

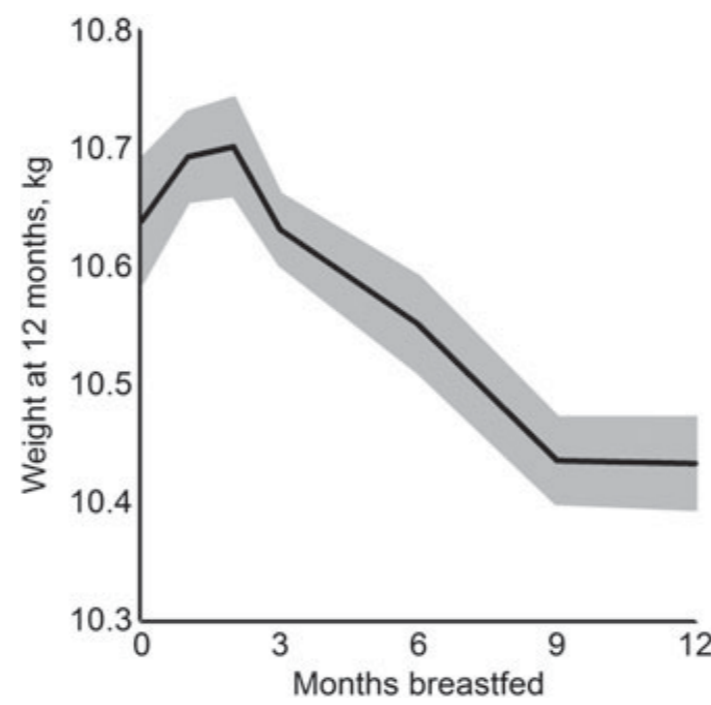
セミパラメトリック因果推論のための C_p 基準

二宮 嘉行 数理・推論研究系 教授

問題例

目的：右図において妥当な
回帰曲線（周辺構造）を得ること

- 横軸のどの値で観測されるかと縦軸の値には相関があり、最小二乗法では得られない
- 相関を考慮して求めた縦軸の値の期待値を結んだものが黒線



目的

因果推論における基本的なモデルである周辺構造モデルに対し、セミパラメトリックアプローチで通常用いられる IPW (inverse probability weight) 推定に対応できる情報量規準を開発すること

- Platt et al. '13 で提案された基準：

$$QIC_w = \sum_{h=1}^H \left\{ \sum_{i=1}^N \frac{t_i^{(h)}}{e_i^{(h)}} \left(\mathbf{y}_i - \mathbf{X}_i^{(h)} \hat{\beta}^{IPW} \right)^t \left(\mathbf{y}_i - \mathbf{X}_i^{(h)} \hat{\beta}^{IPW} \right) \right\} + 2\sigma^2 p$$

上記基準には数理的な妥当性がないため、リスク関数としてある MSE を定義し、その漸近不偏推定量、つまり漸近的な C_p 基準を導出する

モデル

周辺構造モデル (Robins '97)

$$\mathbf{y} = \sum_{h=1}^H t^{(h)} \mathbf{y}^{(h)} = \sum_{h=1}^H t^{(h)} (\mathbf{X}^{(h)} \beta + \varepsilon)$$

- 割り当て変数： $t^{(h)} \in \{0, 1\}$ ($h \in \{1, 2, \dots, H\}$, $\sum_{h=1}^H t_i^{(h)} = 1$)
- 結果変数： $\mathbf{y} \in \mathbb{R}^k$
- 説明変数： $\mathbf{X}^{(h)} \in \mathbb{R}^{k \times p}$ ($\mathbb{E}[\sum_{h=1}^H \mathbf{X}^{(h)t} \mathbf{X}^{(h)}] = \mathbf{I}_p$)
- 周辺構造に関するパラメータ： $\beta \in \mathbb{R}^p$
- 誤差項： $\varepsilon \sim (\mathbf{0}, \sigma^2 \mathbf{I}_k)$

観測値と欠測値の間には偏りがあり、つまり $t^{(h)}$ と $\mathbf{y}^{(h)}$ には相関があり、最小二乗法では一致性をもつ推定量は得られないことを想定

仮定

弱く無視できる割り当て条件

$$\mathbf{y}^{(h)} \perp\!\!\!\perp t^{(h)} \mid \mathbf{z} \quad (\mathbf{z} \in \mathbb{R}^s : \text{交絡変数})$$

サンプル用の表記（添え字はサンプルの番号）

$$\tilde{\mathbf{y}} = \sum_{h=1}^H \tilde{\mathbf{T}}^{(h)} \tilde{\mathbf{y}}^{(h)} = \sum_{h=1}^H \tilde{\mathbf{T}}^{(h)} \left(\tilde{\mathbf{X}}^{(h)} \beta + \tilde{\varepsilon} \right)$$

- $\tilde{\mathbf{T}}^{(h)} = \text{diag}\{t_1^{(h)} \mathbf{I}_k, t_2^{(h)} \mathbf{I}_k, \dots, t_N^{(h)} \mathbf{I}_k\}$
- $\tilde{\mathbf{y}} = (\mathbf{y}_1^t, \mathbf{y}_2^t, \dots, \mathbf{y}_N^t)^t$, $\tilde{\mathbf{y}}^{(h)} = (\mathbf{y}_1^{(h)t}, \mathbf{y}_2^{(h)t}, \dots, \mathbf{y}_N^{(h)t})^t$
- $\tilde{\mathbf{X}}^{(h)} = (\mathbf{X}_1^{(h)t}, \mathbf{X}_2^{(h)t}, \dots, \mathbf{X}_N^{(h)t})^t$
- $\tilde{\varepsilon} = (\varepsilon_1^t, \varepsilon_2^t, \dots, \varepsilon_N^t)^t$

