傾向スコアによるセミパラメトリック統計解析

西畑 樹希　総合研究大学院大学 統計科学専攻 5年一貫制博士課程3年

1　因果推論における傾向スコア解析

1.1　欠測と反実仮想モデル

次のモデルを考える.

$$\boldsymbol{y} = z\boldsymbol{y}_1 + (1-z)\boldsymbol{y}_0 \ ; \ z = \begin{cases} 1 & \text{条件}1(\text{処置群}) \\ 0 & \text{条件}2(\text{対照群}) \end{cases}$$

- 結果変数$\boldsymbol{y}$に対して,割り当て条件zが存在し,条件1の時に$\boldsymbol{y}_1$,条件2の時に$\boldsymbol{y}_0$を観測するものとする.
  つまり,$z=1$という条件1の時,$\boldsymbol{y}_0$は観測されない. 観測されたものは条件1の元での$\boldsymbol{y}_1$だが,「もし仮に$z=0$という条件2だったならば」という反実仮想となる.つまり,ここでは,$\boldsymbol{y}_1-\boldsymbol{y}_0$,すなわち因果効果を調べたい.

1.2　因果効果と強く無視できる割り当て条件

平均を考える.

$$\mathrm{E}[\boldsymbol{y}_1-\boldsymbol{y}_0] = \mathrm{E}[\boldsymbol{y}_1] - \mathrm{E}[\boldsymbol{y}_0]$$

- これをルービンの因果効果(または平均処置効果)と呼ぶ.
  因果効果の一番naiveな推定量は

$$\hat{\mathrm{E}}[\boldsymbol{y}_1-\boldsymbol{y}_0] = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y}_{i1}-\boldsymbol{y}_{i0})$$

- $\boldsymbol{y}_{i1}$,$\boldsymbol{y}_{i0}$は対象者iに対する$\boldsymbol{y}_1$,$\boldsymbol{y}_0$の実現値.
- 潜在的な結果変数は必ずどちらかが欠測しているため,計算できない.
  ここで,強く無視できる割り当て条件を考える.

$$(\boldsymbol{y}_1, \boldsymbol{y}_0) \perp\!\!\!\perp z|\boldsymbol{x}$$

これは,各割り当ては共変量に依存するが,結果変数に依存しないことを意味する. 強く無視できる割り当て条件のもとで因果効果の推定として回帰モデルを用いることで,次を得ることができる.

$$\tfrac{1}{N}\sum_{i=1}^{N}(g(\boldsymbol{x}_i,\beta_1)-g(\boldsymbol{x}_i,\beta_0))$$

- 「$z=1$ でのy_1の$\boldsymbol{x}$への回帰関数」として$\mathrm{E}[\boldsymbol{y}_1|\boldsymbol{x}]$を$g(\boldsymbol{x},\beta_1)$
- 「$z=0$ でのy_0の$\boldsymbol{x}$への回帰関数」として$\mathrm{E}[\boldsymbol{y}_0|\boldsymbol{x}]$を$g(\boldsymbol{x},\beta_0)$
- ただし, β_1,β_0はそれぞれの回帰関数のパラメータ

以上より,因果効果の推定ができる.しかし, 実際には共変量の次元が多く,モデルを誤特定してしまう可能性がある.そこで,割り当てzは1次元であるため,傾向スコアを考える.

1.3　傾向スコア

傾向スコアを共変量$\boldsymbol{X}$を条件つけた割り当ての確率と定義する.

$$w(\boldsymbol{x};\boldsymbol{\alpha}) \equiv \mathrm{P}(z=1|\boldsymbol{X}=\boldsymbol{x})$$

- $\boldsymbol{\alpha}$は共変量から割り当てへの回帰モデルにおけるパラメータとする.

傾向スコアは例えばロジスティック回帰でパラメータを推定する.

$$w(\boldsymbol{x},\alpha) = \mathrm{P}(z=1|\boldsymbol{x},\alpha) = \tfrac{\exp(\alpha'x)}{1+\exp(\alpha' x)}$$

- 例えば,$\mathrm{P}(z=1|\boldsymbol{x})=0.7$の時, 70%の確率で条件1(処置群), 30%の確率で条件2(対照群)に割り当てる. 正の効果が大きくなるように,共変量の条件を決めることが多いため,逆数によって重みをつける.

1.4　Inverse Probability Weighting推定(IPW推定)と二重頑健推定

傾向スコアを用いたM推定として次のものがある.

