2次判別関数に関する高次元漸近近似の誤差限界

広島大・理・名誉教授　藤越　康祝

本研究では, 同一の共分散行列をもつ2群の判別における2次判別関数 Z の高次元大標本漸近分布の誤差限界を導出する. p 次元変量 $\boldsymbol{x}$ の2群 $\Pi_i, i=1,2$ における母集団分布として, 平均 $\boldsymbol{\mu}_i$, 共分散行列 $\boldsymbol{\Sigma}$ の p 変量正規分布 $\mathrm{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ を仮定する. 各群から抽出された N_i 個の独立な観測ベクトル $\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,N_i}$ を用いて, 観測値ベクトル $\boldsymbol{x}$ を判別する規準として, 線形および2次判別関数があるが, ここでは次の2次判別関数を考える.

$$Z = \frac{1}{2(1+N_2^{-1})}(\boldsymbol{x}-\bar{\boldsymbol{x}}_2)'\mathbf{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_2) - \frac{1}{2(1+N_2^{-1})}(\boldsymbol{x}-\bar{\boldsymbol{x}}_1)'\mathbf{S}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}_1).$$

ここに, $\bar{\boldsymbol{x}}_i, i=1,2$ は各群の標本平均ベクトル, $\mathbf{S}$ は不偏合併共分散行列である. 判別方式は, $Z > c$ ならば, $\boldsymbol{x}$ を Π_1 に属するとし, $Z \le c$ ならば, $\boldsymbol{x}$ を Π_2 に属するとする. このとき, 2種類の誤判別確率

$$e(2|1) = P(Z \le c \mid \boldsymbol{x} \in \Pi_1), \quad e(2|1) = P(Z \ge c \mid \boldsymbol{x} \in \Pi_2)$$

の評価が重要となる. $e(1|2)$ は $e(2|1)$ から求められるので, 以下では, $e(1|2)$ の一般化である $P(Z \le z \mid \boldsymbol{x} \in \Pi_1)$ の高次元大標本漸近的枠組み

$$p/N \to r \in (0,1), \quad N_i/N \to k_i \in (0,1), \quad i=1,2$$

のもとでの漸近近似に対する誤差限界を考える. 本報告で与えられる誤差限界式はマハラノビスの距離 Δ に依存するが, 計算可能であり, 正しいオーダー評価にもなっている.

導出法は, 線形判別関数の場合の Fujikoshi (2000) に基づいている. すなわち, まず, Z が位置尺度変数

$$\widetilde{Z} = aZ = V^{1/2}Z - U, \quad Z \sim \mathrm{N}(0,1), \quad Z \perp (U,V), \quad V > 0$$

と表せることを用いる (Yamada et al. (2017)). ここに, $a = (1+N^{-1})/\{(1+N_1^{-1})(1+N_2^{-1})\}$, U, V については省略. 次に, 尺度混合変数に関する次の評価式 (Fujikoshi (2000), Fujikoshi and Ulyanov (2008)) などを適用する.

$$|P(\widetilde{Z} \le x) - \Phi(y)| \le B_2, \quad y = v_0^{-1/2}(x+u_0), \quad u_0 = \mathrm{E}(U), \quad v_0 = \mathrm{E}(V).$$

ここで, Φ は $\mathrm{N}(0,1)$ の分布関数で, 誤差限界 B_2 は次のように与えられる.

$$B_2 = \frac{1}{2\sqrt{2\pi e}}v_0^{-1}\mathrm{Var}(U) + \frac{1}{2}v_0^{-2}\mathrm{Var}(V) + \frac{1}{\sqrt{2\pi}}v_0^{-3/2}\{\mathrm{Var}(U)\mathrm{Var}(V)\}^{1/2}$$

限界式に関する数値例や, Studentized Z への応用についても報告する.

参考文献

1. Fujikoshi, Y. (2000). Error bounds for asymptotic approximations of the linear discriminant function when the sample size and dimensionality are large. *J. Multivariate Anal.*, **73**, 1–17.

2. Fujikoshi, Y. and Ulyanov, V.V. (2006). On accuracy of approximations for location and scale mixture. *J. Math. Sci.*, **138**, 5390–5395.

3. Yamada, T., Sakurai, T., and Fujikoshi, Y. (2017). Asymptotic approximations of EPMC for discriminant analyses for large dimension. *Hiroshima Statistical Research Group*, *Technical Report*, **17-12**.