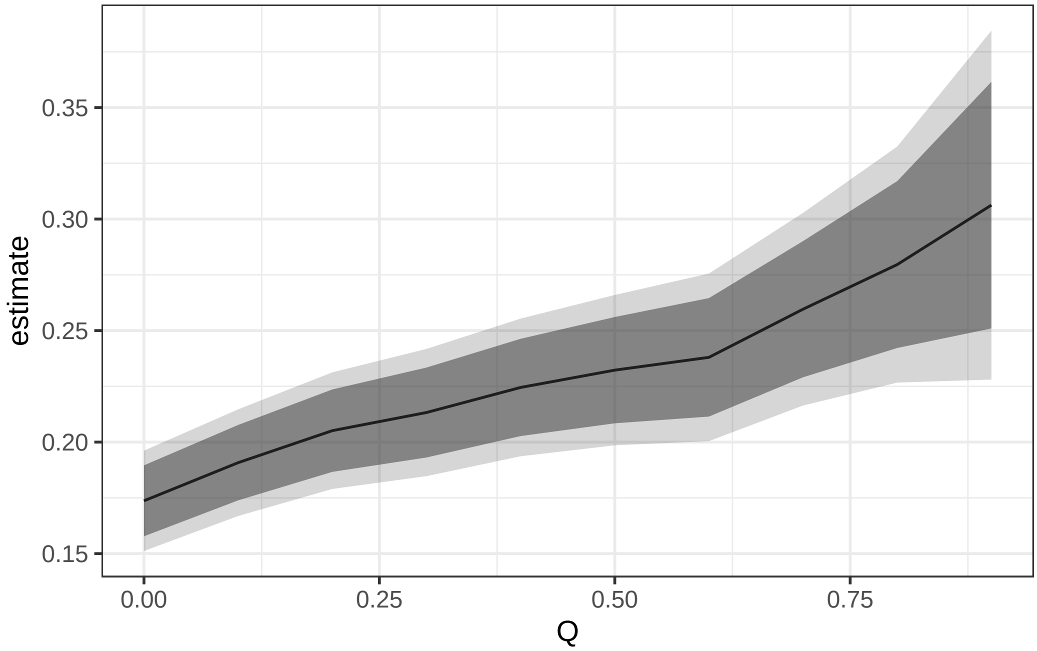
性質

- **Neyman の直行条件**を満たすので、g が $n^{-1/4}$ 以上の速度で収束するのであれば、g の推定誤差は漸近分布を計算する際に無視できる
- 信頼区間が形成可能

実例



実例



Neyman の直行条件

Estimand の一般的定義

- 研究課題 -> 識別 -> 要約から、Estimand θ を**定義**する分布 $Distribution$ の関数 T を**定義**
- $$T(Distribution) = \theta$$
- 例: $\theta =$ 平均
- $$T(Distribution) = \int Y \times f(Y) \times dY$$

Estimand の一般的定義

- ポイント: $T =$ データ/母集団分布、どちらについても定義できる" 記述統計量"

仮定の区分

- 推定には、Distribution/Estimand 双方について、“仮定”(定式化) が必要

- Parametric Assumption (有限個のパラメタで記述できるとするかどうか) の有無で区分できる
  - ポイント: Distribution/Estimand　への仮定をしっかり区別

Parametric Assumption の有無

Estimand/Distribution	有	無
有	教科書的な構造推定	セミパラメトリック
無	古典的な予測研究?	教師付き学習による予測研究

練習問題

- 母平均をサンプル平均で推定
- OLS

モーメント法: 単純なケース

- セミパラメトリック推定の代表的な方法
- 母集団上で Estimand を**定義**する式を設定
- $$E_P[m(\theta_P, Y, D, X)] = 0$$
- 設定可能であることは**仮定**
- 例: 母平均 $m = Y - \theta$

モーメント法: 複雑なケース

- $$E_P[m(\theta_P, Y, D, X, g)] = 0$$
- g Nuisance 関数
  - $E_P[Y|X], E_P[D|X], E_P[\tau|X]$ の Nonparametric な推定値

Neyman's ohthogonal condition

- $$\partial E_P[m(\theta, Y, D, X, g)]/\partial g|_{g=E_P} = 0$$

– E_P : “真” の関数

– 注: 関数微分

- g が真の値から微妙にずれていたとしても、m は影響をほとんど受けない

  – m^{Risk} は満たす

- 漸近性質が、$E[m(\theta, Y, D, X, E_P)]$ と等しくなる

まとめ

- Neyman の直行条件を満たすように設定することは、幅広い Estimand について可能

  – 緩い条件のもとで、Nuisance 関数の推定誤差を” 無視” できる

- 応用例は幅広い

  – 操作変数方 (Chernozhukov, Hansen, and Spindler 2015)、差の差の推定 (Chang 2020) や、Mediation 分析 (Farbmacher et al. 2022; Díaz et al. 2021), Sensitvity 分析 (Chernozhukov et al. 2022)

Reference

Chang, Neng-Chieh. 2020. “Double/Debiased Machine Learning for Difference-in-Differences Models.” *The Econometrics Journal*.

Chernozhukov, Victor, Carlos Cinelli, Whitney Newey, Amit Sharma, and Vasilis Syrgkanis. 2022. “Long Story Short: Omitted Variable Bias in Causal Machine Learning.” National Bureau of Economic Research.

CHERNOZHUKOV, VICTOR, MERT DEMIRER, ESTHER DUFLO, and IV AN FERN ANDEZ-VAL. 2022. “Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India.”

Chernozhukov, Victor, Christian Hansen, and Martin Spindler. 2015. “Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach.” *Annu. Rev. Econ.* 7 (1): 649–88.

Díaz, Iván, Nima S Hejazi, Kara E Rudolph, and Mark J van Der Laan. 2021. “Nonparametric Efficient Causal Mediation with Intermediate Confounders.” *Biometrika* 108 (3): 627–41.

Farbmacher, Helmut, Martin Huber, Lukáš Lafférs, Henrika Langen, and Martin Spindler. 2022. “Causal Mediation Analysis with Double Machine Learning.” *The Econometrics Journal* 25 (2): 277–300.

Kallus, Nathan. 2022. “Treatment Effect Risk: Bounds and Inference.” *2022 ACM Conference on Fairness, Accountability, and Transparency (Minor Revision in Management Science).*