高次元漸近近似に関する最近の発展

広島大学大学院理学研究科　藤越　康祝
広島大学大学院理学研究科　若木　宏文

1. はじめに

多変量統計的推測における漸近理論は、標本の大きさ n を無限大としたときの極限理論，あるいは極限理論を第 1 次近似とする漸近展開を基にした大標本漸近理論が重要な位置を占めている。これは変量の次元 p を固定したままでの漸近理論であり、当然のことであるかも知れないが次元 p が大きくなるにつれて近似が悪くなるという問題点が生じる。一方、情報化の進展と相俟って、次元 p が大であるデータが入手しやすくなり、標本数に比べ変量の次元が高い場合の高次元データ特有の解析法が必要になってきている。

本報告では、標本の大きさ n と変量の次元 p が共に大であるとした高次元枠組みのもとでの漸近近似についてのいくつかの結果を取り上げる。このような枠組みでの近似は伝統的な大標本の枠組みでの近似の弱点を補うのみならず、多くの場合次元が小さい場合にも良い近似であることは注目すべきことである。取り上げる問題は、(1) 標本共分散行列の固有値に関する経験分布と最大固有値の分布、(2) 多変量線形仮説に関する尤度比検定統計量の分布、(3) 判別における誤判別確率、(4) 判別解析における AIC 型基準の構成、である。

なお、多変量独立和の漸近分布および漸近展開についての基礎的成果（Portony (1986)、Anderson, Hall and Titterington (1998)）もあるが、その有用性については今後検討される必要がある。高次元の場合の平均パラメータの検定については、Dempster (1958, 1960) による方法が提案されている。同様な考え方に基づく判別法が Saranadasa (1993) によって考察されている。また、関数データの解析（Ramsay and Silverman (1997) など）, 高次元空間上の回帰（Owen (2000) など）も高次元データ解析として取り上げられるべき内容であるが、この報告では取り上げていない。

2. 固有値の分布

p 次元対称行列 S_n の固有値を $\ell_1 \geq \ldots \geq \ell_p$ とし、これらの固有値に関する経験分布関数を

$$F_n(x) = \frac{1}{p}\sharp\{\ell_i : \ell_i \leq x\}, \tag{2.1}$$

とする。ここに、$\sharp\{\cdot\}$ は指示された集合内の要素の数を表す。経験分布関数 $F_n(x)$ が F に収束するとき、F は S_n の極限スペクトル分布関数とよばれる。

行列 S_n は

$$S_n = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{x}_j \boldsymbol{x}_j', \tag{2.2}$$

で与えられるものとする。ここに、$\boldsymbol{x}_j = (x_{1j}, \ldots, x_{pj})'$ 、x_{ij} は独立同一分布に従う確率変数で、平均 0、分散 σ^2 をもつものとする。高次元確率行列のスペクトル解析においては、標本分散行列は $S_n = \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})'$ ではなく、(2.2) で定義される。これは正規性のもとでは本質的ではないが、非正規性のもとでは本質的である。

定理 2.1. (2.2) で定義される標本分散行列 S_n の経験スペクトル分布を F_n とする。pが大のとき $p/n \to c \in (0,1)$ とする。このとき、

$$F_n \to F, \quad a.s. \tag{2.3}$$

が成り立つ。ここに

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \frac{1}{2\pi c x \sigma^2}\sqrt{(x-a)(b-x)}, & (a < x < b), \\ 0, & (\text{otherwise}), \end{cases}$$

で、定数 a, b は $a = (1-\sqrt{c})^2\sigma^2,\ b = (1+\sqrt{c})^2\sigma^2$ で与えられる。

この結果は、Wachter (1978), Jonsson (1982) and Yin (1986) 等により求められたものであるが、モーメント存在に関するより強い条件もとでは Marčenko and Pastur (1967) により求められた。スペクトル分布は、経験分布関数を用いて表せる統計量

$$\begin{aligned} T_n &\equiv \frac{1}{p}\{\phi(\ell_1) + \ldots + \phi(\ell_p)\} \\ &= \int_0^\infty \phi(x) dF_n(x). \end{aligned} \tag{2.4}$$

