

# Approximation of Conditional Average Difference

機械学習の経済学への応用

川田恵介

## 冬休みのおすすめ

- [Modern Business Analytics](#)
  - 学部初級/中級?
  - R のコード付き
  - データ分析最初の一冊としても OK
  - 社研の図書館に一冊あります

## ここまでの復習

### よくある書き方

- いきなり統計モデルの定式化から議論が始まる
- 例

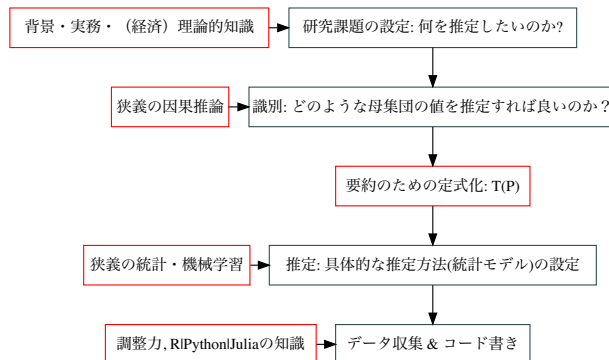
$$Y_i = \beta_0 + \beta_D D_i + \beta_1 X_i + u_i$$

- $u_i \sim N(0, \delta)$
- 全ての仮定 (識別の仮定、推定の仮定、**要約 (parsimonious) ための仮定**) が、ごちゃごちゃに定式化
  - 仮定  $\rightarrow$  結論が見えにくい

### おすすめ: RoadMap

- 経済学と同様に、必要な仮定を少しずつ設定する
- RoadMap: [Chapter 4 in Target Learning in R](#)

## 簡易版



## Conditional Average Difference

- 母集団における  $f_{\tau}(x) = f(Y|D = d, X = x) - f(Y|D = d', X = x)$  について議論できれば、研究課題に答えられる
- 一般に  $f_{\tau}(x)$  を直接推定することは難しい（次元の呪い、密度の推定）、推定結果が人間の認知能力を超える（解釈できない）
  - 社会を理解する上で too much な目標かもしれない
- 要約のための定式化: Partialling-Out | AIPW:  $T(\tau(d, d', x)) = \int [\tau(x) \times weight(x)] dx$  where  $\tau(X) = \int Y f_{\tau}(X) dX$  スカラー（平均差）
  - かなり節約的な目標: 推定容易
- 推定のための仮定: Nuisance function の推定値が  $n^{1/4}$  以上の速度で収束

## 異質性

- $X$  と  $f_{\tau}(X)$  の間にどのような関係性があるのか?

- 非常に重要
- 実務的含意: どの様な層に効果的なのか、負の効果が生じる層はないのか、“格差”はどの様な層で深刻なのか?
- 理論的含意: 理論への反証力を高める

## 丸暗記モデル

- 節約のための仮定:  $\tau(X) = \int Y f_\tau(X) dX$  を推定
- 各サブグループ  $X$  について個別に Partiallingout、AIPW 推定
  - $X$  が多次元のケースでは、サンプルサイズが減りすぎる
  - 計算負荷が大きい

## 節約のための定式化

- 節約のための仮定:  $\tau(X) \sim \beta_0 + \beta_1 Z_1 + \dots + \beta_L Z_L$ 
  - $Z_1, \dots, Z_L = X$  の一部: 研究課題に応じて指定
- 複数の推定アルゴリズム
  - 極力推定のための仮定を減らしたい

## 伝統的交差項

- 線形近似モデルを推定

$$E[Y|D = d, X = x] = \tau_D \times D + \tau_1 \times D \times \bar{Z}_1 + \dots + \tau_L \times D \times \bar{Z}_L$$

$$+ \underbrace{f(X)}_{Linear Model}$$

- $f(X)$  の定式化の影響が大きい

## 補論: Centering

- $\beta_0 =$  全ての  $X$  が 0
  - 経済学への応用において、 $X = 0$  がナンセンスなケースは多い (年齢、身長、就学年数など)

