正準判別分析における一致性を持つ高次元変数の選択法

広島大・理　鈴木 裕也　小田 凌也　柳原 宏和　藤越 康祝

本発表では, p 変量正規分布 $N_p(\boldsymbol{\mu}^{(i)}, \boldsymbol{\Sigma})$ $(i = 1, \ldots, q+1)$ からなる $(q+1)$ 群の正準判別分析における変数選択法の高次元大標本漸近性質を扱う．いま, $\omega = \{1, 2, \ldots, p\}$ とし, j を ω の部分集合とする．集合 j は変数 $\boldsymbol{x}$ の添え字の集合に用いられる．ここで, $\boldsymbol{x}_j$ を j に対応した $\boldsymbol{x}$ の成分を並べたベクトルとし, $\boldsymbol{x} = (\boldsymbol{x}'_j, \boldsymbol{x}'_{\bar{j}})'$ と書き, $m = \min(p, q)$ 個の線形判別関数を $\boldsymbol{\beta}'_k \boldsymbol{x}$ $(k = 1, \ldots, m)$ とする．$\boldsymbol{x}_{\bar{j}}$ が冗長であるとは, $\boldsymbol{x}_{\bar{j}}$ に対応する m 個の線形判別関数の係数がすべて 0 となることであり, そのような変数は判別能力をもたないために解析から取り除くことが望ましい．冗長な変数を取り除くことは変数選択問題とみなされ, そのような変数選択の手法の 1 つとして以下のような AIC による全ての変数の組み合わせに対して変数選択規準の計算を行い最適な変数の組み合わせ $\hat{j}$ を求める総当たり法が挙げられる:

$$\hat{j} = \arg\min_{j \subset \omega} \mathrm{AIC}(j). \tag{1}$$

近年, 計算機の発展に伴い高次元データを扱う需要が増えている．特に, 変数の次元数 p は大きいが標本数 n よりは小さいデータを考える．そのような高次元データにおいて (1) 式のような総当たり法は変数の組み合わせが $2^p - 1$ 個もあるため実行は物理的に不可能である．そこで, 本発表では AIC の罰則項を正数 $\alpha \times$ (独立パラメータ数) で置き換えた一般化情報量規準 (Generalized Information Criterion; GIC) を用い, Zhao *et al.* (1986) で提案された以下のような変数選択法を適用する:

$$\hat{j} = \{\ell \in \omega \mid \mathrm{GIC}(\omega \backslash \{\ell\}) > \mathrm{GIC}(\omega)\}. \tag{2}$$

(2) 内の不等式は, $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kp})'$ とおくとき, 仮説: $\beta_{1\ell} = \cdots = \beta_{m\ell} = 0$ の棄却域と関係している．(2) 式による選択法により, 扱う変数の組み合わせを $2^p - 1$ 個から p 個まで減らすことができる．Fujikoshi & Sakurai (2018) では 2 群判別問題に対して (2) 式を適用し GIC が一致性をもつための α に関する十分条件を求めている．一致性とは, 最適な変数の組み合わせ $\hat{j}$ が真の変数の組み合わせ j_* となる確率が 1 に収束する性質, すなわち, $\Pr(\hat{j} = j_*) \to 1$ が成り立つことであり, 変数選択法の良さを評価する 1 つの指標である．本発表ではこれらの結果を多群に拡張し, n, p をともに無限大であるが p は n より小さいという高次元大標本漸近理論: $(n, p) \to \infty$, $p/n \to c \in [0, 1)$ を用いて, (2) 式の変数選択法の下で GIC が一致性をもつための α に関する十分条件を求める．発表当日は, 一致性をもつ様々な α を数値的に比較する．

参考文献

[1] Fujikoshi, Y. & Sakurai, T. (2018). Consistency of test-based criterion for selection of variables in high-dimensional two group-discriminant analysis. TR No. 18–04, *Statistical Research Group*, Hiroshima University.

[2] Zhao, L. C., Krishnaiah, P. R. & Bai, Z. D. (1986). On detection of the number of signals in presence of white noise. *J. Multivariate Anal.*, **20**, 1–25.