の高次元枠組みのもとでの挙動を求めるのに利用できる。すなわち、統計量 T_n は適当な正則条件のもとで漸近的に

$$\int_0^\infty \phi(x) dF(x) \tag{2.5}$$

に収束する。

確率行列 S_1、S_2 は互いに独立で、それぞれウィシャート分布 $W_p(m, I_p)$、$W_p(n, I_p)$ に従うとする。仮定

$$p/m \to c_1 > 0 \quad \text{and} \quad p/n \to c_2 \in (0, \frac{1}{2}). \tag{2.6}$$

のもとでの行列 $F = S_1 S_2^{-1}$ の経験スペクトル分布は、Wachter (1980), Bai et al (1987), Silverstein (1985), Yin et al (1983) によって求められた。Bai (1999) は非正規の場合にも、S_1、S_2 の基礎分布がそれぞれ 2 次、4 次モーメントをもてば、同様な結果が得られることを注意している。

次に、(2.2) で定義された標本共分散行列 S_n の最大固有値 ℓ_1 の漸近近似について考える。Geman (1980) はある種のモーメント条件のもとで

$$\ell_1 \to (1+\sqrt{c})^2 \quad \text{a.s.},$$

すなわち、$n\ell_1 \sim (\sqrt{n}+\sqrt{p})^2$ を示し、この結果はその後 Yin et. all (1988) によって次のように一般化された。

定理 2.2. 確率行列 S_n は (2.2) で定義され、その基礎分布は 4 次モーメントをもとものとする。このとき、$p/n \to c \in (0,1)$ のもとで、次が成り立つ。

$$\ell_1 \to (1+\sqrt{c})^2 \quad \text{a.s.},$$

基礎分布が標準正規分布 $N(0,1)$ に従うとき、nS_n はウィシャー分布 $W_p(n, I_p)$ に従う。このとき、Johnstone (2001) は極限分布を求めることに成功している。中心化および尺度化定数は

$$\mu_{np} = (\sqrt{n-1}+\sqrt{p})^2, \tag{2.7}$$

$$\sigma_{np} = (\sqrt{n-1}+\sqrt{p})\left(\frac{1}{\sqrt{n-1}}+\frac{1}{\sqrt{p}}\right)^{1/3}. \tag{2.8}$$

で与えられる。また、極限分布関数は

$$F_1(s) = \exp\left\{-\frac{1}{2}\int_s^\infty (q(x)+(x-s)q^2(x))dx\right\}. \quad s \in R,$$

として与えられる。ここに、q は非線形 Painlevé II 微分方程式

$$\begin{aligned} q"(x) &= xq(x)+2q^3(x), \\ q(x) &\sim \mathrm{Ai}(x) \text{ as } x \to +\infty \end{aligned}$$

の解で、$\mathrm{Ai}(x)$ は Airy 関数である。この分布関数は Tracy and Widom (1996) によって、$p = n$ の特別な場合の極限関数として求められたものであり、次数 1 の *Tracy-Widom*

法則とよばれる。

Theorem 3.2. 確率行列 nS_n はウィシャート分布 $W_p(n, I_p)$ に従い、S_n の最大固有値を ℓ_1 とする。$p/n \to c \in (0, 1]$ とき

$$w_1 \equiv \frac{n\ell_1 - \mu_{np}}{\sigma_{np}} \to F_1 \quad \text{dist.} \tag{2.9}$$

この極限分布は数値的に求めることができ、数値表のためのソフトウエアが用意されている。

3. 多変量線形モデルにおける尤度比統計量の分布

多変量線形モデルにおける仮説による平方和・積和行列、誤差による平方和・積和行列をそれぞれ S_h、S_e とする。正規性のもとで尤度比検定統計量は $\Lambda = |S_e|/|S_e + S_h|$ の単調関数である。仮説のもとでの Λ の分布を考えるとき、一般性を失うことなく S_h、S_e は互いに独立にウイシャート分布 $W_p(q, I_p)$、$W_p(n, I_p)$ に従うとしてよい。このような Λ の分布を $\Lambda_{p,q,n}$ と表す。