- 近似する上で無視される
- 簡単な解決策は、変数の加工
  - Centering された  $X$   $\tilde{X} = (X - \text{mean}(X))$  の利用
  - $E[\tau(\tilde{X})|\tilde{X} = 0]$  : ” 平均的サンプル ” における平均効果の予測値
  - 経済学での応用では、” 平均 ” 周りにはサンプルが多いことが多いので、高い近似精度が期待できる

## Parital Linear Model with interaction

1.  $E[Y|X], E[D|X]$  を推定
2. 以下のモデルに代入し、OLS 推定

$$Y - E[Y|X] \sim D - E[D|X] + \bar{Z}_1 \times (D - E[D|X]) + .. + \bar{Z}_L \times (D - E[D|X])$$

- Neyman's ohthogonality を満たし、 $E[Y|X], E[D|X]$  の推定誤差を漸近的に無視できる

## BLP on Conditional Average Difference

- Semenova and Chernozhukov (2021)
1.  $E[Y|D, X], E[D|X]$  を推定
  2. AIPW スコアに代入し、Psude-Outcome  $Y_{AIPW}$  を計算
  3. 以下のモデルに代入し、OLS 推定

$$Y_{AIPW} \sim \tau_1 \bar{Z}_1 + .. + \tau_L \bar{Z}_L$$

## Implementation

- とりあえず DoubleML で平均差を推定すると、必要な Nuisance 推定値も提供してくれる
  - 非常に少ない労力で追加的な含意を得られる!!!
- ただし真面目に論じる場合は多重検定の補正が必要
- **いろんなモデルを推定し、都合のいい結果だけを報告することだけは厳禁**
  - 推定したパラメータは全て報告!!!

## 信頼区間の補正

- 複数のパラメータについて、信頼区間や統計的検定を行いたい
  - 注意しないと大きくミスリードしてしまう
- 多重検定用の修正方法が多く提案されている
  - 優れた入門: [Introduction to Statistical Learning: Chapter 13](#)

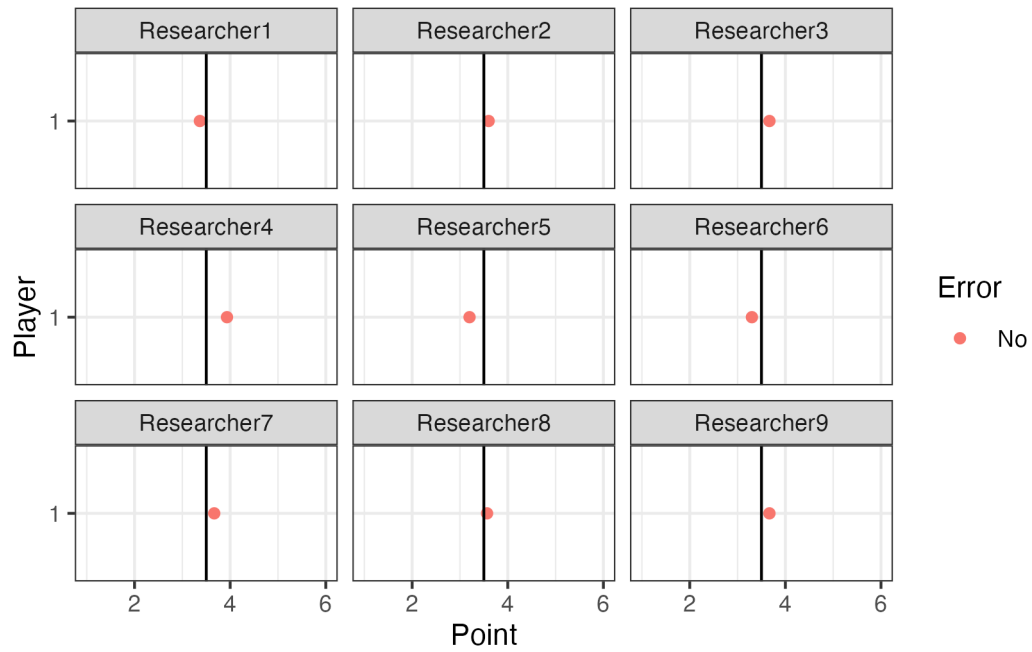
## 復習: Point-wise Confidence Interval

- 信頼区間: “同じデータ活用・サンプリング法を用いる研究者” の  $1 - \alpha$  が正しい値を含む信頼区間を獲得できる
  - “間違った” 区間を観察する研究者割合 (確率)  $\alpha$  をコントロール
- 特定の (Point-wise) 推定値について” 信頼できる” 信頼区間
  - 大量の推定値について適用できるか?

