自主ゼミ
『データ駆動型回帰分析』勉強会 概要説明

前川 大空 (一橋大学経済学部 4 年)
佐藤ゼミ 24 期, 五年一貫専修コース公共経済プログラム

May 11, 2025

勉強会の目標

- ML に興味がある
- 『R 勉強会』をめでたく走破したのでその続編として
- 参加者の皆さんの研究に活用できる可能性
- 将来エコノミストとして働く際の強力な分析ツールとしての可能性

テキスト

regression analysis
データ駆動型回帰分析
計量経済学と機械学習の融合
末石直也
Naoya Sueishi
日本評論社

テキスト

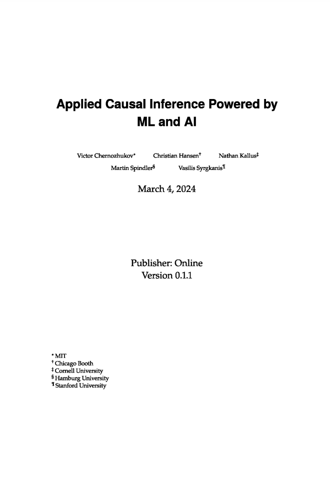
メインテキスト:

1. **末石 (2024)『データ駆動型回帰分析』**

サブテキスト:

1. 末石 (2015)『計量経済学』
2. 黒住 (2016)『計量経済学』
3. Chernozhukov 他 (2024) "Applied Causal Inference Powered by ML and AI"

テキスト


Applied Causal Inference Powered by ML and AI

Victor Chernozhukov* Christian Hansen† Nathan Kallus‡
Martin Spindler§ Vasilis Syrgkanis¶

March 4, 2024

Publisher: Online
Version 0.1.1

* MIT
† Chicago Booth
‡ Cornell University
§ Hamburg University
¶ Stanford University

テキスト

メインテキスト:

1. **末石 (2024)『データ駆動型回帰分析』**

サブテキスト:

1. 末石 (2015) 『計量経済学』
    - 院レベル, 記述スタイルが末石 (2024) と同一 (証明は省かれ, 意義を端的に説明)
    - **行間埋めは必須.** 標準的なコアコース (上級計量) の内容復習に
2. 黒住 (2016) 『計量経済学』
    - 学部中上級レベル, 一般的な計量経済学の知識 (中級計量) の復習用に
3. Chernozhukov 他 (2024) "Applied Causal Inference Powered by ML and AI"
    - 院レベル, 機械学習と因果推論を組み合わせた手法について. 無料.
    - 英語の本で長いが, その分解説は丁寧なので詰まったときに参照できるかも
    - **R と python のコードが実装**されており (!), 演習にはとても良い

余談: 末石本は端的ゆえに難しい

末石のスタイル: 証明を省き, 意義のみを端的に説明する

- よくまとまっているし, 評価されるのも分かる
- ちゃんと読んで, 時には計算しないと流れを見失う

余談: 末石本は端的ゆえに難しい

末石 (2015) から例を挙げてみよう:

- 単回帰モデルの古典的仮定のもとでは OLSE $\hat{\beta}$ は以下の厳密分布に従う:

$$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\Sigma_i (X_i - \bar{X})^2})$$

- ところが, より一般化した現代的な仮定の下では, 特に不均一分散を考えると, この議論では正しい標準誤差 (≈ 標準偏差) を求められず, 推定が失敗する.

Q. 何を言っているのだろうか???

A. 実際に計算すれば分かる (次頁)

$$Var(\hat{\beta}_1) = \mathbb{E}[(\hat{\beta}_1 - \beta_1)^2] = \mathbb{E}[(\frac{\sum_{i=1}^n (X_i - \bar{X}) u_i}{\sum_{i=1}^n (X_i - \bar{X})^2})^2] \overset{\text{LIE}}{=} \mathbb{E}[\frac{\mathbb{E}[(\sum_{i=1}^n (X_i - \bar{X}) u_i)^2 | \mathbf{X}]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2}].$$ 期待値内の分母を考えよう.

$$(\text{denominator}) = \mathbb{E}[(\sum_{i=1}^n (X_i - \bar{X}) u_i)^2 | \mathbf{X}] = \mathbb{E}\left[S^2 | \mathbf{X}\right] \text{ where } S = \sum_{i=1}^n (X_i - \bar{X}) u_i$$

$$\mathbb{E}\left[S^2 | \mathbf{X}\right] = \sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}] + 2 \sum \sum_{i<j} (X_i - \bar{X})(X_j - \bar{X}) \mathbb{E}[u_i u_j | \mathbf{X}].$$

先述の議論より, A1 より**誤差項同士が独立**, つまり**相関がない** (∵独立性の必要条件) ので:

$$\mathbb{E}[u_i u_j | \mathbf{X}] = \mathbb{E}[u_i | \mathbf{X}] \mathbb{E}[u_j | \mathbf{X}] = 0 \;,\; \therefore \mathbb{E}\left[S^2 | \mathbf{X}\right] = \sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}] = \sigma_i^2 \;\forall i \neq j$$

$$Var(\hat{\beta}_1) = \mathbb{E}[\frac{\sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2}]$$

$$Var(\hat{\beta}_1 | \mathbf{X}) \overset{\text{LIE}}{=} \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \mathbb{E}[u_i^2 | \mathbf{X}]}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 \sigma_i^2}{[\sum_{i=1}^n (X_i - \bar{X})^2]^2}$$

更に均一分散の仮定を置けば:

$$Var(u_i | X_i) = \mathbb{E}[u_i^2 | \mathbf{X}] = \sigma^2 \;\forall i \;,\; \therefore \mathbb{E}\left[S^2 | \mathbf{X}\right] = \sigma^2 \sum_{i=1}^n (X_i - \bar{X})^2$$

$$Var(\hat{\beta}_1) = \sigma^2 \mathbb{E}[\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}], \; Var(\hat{\beta}_1 | \mathbf{X}) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

余談: 末石本は端的ゆえに難しい

末石 (2015) の例:

- 計算過程を文字に直すとしたらまさにその通り, 非常に良くまとまっている
- だが基本軽く流すだけでは意味が分からず, 吸収できない

前提知識

曰く:

- 『本書では**学部上級レベルの計量経済学**の知識を前提としています。『計量経済学』 (西山慶彦 他, 2019) くらいを想定していますが、それにプラスして、**線形代数**の知識も多少必要です。』

1. 基礎計量の内容は既知とする
2. 中級計量相当の内容は, 履修済みまたは履修中であること
3. 線形代数の rank についての議論程度は備わっていると望ましい
4. **議論をちゃんと追う姿勢.** ごまかして読むと途端に分からなくなります…

日程・進め方

- 6 月下旬あたりに開始
- しっかりと難しいので, 焦らずゆっくりと進めましょう
- 半年ぐらいで終われるとうれしい
  - 1 か月あたりおよそ 1 章
  - ひと月に 2 回, 一度の勉強会で 2〜3 節扱う
- というか 5 章までたどり着けるだけでも嬉しい
- 対面・オンライン併用
  - 8 月までは前川がドイツにいるので, 日本時間の休日夜開催になると思います
  - 日本に帰ったらぜひ対面でやりましょう!

さいごに

- 皆さまのやる気, 大歓迎です
- マジで難しいと思うので, 適度に妥協しながら頑張りましょう!

参考文献

- 末石 直也 (2024), データ駆動型回帰分析 計量経済学と機械学習の融合, 第一版, 日本評論社