

スパース推定と条件付きAIC、selective inferenceの融合

小松 慧

総合研究大学院大学 5年一貫制(5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

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

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

$$\rho = \text{tr} \left(\frac{\partial \hat{\mu}_\lambda}{\partial \mathbf{y}} \right) = (1 + \lambda_2) \text{tr} \left((\mathbf{X}_A^T \mathbf{X}_A + \lambda_2 \mathbf{I})^{-1} \mathbf{X}_A^T \mathbf{X}_A \right)$$

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

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

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

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

$$\left[(1 + \lambda_2) (\mathbf{X}_A^T \mathbf{Z}) \mathbf{M}^{-1} \begin{pmatrix} \mathbf{X}_A^T \\ \mathbf{Z}^T \end{pmatrix} - \lambda_2 \mathbf{Z} (\mathbf{Z}^T \mathbf{Z} + \sigma^2 \mathbf{G}_0^{-1})^{-1} \mathbf{Z}^T \right] \mathbf{y}$$

となる. ただし, $\mathbf{M} = \begin{pmatrix} \mathbf{X}_A^T \mathbf{X}_A + \sigma^2 \lambda_2 \mathbf{I} & \mathbf{X}_A^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X}_A & \mathbf{Z}^T \mathbf{Z} + \sigma^2 \mathbf{G}_0^{-1} \end{pmatrix}$.

ここで, 上述の導出が適用できることを用いると, $\hat{\mu}_\lambda$ の条件付きAICにおける自由度 ρ は

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

と求まり, elastic net推定量に対する条件付きAICは

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

と導出される. ただし, $\mathbf{M}_0 = \begin{pmatrix} \mathbf{X}_A^T \mathbf{X}_A & \mathbf{X}_A^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X}_A & \mathbf{Z}^T \mathbf{Z} \end{pmatrix}$.

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.
 - \mathbf{y} が変化するととも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次元に落とし込み, 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)

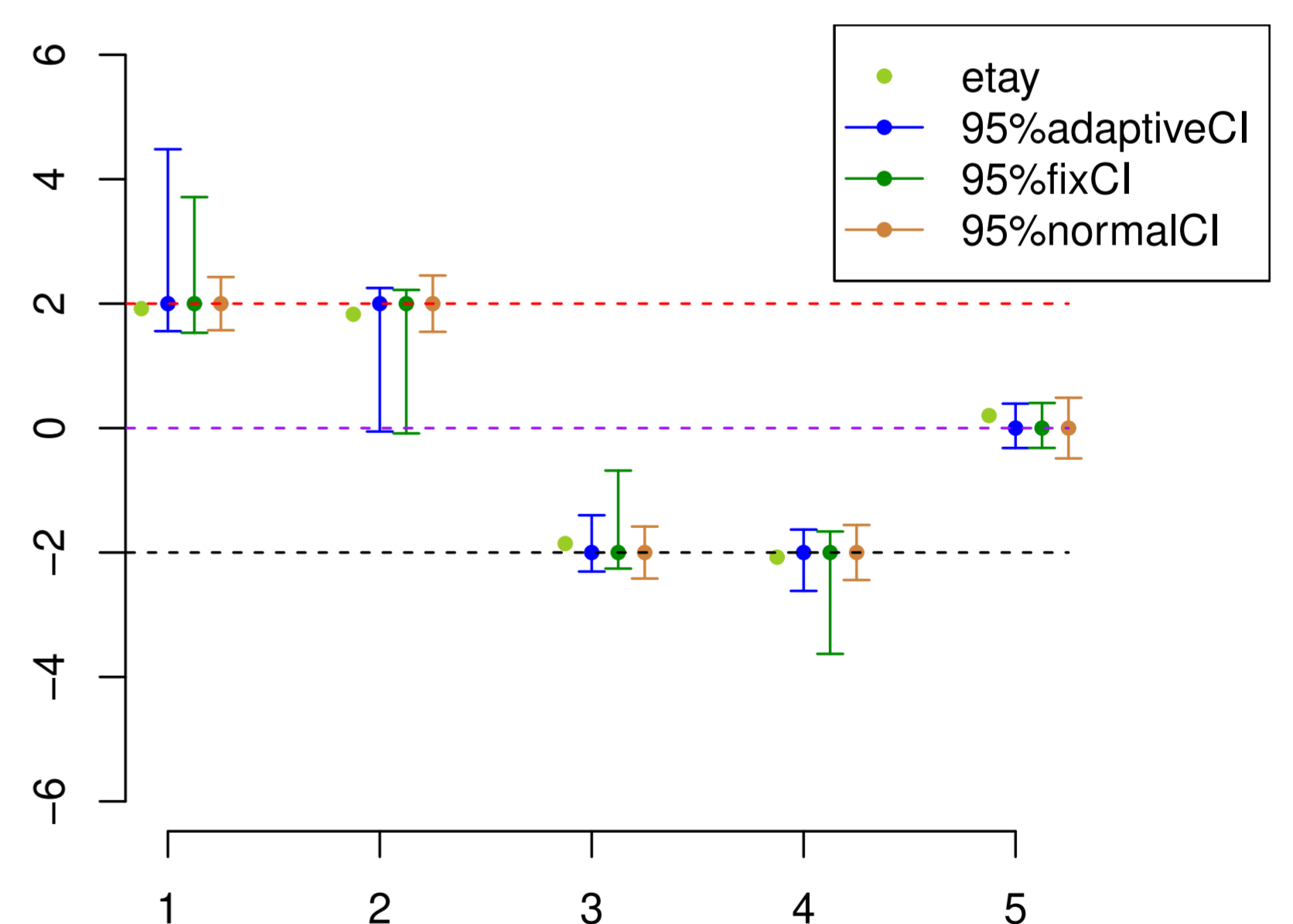
β_j については0か否か, b_h が全てのグループで同一か異なるか, のモデルの比較を行った.

- b_h が全てのグループで同一 $\rightarrow b_1 = b_2 = \dots = b_m \sim \mathcal{N}(0, \tau^2)$
 - b_h が全てのグループで異なる $\rightarrow b_1, b_2, \dots, b_m \stackrel{iid}{\sim} \mathcal{N}(0, \tau^2)$
- 他の設定は以下である.
- $\mathbf{y} = \mathbf{X}\beta + \mathbf{b} + \varepsilon$, $\mathbf{b} = (b_h \mathbf{1}_n)_{h=1, \dots, m}$, $\varepsilon \sim \mathcal{N}_{mn}(\mathbf{0}, \sigma^2 \mathbf{I})$
 - \mathbf{X} : 各成分は $\mathcal{N}(0, 1)$ の実現値, $\beta = (\beta_1, \beta_1, \beta_2, \beta_2, 0, 0)$
 - [1] $\rightarrow b_h$ が全てのグループで同一, [2] \rightarrow 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.