2023 年 5 月 26 日 統計数理研究所 オープンハウス

傾向スコア解析における SURE 理論

二宮 嘉行　数理・推論研究系 教授

ノイズが条件付き正規分布にしたがうモデル

$$y_i = \sum_{h=1}^{H} t_i^{[h]} y_i^{[h]} = \sum_{h=1}^{H} t_i^{[h]} \{ \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\theta}^{[h]} + \eta(\boldsymbol{z}_i) + \varepsilon_i^{[h]} \} \qquad i \in \{1, 2, \ldots, N\}$$

- $t_i^{[h]} \in \{0, 1\}$：第 h 群の割り当て変数 $(\sum_{h=1}^{H} t_i^{[h]} = 1)$
- $y_i^{[h]} \in \mathbb{R}$：第 h 群の潜在結果変数（y_i：観測される結果変数）
- $\boldsymbol{z}_i \in \mathbb{R}^q$：共変量（説明変数 $\boldsymbol{x}_i \in \mathbb{R}^p$ や交絡変数を含む）
- $\boldsymbol{\theta}^{[h]} \in \mathbb{R}^p$：第 h 群のパラメータ
- $\eta(\cdot)$：未知の局外ノンパラメトリック関数
- $\varepsilon_i^{[h]} \sim \mathrm{N}(0, \sigma^2)$：第 h 群の誤差変数（$t_i^{[h]}$ や $\boldsymbol{z}_i$ とは独立）
- $(c^{[1]}, c^{[2]}, \ldots, c^{[H]})$ を $\sum_{h=1}^{H} c^{[h]} = 0$ と $\sum_{h=1}^{H} c^{[h]2} = 1$ を満たすコントラスト，$\boldsymbol{\theta} \equiv \sum_{h=1}^{H} c^{[h]} \boldsymbol{\theta}^{[h]}$ で定義される因果効果を推定対象とする

傾向スコア解析における通常の仮定

- $t_i^{[h]}$ と $y_i^{[h]}$ には相関があるので，観測値から $\boldsymbol{\theta}$ をナイーブに（最小二乗法とかで）推定すると，一般にバイアスが生じてしまう
- 条件付き無視可能性：$\{y_i^{[1]}, y_i^{[2]}, \ldots, y_i^{[H]}\} \perp\!\!\!\perp \{t_i^{[1]}, t_i^{[2]}, \ldots, t_i^{[H]}\} \mid \boldsymbol{z}_i$
- 正値性：$\mathrm{P}(t_i^{[h]} = 1) > 0$
- 独立性：$i \neq i'$ $(i, i' \in \{1, 2, \ldots, N\})$ に対して

$$(y_i^{[1]}, \ldots, y_i^{[H]}, t_i^{[1]}, \ldots, t_i^{[H]}, \boldsymbol{z}_i) \perp\!\!\!\perp (y_{i'}^{[1]}, \ldots, y_{i'}^{[H]}, t_{i'}^{[1]}, \ldots, t_{i'}^{[H]}, \boldsymbol{z}_{i'})$$

Inverse Probability Weighted 推定の紹介

- $y^{[h]}$ と $\boldsymbol{z}$ の関係をモデリングするのは一般に難しく，近年は，傾向スコア $e^{[h]}(\boldsymbol{z}) \equiv \mathrm{P}(t^{[h]} = 1 \mid \boldsymbol{z})$ を用い，一致性を保証するようなセミパラメトリック推定量 $\hat{\boldsymbol{\theta}}$ を与えることが標準となっている
- IPW（逆確率重み付け）推定：以下を解いたもの

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left[\sum_{i=1}^{N} \left\{ \sum_{h=1}^{H} \frac{c^{[h]} t_i^{[h]} y_i^{[h]}}{e^{[h]}(\boldsymbol{z}_i)} - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\theta} \right\}^2 \right] = \mathbf{0}_p$$

