

スパース推定に対する条件付き AIC の導出

九州大学 大学院数理学府 小松 慧
統計数理研究所 数理・推論研究系 二宮 嘉行

概要: 混合効果モデルに対する情報量規準として、一般に条件付き AIC (Vaida & Blanchard, 2005, Biometrika) が用いられており、近年では線形回帰モデルに適用されていたスパース推定法が混合効果モデルに対しても適用されている。本講演では混合効果モデルにおける固定効果項とランダム効果項の両方に対し、スパース推定に基づくモデル選択のための条件付き AIC、特にランダム効果項の分散共分散行列に関するパラメータが既知という仮定の下での lasso (Tibshirani, 1996, JRSSB) に対する条件付き AIC を導出する。混合効果モデルに対するスパース推定は、Bondell et al. (2010, Biometrics) などで扱われているが、それらがランダム効果項の分散共分散行列に対する罰則を考えているのに対し、本講演ではランダム効果項自体に対する罰則を考えている。

サンプル i に対し、 $\mathbf{y}_i$ を目的変数ベクトル、 $\mathbf{X}_i$ と $\mathbf{Z}_i$ をそれぞれ非ランダム係数 $\boldsymbol{\beta}$ とランダム係数 $\mathbf{b}_i$ が付く説明変数行列、 $\boldsymbol{\varepsilon}_i$ を誤差変数ベクトルとし、 $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i$ というモデルを以下では考える ($1 \leq i \leq n$)。ここで、 $\mathbf{b}_i$ と $\boldsymbol{\varepsilon}_i$ はそれぞれ $N(\mathbf{0}, \mathbf{G})$ と $N(\mathbf{0}, \sigma^2 \mathbf{I})$ に独立に従うとする。ただし、 $\mathbf{0}$ と $\mathbf{I}$ はそれぞれ適切なサイズの零ベクトルと単位行列とする。このとき、 $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)^T$, $\mathbf{X} = (\mathbf{X}_1^T, \dots, \mathbf{X}_n^T)^T$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_n)$, $\mathbf{b} = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)^T$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^T, \dots, \boldsymbol{\varepsilon}_n^T)^T$ とすれば、このモデルは $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$ とかける。ただし、 $\mathbf{b}$ と $\boldsymbol{\varepsilon}$ はそれぞれ $N(\mathbf{0}, \mathbf{G} \otimes \mathbf{I})$ と $N(\mathbf{0}, \sigma^2 \mathbf{I})$ に独立に従う。

Hodges & Sargent (2001, Biometrika) の方式を用いると、 $\boldsymbol{\beta}$ の MLE と $\mathbf{b}$ の BLUP は重み付き二乗誤差の最小解として得られるが、それに L_1 罰則をつけたものの最小解

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = \arg \min_{\boldsymbol{\beta}, \mathbf{b}} \left[\left\| \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{O} & -\sigma^2 \mathbf{D} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{pmatrix} \right\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 + \eta \|\mathbf{b}\|_1 \right]$$

を考える。ここで、 $\mathbf{O}$ は適切なサイズの零行列、 $\mathbf{D}$ は $\mathbf{D}^T \mathbf{D} = \mathbf{G}^{-1} \otimes \mathbf{I}$ をみたす行列である。正則化パラメータである λ と η の妥当な値を与えるための条件付き AIC のペナルティ項は

$$2 \text{tr} \left[\begin{pmatrix} \mathbf{X}_A^T \mathbf{X}_A & \mathbf{X}_A^T \mathbf{Z}_B \\ \mathbf{Z}_B^T \mathbf{X}_A & \mathbf{Z}_B^T \mathbf{Z}_B + \sigma^2 \mathbf{D}_B^T \mathbf{D}_B \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_A^T \mathbf{X}_A & \mathbf{X}_A^T \mathbf{Z}_B \\ \mathbf{Z}_B^T \mathbf{X}_A & \mathbf{Z}_B^T \mathbf{Z}_B \end{pmatrix} \right] = 2|\mathcal{A}| + 2 \sum_{i=1}^{r_0} \frac{\gamma_i}{1 + \gamma_i}$$

で与えられる。

ここで、 $\mathcal{A}$ と $\mathcal{B}$ は $\boldsymbol{\beta}$ と $\mathbf{b}$ のアクティブセットであり、 $\mathbf{X}_A$ は $\mathcal{A}$ に対応する列ベクトルを取り出してできる行列。 $\mathbf{Z}_B$, $\mathbf{D}_B$ も同様の手法で出来た行列である。

また、 γ_i ($i = 1, \dots, r_0$) は $\mathbf{D}^{*1/2} \mathbf{Z}_B^T (\mathbf{I} - \mathbf{X}_A (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \mathbf{X}_A^T) \mathbf{Z}_B \mathbf{D}^{*1/2}$ の非ゼロ固有値である。ただし、 $\mathbf{D}^* = \sigma^{-2} ((\mathbf{G} \otimes \mathbf{I})_{jk})_{j \in \mathcal{B}, k \in \mathcal{B}}$ である。 $\mathcal{A}$ と $\mathcal{B}$ がフルセットならば、このペナルティ項はそれぞれ Vaida & Blanchard (2005, Biometrika) と Liang et al. (2008, Biometrika) のそれと一致する。