仮定

- $(\varepsilon_i, t_i^{(h)}, z_i, \mathbf{X}_i^{(h)}) \perp\!\!\!\perp (\varepsilon_j, t_j^{(h)}, z_j, \mathbf{X}_j^{(h)})$ ($i \neq j$, $h \in \{1, 2, \dots, H\}$)
- $\mathbf{X}_i^{(h)}$ と ε_i は無相関 ($h \in \{1, 2, \dots, H\}$)

IPW推定量

Inverse probability weighted 推定量 (Rubin '85)

$$\hat{\beta}^{IPW} = \left(\sum_{h=1}^H \tilde{\mathbf{X}}^{(h)t} \tilde{\mathbf{W}}^{(h)} \tilde{\mathbf{X}}^{(h)} \right)^{-1} \sum_{h=1}^H \tilde{\mathbf{X}}^{(h)t} \tilde{\mathbf{W}}^{(h)} \tilde{\mathbf{y}}$$

- $\tilde{\mathbf{W}}^{(h)} = \text{diag}\{t_1^{(h)} \mathbf{I}_k / e_1^{(h)}, t_2^{(h)} \mathbf{I}_k / e_2^{(h)}, \dots, t_N^{(h)} \mathbf{I}_k / e_N^{(h)}\}$
- 一般化傾向スコア： $e_i^{(h)} = P(t_i^{(h)} = 1 \mid z_i)$
- 弱く無視できる割り当て条件のもと一致性を持つ

MwSE とそれに基づく C_p 基準

Mean weighted squared error

$$\sum_{h=1}^H \mathbb{E} \left[\left(\tilde{\mathbf{X}}^{(h)} \hat{\beta} - \mathbb{E}[\tilde{\mathbf{y}}^{(h)} \mid \tilde{\mathbf{X}}^{(h)}] \right)^t \tilde{\mathbf{W}}^{(h)} \left(\tilde{\mathbf{X}}^{(h)} \hat{\beta} - \mathbb{E}[\tilde{\mathbf{y}}^{(h)} \mid \tilde{\mathbf{X}}^{(h)}] \right) \right]$$

- 偏りなくデータが得られたときの二乗誤差を考えていることに相当

MwSE の漸近評価 (Baba et al. '17)

$$wC_p = \sum_{h=1}^H \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}^{(h)} \hat{\beta}^{IPW}(\alpha) \right)^t \tilde{\mathbf{W}}^{(h)} \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}^{(h)} \hat{\beta}^{IPW}(\alpha) \right) + 2 \sum_{h=1}^H \mathbb{E} \left[\frac{1}{e^{(h)}} \varepsilon^t \mathbf{X}^{(h)} \mathbf{X}^{(h)t} \varepsilon \right]$$

数値実験

Platt et al. '13 の例の設定で $\mathbb{E}[y^{(h)}] = \sum_{\ell=0}^k \beta_{k,\ell} \tau^{(h)\ell}$ という k 次多項式モデルを考え、この k の選択問題を扱う

- $t^{(h)} = 1$ という割り当てだと $\tau^{(h)}$ なる横軸の値でデータが観測される
- $\beta_{2,0} = \beta_{2,1} = 1$, $\beta_{2,2} = 0.1/0.3/0.5$ という 2 次多項式モデルを真とし、100/200 サンプルにおいて 5000 回のシミュレーションをおこなう

		モデル選択確率					
$\beta_{2,2}$	N	基準	MwSE	0-1次	2次	3次	4-5次
0.5	100	wC_p	62.55	0.00	68.52	14.40	17.08
		QIC_w	81.00	0.00	10.62	12.36	77.02
	200	wC_p	62.35	0.00	72.20	13.92	13.88
		QIC_w	82.41	0.00	11.24	12.76	76.00
0.3	100	wC_p	62.57	0.02	68.50	14.40	17.08
		QIC_w	81.00	0.00	10.62	12.36	77.02
	200	wC_p	62.35	0.00	72.20	13.92	13.88
		QIC_w	82.41	0.00	11.24	12.76	76.00
0.1	100	wC_p	65.38	30.60	42.36	12.22	14.82
		QIC_w	81.05	2.20	9.10	12.02	76.68
	200	wC_p	66.90	15.64	58.54	12.94	12.88
		QIC_w	82.41	0.72	10.84	12.66	75.78

引用文献

- Baba, T., Kanemori, T. and Ninomiya, Y. (2017). Biometrika, 104, 845-861.
 Platt, R. W., Brookhart, M. A., Cole, S. R., Westreich, D. and Schisterman, E. F. (2013). Statistics in Medicine, 32, 1383-1393.
 Robins, J. M. (1997). Proceedings of the Section on Bayesian Statistical Science. American Statistical Association, 1-10.
 Rubin, D. B. (1985). Bayesian statistics, 2, 463-472.

本内容は、馬場崇充氏（塩野義製薬解析センター）、金盛誠之氏（トリア再保険）との共同研究に基づく成果である