2005年度統計関連学会連合大会
2005年9月12日～15日 (広島)

多変量逆回帰モデルにおける変数選択法の漸近的性質と改良

東京学芸大学・教育　横山　隆久
福山暁の星女子高校　美馬　芳江
中央大学・理工　藤越　康祝

p 次元目的変数 $\boldsymbol{y} = (y_1, \ldots, y_p)'$ と q 次元説明変数 $\boldsymbol{x} = (x_1, \ldots, x_q)'$ が N 個の個体について観測され, これらの初期標本を

$$Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)', \quad X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)'$$

とする. Y について正規性を仮定するが, 一般に

$$M^* : Y \sim N_{N\times p}(\eta^*,\ \Sigma^* \otimes I_N)$$

であるとする. ここで, $\eta^* : N \times p,\ \Sigma^* : p \times p$ は真の平均, 共分散行列である. 多変量逆推定問題では, 新たな観測値 $\boldsymbol{y} = (y_1, \ldots, y_p)'$ が得られたとき, 対応する $\boldsymbol{x}$ は未知であるとして, これを推測することを問題にする. ここでは, Y と $\boldsymbol{y}$ に同じ線形回帰構造

$$\begin{aligned} &Y \sim N_{N\times p}(\mathbf{1}_N \boldsymbol{\alpha}' + XB,\ \Sigma \otimes I_N), \\ &\boldsymbol{y} \sim N_p(\boldsymbol{\alpha} + B\boldsymbol{x}, \Sigma) \end{aligned}$$

を想定する. このとき, $\boldsymbol{x}$ の推定量として, 初期標本によりパラメータの最尤推定量 $\hat{\Sigma},\ \hat{\boldsymbol{\alpha}},\ \hat{B}$ を求め,

$$\hat{\boldsymbol{x}}_c = (\hat{B}\ \hat{\Sigma}^{-1} \hat{B}')^{-1} \hat{B}\ \hat{\Sigma}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\alpha}})$$

と推定する伝統的な推定量や, 逆推定量などが用いられる. このような多変量逆回帰推定問題において, $y_1, \ldots, y_p$ に関する変数選択問題を考える.

一つのアプローチは, たとえば, 変数の組 $\{y_1, \ldots, y_k\}$ が十分で, 残りの変数の組 $\{y_{k+1}, \ldots, y_p\}$ は追加情報をもたないというモデル

$$M_{1\cdots k}; B_2 = B_1 \Sigma_{11}^{-1} \Sigma_{12}$$

を導入し, AIC 基準（Fujikoshi and Nishii (1986)）や C_p 基準などを用いることである. ここに, $B = (B_1, B_2), B_; q \times k$ で, Σ_{ij} は Σ の分割行列である. 他の変数選択法として, 逆回帰推定量の平均 2 乗誤差の推定量を構成することも考えられる. 本報告では, これらの変数選択法の漸近的性質を調べ, 改良基準を提案する.

参考文献

1. Fujikoshi, Y. and Nishii, R. (1986). Selection of variables in multivariate inverse regression. *Hiroshima Math.J.*, **13**, 269-277.