- 傾向スコア未知のケースや，DR（二重頑健）推定のケースは省略する

LASSO および SURE 理論の紹介

- 正則化パラメータ λ を適切に選ぶことが重要であり，交差検証法がしばしば用いられるものの，Efron et al. ('04 AS) や Zou et al. ('07 AS) は，SURE 理論を用いて弱点のほぼない情報量規準を開発した
- LASSO は微分不可能の点をもつ推定関数を用いているので，慣習的な Cp 基準を導くことはできないが，SURE 理論を用いれば，漸近理論に頼ることなく一般化 Cp 基準を導くことができる
- $\hat{p}$ を LASSO の非ゼロ推定量（アクティブセットの要素）数とすると

$$\mathrm{Cp} \equiv \sum_{i=1}^{N} \left(\sum_{h=1}^{H} c^{[h]} y_i^{[h]} - \boldsymbol{x}_i^{\mathrm{T}} \hat{\boldsymbol{\theta}}_\lambda \right)^2 + 2\hat{p}\sigma^2$$

は通常の平均二乗誤差から定数をひいたものであるリスクの不偏推定量であり，これを最小にする λ を最適値として選べばよい

- LASSO は「適応的に変数を選んでるので罰則は大きくなりそう」「縮小推定なので罰則は小さくなりそう」の二つがキャンセルし，ちょうど $\hat{p}$ が残るという驚きある結果である

引用文献

- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32, 407–451.
- Ninomiya, Y. (2022). Information criteria for sparse methods in causal inference. *arXiv preprint*, 2203.15308.
- Zou, H., Hastie, T., & Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. *The Annals of Statistics*, 35, 2173–2192.

スパース IPW 推定とリスクの提案

- 説明変数の候補数 p が大きめのときは，次の推定量を用いるのが自然

$$\hat{\boldsymbol{\theta}}_\lambda \equiv \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \left[\sum_{i=1}^{N} \left\{ \sum_{h=1}^{H} \frac{c^{[h]} t_i^{[h]} y_i^{[h]}}{e^{[h]}(\boldsymbol{z}_i)} - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\theta} \right\}^2 + \lambda \|\boldsymbol{\theta}\|_1 \right]$$

- 推定の際の損失関数を考慮し，情報量規準を導出するためのリスクとしては，次の平均二乗誤差を用いることを提案する

$$\mathrm{MSE} \equiv \mathrm{E}\left(\sum_{i=1}^{N} \left[\sum_{h=1}^{H} \mathrm{E}\left\{ \frac{c^{[h]} t_i^{[h]} y_i^{[h]}}{e^{[h]}(\boldsymbol{z}_i)} \,\middle|\, t_i^{[h]}, \boldsymbol{z}_i \right\} - \boldsymbol{x}_i^{\mathrm{T}} \hat{\boldsymbol{\theta}}_\lambda \right]^2 \right)$$

- $y^{[h]}$ は非正規なので SURE 理論は適用できないと思うかもしれないが，$t^{[h]}$ を先に条件付けるという，傾向スコア解析では通常考えないようなことをすると，条件付き正規性より SURE 理論を適用できる

スパース IPW 推定のための一般化 Cp 基準の導出

- 定理：$\hat{\boldsymbol{\theta}}_\lambda^{(2)} \equiv (\hat{\theta}_{\lambda,j})_{j:\hat{\theta}_{\lambda,j} \neq 0}$ や $\boldsymbol{x}_i^{(2)} \equiv (x_{i,j})_{j:\hat{\theta}_{\lambda,j} \neq 0}$ のように定義すれば，リスクの不偏推定量は以下で与えられる

$$\mathrm{IPCp} \equiv \sum_{i=1}^{N} \left\{ \sum_{h=1}^{H} \frac{c^{[h]} t_i^{[h]} y_i^{[h]}}{e^{[h]}(\boldsymbol{z}_i)} - \boldsymbol{x}_i^{(2)\mathrm{T}} \hat{\boldsymbol{\theta}}_\lambda^{(2)} \right\}^2 + 2\sigma^2 \mathrm{tr}\left[\left(\sum_{i=1}^{N} \boldsymbol{x}_i^{(2)} \boldsymbol{x}_i^{(2)\mathrm{T}} \right)^{-1} \left\{ \sum_{i=1}^{N} \sum_{h=1}^{H} \frac{c^{[h]2}}{e^{[h]}(\boldsymbol{z}_i)} \boldsymbol{x}_i^{(2)} \boldsymbol{x}_i^{(2)\mathrm{T}} \right\} \right]$$

