2020年10月27日 統計数理研究所 オープンハウス

スパース推定に対するcAICの開発と選択的推論への貢献

小松 慧 総合研究大学院大学 5年一貫制(4年) skomatsu@ism.ac.jp

Elastic netに対する自由度の導出

線形回帰モデルにおいてのelastic net推定量は

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}} = (1+\lambda_2)\arg\min_{\boldsymbol{\beta}}\left[\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2\right]$$

であり, $\hat{\boldsymbol{\mu}}_{\boldsymbol{\lambda}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}}$の自由度$\rho$は, SURE理論を用いたZou *et al.* (2007)の話を拡張すると,

$$\rho = \mathrm{tr}\left(\frac{\partial\hat{\boldsymbol{\mu}}_{\boldsymbol{\lambda}}}{\partial\boldsymbol{y}}\right) = (1+\lambda_2)\mathrm{tr}\left((\boldsymbol{X}_{\mathcal{A}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{A}} + \lambda_2\boldsymbol{I})^{-1}\boldsymbol{X}_{\mathcal{A}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{A}}\right)$$

と求まる.ただし, $\mathcal{A} = \{j \mid \hat{\beta}_{\boldsymbol{\lambda},j} \neq 0\}$

Elastic netに対する条件付きAICの開発

以下の表記法はVaida & Blanchard (2005)に準じる.
混合効果モデルにおいてのelastic net推定量は

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}} = (1+\lambda_2)\arg\min_{\boldsymbol{\beta}}\left[\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\|_{\boldsymbol{\Sigma}} + \lambda_1\|\boldsymbol{\beta}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_2^2\right]$$

となり,経験ベイズ推定量$\hat{\boldsymbol{b}}_{\boldsymbol{\lambda}} = (\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} + \sigma^2\boldsymbol{G}_0^{-1})^{-1}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}})$を用いると, $\hat{\boldsymbol{\mu}}_{\boldsymbol{\lambda}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}} + \boldsymbol{Z}\hat{\boldsymbol{b}}_{\boldsymbol{\lambda}}$は ($\mathcal{A} = \{j \mid \hat{\beta}_{\boldsymbol{\lambda},j} \neq 0\}$)

$$\left[(1+\lambda_2)(\boldsymbol{X}_{\mathcal{A}}\ \boldsymbol{Z})\boldsymbol{M}^{-1}\begin{pmatrix}\boldsymbol{X}_{\mathcal{A}}^{\mathrm{T}}\\ \boldsymbol{Z}^{\mathrm{T}}\end{pmatrix} - \lambda_2\boldsymbol{Z}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} + \sigma^2\boldsymbol{G}_0^{-1})^{-1}\boldsymbol{Z}^{\mathrm{T}}\right]\boldsymbol{y}$$

となる.ただし, $\boldsymbol{M} = \begin{pmatrix}\boldsymbol{X}_{\mathcal{A}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{A}} + \sigma^2\lambda_2\boldsymbol{I} & \boldsymbol{X}_{\mathcal{A}}^{\mathrm{T}}\boldsymbol{Z}\\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{A}} & \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} + \sigma^2\boldsymbol{G}_0^{-1}\end{pmatrix}$.
ここで,上述の導出が適用できることを用いると, $\hat{\boldsymbol{\mu}}_{\boldsymbol{\lambda}}$の条件付きAICにおける自由度$\rho$は

$$\rho = (1+\lambda_2)\mathrm{tr}\left[\boldsymbol{M}^{-1}\boldsymbol{M}_0\right] - \lambda_2\mathrm{tr}[(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} + \sigma^2\boldsymbol{G}_0^{-1})^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}]$$

と求まり, elastic net推定量に対する条件付きAICは

$$\mathrm{cAIC}^{\mathrm{ela}} = -2\log f(\boldsymbol{y} \mid \hat{\boldsymbol{b}}_{\boldsymbol{\lambda}}; \hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}}) + 2\rho$$

と導出される.ただし, $\boldsymbol{M}_0 = \begin{pmatrix}\boldsymbol{X}_{\mathcal{A}}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{A}} & \boldsymbol{X}_{\mathcal{A}}^{\mathrm{T}}\boldsymbol{Z}\\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{X}_{\mathcal{A}} & \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\end{pmatrix}$.

Selective inferenceへの貢献

- Lee *et al.* (2016)
  - 正則化パラメータを固定して,線形回帰モデルにおけるLASSOに対するselective inferenceを構築.
  - 他の正則化法に応用可能だが固定の仕方に恣意性あり.
- 本研究
  - 正則化パラメータを選択したときの混合効果モデルにおけるelastic netに対するselective inferenceを構築.
  - $\boldsymbol{y}$が変化するごとに選択される正則化パラメータも変化するため,逐一selective eventに入るか調べる必要があるが,そのプログラムは1次元に落とし込め,開発可能.
  - 本提案は,一般のselective inferenceにも適用可能である.

数値実験(リスク評価)

$\boldsymbol{\beta}_j$については0か否か, b_hが全てのグループで同一か異なるか,のモデルの比較を行った.