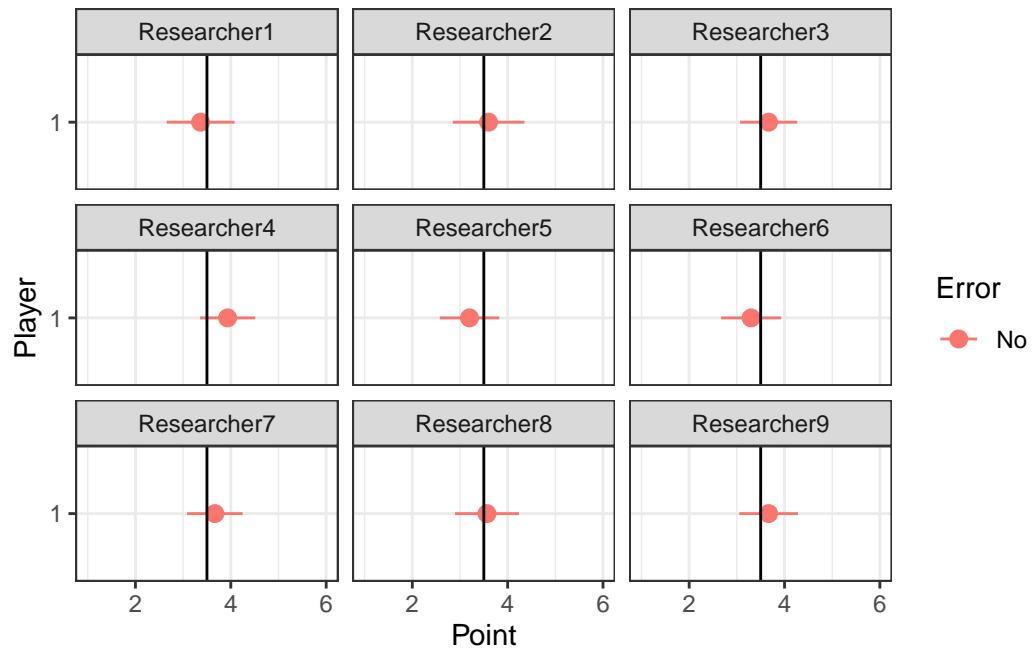
## 例: 適切

- Research question: 「サイコロの出目をコントロールできるか？」
  - サイコロの出目の平均値を操作できるか?
- 各研究者が一人のディーラーからのみデータを収集するのであれば、問題ない
  - 母集団 = ” 仕掛け” はない
  - 1 人につき 30 回サイコロを振らせる

