ランダム係数を持つGMANOVAモデルの変数選択規準

広島大　理　　若木　宏文

$Y_{ij}^{(l)}$ $(l = 1, \ldots, g; i = 1, \ldots, n_l; j = 1, \ldots, p)$ を第 l 群に属する個体 i の t_j 時点での観測値とし, 次の成長曲線モデルを考える.

$$Y_{ij}^{(l)} = b_i^{(l)} + \sum_{k=1}^{q} \beta_k^{(l)} x_k(t_j) + \varepsilon_{ij}^{(l)}$$

ここで, $x_k(t)$ $(k = 1, \ldots, q)$ は基底関数, $b_i^{(l)}$ は個体変動を表わすランダム係数, $\varepsilon_{ij}^{(l)}$ は観測誤差で,

$$\begin{aligned} &b_i^{(l)} \ (i = 1, \ldots, n_l; l = 1, \ldots, g) \overset{i.i.d.}{\sim} N(\beta_0^{(l)}), \tau^2, \\ &\varepsilon_{ij}^{(l)} \ (l = 1, \ldots, g; i = 1, \ldots, n_l; j = 1, \ldots, p) \overset{i.i.d.}{\sim} N(0, \sigma^2) \end{aligned}$$

とする.

$\boldsymbol{\theta}$ をすべての未知母数を並べたベクトル, $\mathbf{Y}$ をすべての観測変量を並べた確率行列とし, $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ を $\mathbf{Y}$ による最尤推定量とすると, カルバック-ライブラーダイヴァージェンスに基づくモデルの良さは, $\mathbf{Z}$ を $\mathbf{Y}$ と独立に同じ分布に従う確率行列として

$$-2\mathrm{E}_{\mathbf{Y}}\mathrm{E}_{\mathbf{Z}}[\log f(\mathbf{Z}; \hat{\boldsymbol{\theta}}(\mathbf{Y}))] \tag{1}$$

で与えられる. ここで, $f(\mathbf{Y}, \boldsymbol{\theta})$ は, $\mathbf{Y}$ の同時確率密度関数である. AIC 規準は, (1) を $(-2)\times$ 最大対数尤度

$$\log f(\mathbf{Y}; \hat{\boldsymbol{\theta}}(\mathbf{Y}))$$

によって推定するときの, 漸近バイアスを補正することによって与えられるが, τ^2 の値が 0 に近いとき, $\hat{\boldsymbol{\theta}}$ はパラメータ空間の境界上に値をとり, 尤度方程式の解とならないことがある. この場合, 最大対数尤度によるリスクの推定量の漸近バイアスは未知パラメータ数の 2 倍とはならない.

本報告では, ラプラス近似を用いて漸近バイアスを評価することにより, ランダム係数を持つ GMANOVA モデルに対する変数選択規準を導出した.