- b_hが全てのグループで同一→$b_1 = b_2 = \cdots = b_m \sim \mathcal{N}(0,\tau^2)$
- b_hが全てのグループで異なる→$b_1, b_2, \ldots, b_m \overset{iid}{\sim} \mathcal{N}(0,\tau^2)$

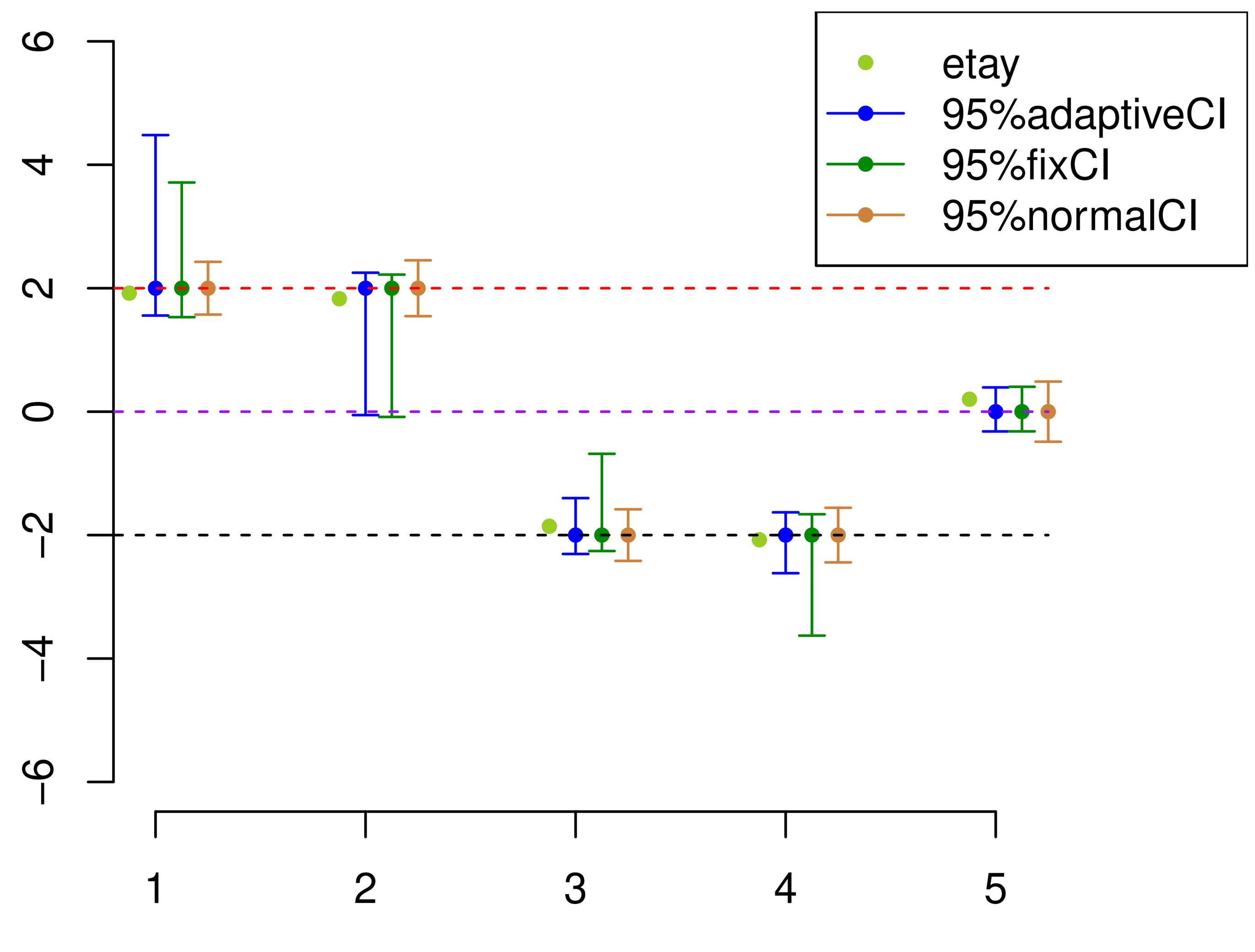
他の設定は以下である.

- $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{b} + \boldsymbol{\varepsilon}, \boldsymbol{b} = (b_h\boldsymbol{1}_n)_{h=1,\ldots,m}, \boldsymbol{\varepsilon} \sim \mathcal{N}_{mn}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$
- $\boldsymbol{X}$: 各成分は$\mathcal{N}(0,1)$の実現値, $\boldsymbol{\beta} = (\beta_1, \beta_1, \beta_2, \beta_2, 0, 0)$
- [1]→b_hが全てのグループで同一, [2]→全てのグループで別

$(\tau^2, \beta_1, \beta_2)$	[1]		[2]	
	cAIC	mAIC	cAIC	mAIC
$(1, 0.5, -0.5)$	12.68	31.69	12.18	19.55
$(1, 0.5, 0.1)$	8.34	30.77	7.28	17.42
$(1, 0.1, -0.1)$	6.86	6.06	3.77	2.86
$(2, 0.5, -0.5)$	13.74	32.15	8.50	31.49
$(2, 0.5, 0.1)$	13.14	20.31	7.37	17.61
$(2, 0.1, -0.1)$	9.93	9.09	3.91	2.70

数値実験(信頼区間)

本研究によって得た信頼区間(adaptive CI), Lee *et al.* (2016)によって得られた信頼区間(fix CI),通常の信頼区間(normal CI)の比較を行い,従来の手法との明らかな違いが示された.



参考文献

- Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. The Annals of Statistics, 44(3), 907-927.
- Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. Biometrika, 92(2), 351-370.
- Zou, H., Hastie, T., & Tibshirani, R. (2007). On the "degrees of freedom" of the lasso. The Annals of Statistics, 35(5), 2173-2192.