適切: 1 player



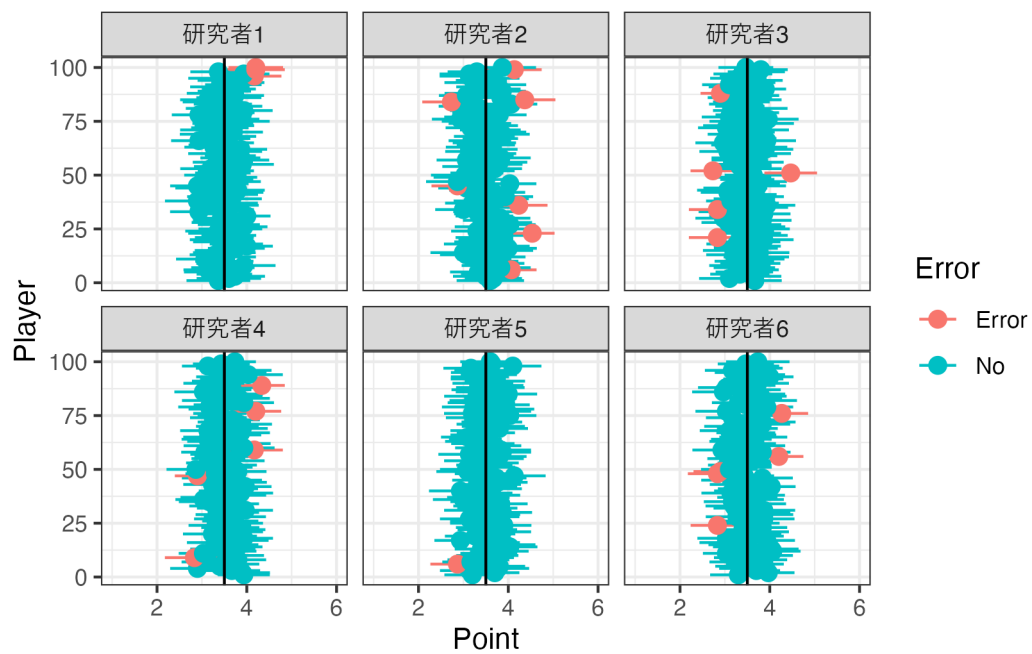
適切: 1 Player



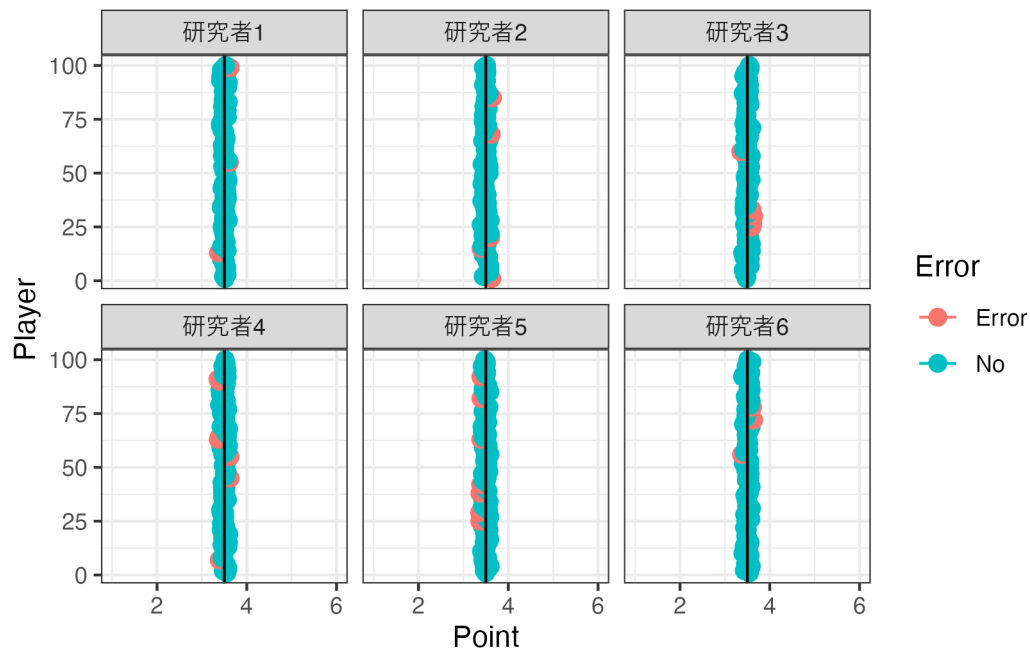
## 例: 不適切

- 各研究者が、「100名のディーラーから、操作できるディーラーを見つけられるか？」であれば、不適切
- 他の不適切な例
  - 影響を受ける指標 (所得、健康状態、主観的幸福の中から探索) はあるか？
  - 影響を与える特定のグループがあるか？

## 不適切: Small sample



不適切: Larger sample



## 問題点

- 「本来はほとんど差がないのに、サンプルが偶然偏った結果、大きな差が推定されてしまう」
  - 注目を浴びたい研究者にとっての当たりくじ
- 「どんなに当たり確率が低いくじであったとしても、無限回引けば”絶対”に当たる」
  - あやまって有意

## 基本戦略

- あやまって「サイコロを操作するディーラー」を発見する研究者の割合をコントロールするには、
- 信頼区間を適切に”広げる”