Theorem 1 (IPW推定(Hoshino et al. (2006))). 目的関数を次のよう置く

$$Q_N(\boldsymbol{x},\boldsymbol{y},z;\boldsymbol{\theta},\boldsymbol{\alpha}) = \tfrac{1}{N}\sum_{i=1}^{N}\frac{z_i}{w(\boldsymbol{x}_i;\alpha)}m(\boldsymbol{y}_i;\boldsymbol{\theta})$$

$$\tilde{\boldsymbol{\theta}}(\boldsymbol{\alpha}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} Q_N(\boldsymbol{x},\boldsymbol{y},z;\boldsymbol{\theta},\boldsymbol{\alpha})$$

$\boldsymbol{x}_i,\boldsymbol{y}_i,z_i$を$\boldsymbol{x},\boldsymbol{y},z$の実現値とする. このとき, $\boldsymbol{\alpha}_0$を真値として, 推定量$\tilde{\boldsymbol{\theta}}(\boldsymbol{\alpha}_0)$は一致性と漸近正規性をもつ

Theorem 2 (IPW推定(Hoshino et al. (2006))). $\hat{\boldsymbol{\alpha}}$を漸近有効な推定量とする. 推定量$\tilde{\boldsymbol{\theta}}(\hat{\boldsymbol{\alpha}})$ は 一致性と漸近正規性を持つ. ただし,漸近分散は推定量$\tilde{\boldsymbol{\theta}}(\boldsymbol{\alpha}_0)$のものより小さくなる.

Theorem 3 (二重頑健推定Hoshino(2007)). 推定方程式を次のように置く

$$\sum_{i=1}^{N}\mathrm{S}_i(\boldsymbol{\theta};\boldsymbol{\alpha},\boldsymbol{\beta}) =$$

$$\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\theta}}\{\frac{z_i}{w(\boldsymbol{x}_i;\boldsymbol{\alpha})}m(\boldsymbol{y}_i;\boldsymbol{\theta}) + (1-\frac{z_i}{w(\boldsymbol{x}_i;\boldsymbol{\alpha})})\int m(\boldsymbol{y}_i;\boldsymbol{\theta})p(\boldsymbol{y}_i|\boldsymbol{x}_i,\boldsymbol{\beta})d\boldsymbol{y}_i\} = 0$$

モデルAを共変量$\boldsymbol{x}$と割り当てzのモデル(パラメータは$\boldsymbol{\alpha}$), モデルBを共変量$\boldsymbol{x}$で結果変数$\boldsymbol{y}$を説明する回帰モデル(パラメータは$\boldsymbol{\beta}$)とする.モデルA,モデルBの一方が正しければ,推定方程式の$\boldsymbol{\theta}$の推定量は一致性を持つ.

2　今後

M推定の満たす性質を考える.Asymptotically Linear Estimatorと呼ばれるクラスは次を満たす.

$$\sqrt{N}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}^0) = \tfrac{1}{\sqrt{N}}\sum_{i=1}^{N}\boldsymbol{\varphi}(\boldsymbol{y}_i,\boldsymbol{\theta}) + o_p(1)$$

$o_p(1)$はオーダー1で確率収束する項, $\boldsymbol{\varphi}$を影響関数と呼ぶ. Asymptotically Linear Estimatorは一致性や漸近正規性を持つ. また, Hilbert空間における射影定理を用いて漸近分散を小さくすることを考える(Tsiatis. 2006). Double/debiased machine learning (Chernozhukov et al.2016)と呼ばれる手法もAsymptotically Linear Estimatorである.次のモデルを考える.

$$\boldsymbol{y} = \boldsymbol{g}(z,\boldsymbol{x}) + \boldsymbol{u}, \quad \mathrm{E}[\boldsymbol{u}|\boldsymbol{x},z] = 0$$
$$z = w(\boldsymbol{x}) + v, \quad \mathrm{E}[v|\boldsymbol{x}] = 0$$

興味のないパラメータ, 例えば傾向スコアwや関数$\boldsymbol{g}$にMLの手法を用いて, 興味のあるパラメータ, ここでは平均処置効果$\mathrm{E}[\boldsymbol{g}(1,\boldsymbol{x})-\boldsymbol{g}(0,\boldsymbol{x})]$を推定する. このようなモデルに対して, モデル選択を考える.

参考文献

- Chernozhukov, V. et al. (2016). Double/debiased machine learning for treatment and causal parameters. *The Econometrics Journal*, **21(1)**, C1-C68.
- Hoshino, T., Kurata, H., & Shigemasu, K. (2006). A propensity score adjustment for multiple group structural equation modeling. *Psychometrika*, **71**, 691–712.
- Hoshino, T. (2007). Doubly robust-type estimation for covariate adjustment in latent variable modeling. *Psychometrika*, **72**, 535-549.
- Tsiatis, A. A.(2006). *Semiparametric Theory and Missing Data*, Springer, New York.