- もし既存の情報量規準 QICw のアイディアをナイーブに用いると

$$\mathrm{QICw} \equiv \sum_{i=1}^{N} \left\{ \sum_{h=1}^{H} \frac{c^{[h]} t_i^{[h]} y_i^{[h]}}{e^{[h]}(\boldsymbol{z}_i)} - \boldsymbol{x}_i^{(2)\mathrm{T}} \hat{\boldsymbol{\theta}}_\lambda^{(2)} \right\}^2 + 2\sigma^2 \dim(\hat{\boldsymbol{\theta}}_\lambda^{(2)})$$

が得られ，QICw はスパース推定のためのものではないが，そう呼ぶことにする（罰則は IPCp より相当小さいものとなっている）

数値実験の設定

- $H = 2$ の適当なモデルにおいて推定対象は $(c^{[1]}, c^{[2]}) = (-1/\sqrt{2}, 1/\sqrt{2})$ で，パラメータの真値は $\boldsymbol{\theta}^{[1]} = -\boldsymbol{\theta}^{[2]} = (\theta^* \mathbf{1}_{p/2}{}^{\mathrm{T}}, \mathbf{0}_{p/2}{}^{\mathrm{T}})^{\mathrm{T}}$ で与える
- バイアス評価実験：$j \in \{1, 2, \ldots, 8\}$ の列は，非ゼロ推定量数が $\hat{p} = j \times p/8$ のバイアスから非ゼロ推定量数が $\hat{p} = (j-1) \times p/8$ のバイアスを引いたもの（QICw はこれらの値を $p/4$ と評価する）

$(p,\ N,\ \theta^*)$		1	2	3	4	5	6	7	8
(8, 40, 0.2)	True	3.36	3.94	3.90	4.18	4.42	5.32	3.90	5.71
	IPCp	2.90	3.95	3.88	4.07	4.92	4.86	4.68	5.41
(16, 40, 0.2)	True	3.39	3.82	4.24	4.08	4.99	4.65	4.04	6.39
	IPCp	2.98	4.08	3.78	4.39	4.59	4.72	4.14	5.90
(32, 40, 0.2)	True	5.59	5.80	6.51	7.98	7.92	11.14	12.22	10.90
	IPCp	5.80	6.19	6.54	8.29	7.94	11.25	11.93	11.32

- QICw と IPCp の性能比較実験：$\hat{p}_1$ と $\hat{p}_2$ は非ゼロ推定量とゼロ推定量の数，MSE は全パラメータに対するもの（括弧内の値は標準偏差）

$(p,\ \theta^*)$		$N = 40$			$N = 120$		
		$\hat{p}_1$	$\hat{p}_2$	$\sqrt{\mathrm{MSE}}$	$\hat{p}_1$	$\hat{p}_2$	$\sqrt{\mathrm{MSE}}$
(8, 0.2)	QICw	3.1	3.0	1.86 (0.67)	3.1	2.7	1.04 (0.32)
	IPCp	1.1	1.0	1.21 (0.60)	1.3	0.7	0.84 (0.22)
(16, 0.2)	QICw	7.0	7.0	3.27 (1.11)	6.8	6.2	1.53 (0.35)
	IPCp	4.3	3.9	2.32 (1.21)	2.5	1.6	1.19 (0.23)
(32, 0.4)	QICw	15.6	15.6	8.65 (3.07)	14.4	13.7	2.42 (0.48)
	IPCp	14.8	14.9	8.12 (3.21)	8.4	6.3	1.78 (0.38)