## 同時 (Family-wise) 信頼区間

- 同時信頼区間: 複数の推定値が前提
  - 一つ以上の信頼区間が真の値を含まない確率を  $\alpha_{Family}$  以内に抑える



- 複数の提案がある
- Bonferroni 法:  $\alpha_{Family} = \alpha_{Point} / \text{推定値数}$ 
  - 最もシンプルかつ仮定が少ない
  - 保守的な結果

### 発展: Bonferroni 法の根拠

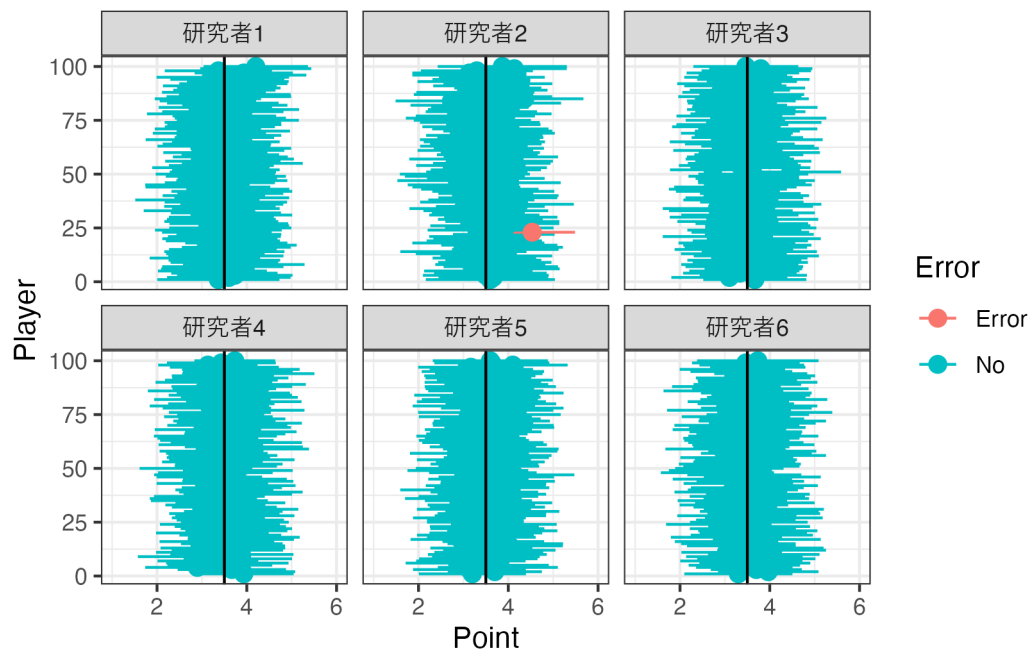
- 以下の一般原則を利用

一つ以上の区間について間違いが起こる確率 ( $= \alpha_{Family}$ )

$\leq$  区間1について間違いが起きる確率

+ ... + 区間Lについて間違いが起きる確率 ( $= L \times \alpha_{Point}$ )

適切: Larger sample with adjustment



### 発展: 他の手法

- 検定したいパラメータが多くなると、信頼区間が非常に広がる
- 多くの改善提案

- False Discovery Rate, Uniform inference などなど
- DoubleML は Uniform Inference も実装 ([DoubleML: Chapter 6](#))

## 実践への含意

- 避けるべきは、「明確な差が出ないので、大量の推定を行い、偶然の差を強引に見つけた」と疑われなこと
- 事前の分析計画が重要
  - 推定量を極力絞る!!!
  - 分析の Priority を明示する

## まとめ

- なんの考慮・対処をせずに多重検定を行なっている事例は極めて多い
  - いくらでもいかさま師を”発見できる”
- 事前の分析計画の作成と共有が重要
- 大規模データである程度の同時推定・検定は実用的

## Reference

Semenova, Vira, and Victor Chernozhukov. 2021. “Debiased Machine Learning of Conditional Average Treatment Effects and Other Causal Functions.” *The Econometrics Journal* 24 (2): 264289. <https://doi.org/10.1093/ectj/utaa027>.