通常の大標本の枠組みでは、p と q を固定し、n を無限大に近づける。このとき、次の漸近展開が Box (1949) により求められた。

$$\begin{aligned} \mathrm{P}(-m \log \Lambda) = G_f(x) &+ \frac{\gamma_2}{m^2}\{G_{f+4}(x) - G_f(x)\} \\ &+ \frac{1}{m^4}[\gamma_4\{G_{f+8}(x) - G_f(x)\} \\ &- \gamma^2\{G_{f+4}(x) - G_f(x)\}] + \mathrm{o}(n^{-4}), \end{aligned} \tag{3.1}$$

ここに、$f = pq$, $m = n - (p - q + 1)/2$, $\gamma_2 = f(p^2 + q^2 - 5)/48$,

$$\gamma_4 = \frac{1}{2}\gamma_2^2 + \frac{f}{1920}\left\{3(p^4 + q^4) + 10f^2 - 20(p^2 + q^2) + 159\right\},$$

$G_f(x)$ は自由度 f のカイ 2 乗分布の分布関数である。

Muldhokar and Trivedi (1980), (1981) は $T = -\log \Lambda_{p,q,n}$ の分布に対して、p かつ（または）qが大のときの漸近正規近似を与えている．この近似は χ^2-分布に対する Wilson-Hilferty の正規近似の考えを適用して求められたものである。すなわち、非負値確率変数列 $\{Y_k\}$ は、$k \to \infty$ のとき $(Y_k - \mu_k)/\sigma_k \to N(0, 1)$ であるとする。ここに、$\mathrm{E}(Y_k) = \mu_k$、$\mathrm{Var}(Y_k) = \sigma_k^2$ で、さらに、$\mu_k \to \infty$ かつ σ_k^2/μ_k は有界であるとする。このとき、変換 $Z_k = (Y_k/\mu_k)^h$ を考えると、

$$\frac{Z_k - 1}{h(\sigma_k/\mu_k)} = \frac{\mu_k(Z_k - 1)}{h\sigma_k} \to N(0, 1)$$

となる。定数 h は正規性への近似が加速するように定められるが、具体的には Z_k の分布が比較的対称になるように 3 次キュミュラントの主要項が 0 になるように定める。

最近、Tonda and Fujikoshi (2001) は高次元枠組み $p/n \to c \in (0,1)$ のもとで $T = -\log\Lambda$ の漸近展開を導出している。極限分布は正規分布となるが、その正規化変換は

$$\tilde{T} = \sqrt{\frac{p}{2}}\left(\frac{T-\mu}{d}\right) \tag{3.2}$$

である。ここに

$$\begin{aligned}
&\mu = \sum_{j=1}^{m} \log \xi_j^2, \quad d = \{\sum_{j=1}^{m} b_j^2\}^{1/2}, \\
&m_j = n+q-p+1-j, \quad \xi_j = (m_j+p)m_j^{-1}, \quad b_j = p(m_j+p)^{-1}, \\
&(j = 1, \ldots, q)
\end{aligned}$$

漸近展開の係数は

$$\begin{aligned}
\tau_1 &= \frac{1}{\sqrt{2}}a\xi^{-1}, \quad \tau_3 = \frac{1}{3\sqrt{2}}(2+3a)\xi^{-1}, \\
\tau_2 &= -\frac{1}{4}a(8+17a)\xi^{-2}, \quad \tau_4 = -\frac{7}{6}a(2+3a)\xi^{-2}, \quad \tau_6 = -\frac{1}{12}a(4+5a)\xi^{-2}, \\
\tau_{\alpha j} &= \tau_\alpha(a_j, \xi_j)
\end{aligned}$$

を用いて表せる。

Theorem 3.1. 統計量 $T = -\log\Lambda$ の (3.2) で与えられる正規化変量 $\tilde{T}$ の分布は次のように展開される。

$$\mathrm{P}(\tilde{T} \le x) = \Phi(x) - \phi(x)\left[\frac{1}{\sqrt{p}}p_1(x) + \frac{1}{p}p_2(x)\right] + \mathrm{o}(p^{-1}). \tag{3.3}$$

ここに、$\Phi(x)$、$\phi(x)$ はそれぞれ $N(0,1)$ の分布関数、密度関数であり、$p_j(x)$'s はエルミート多項式 $h_j(x)$ を用いて次のように与えられる。

$$\begin{aligned}
p_1(x) &= \tau_{1\cdot} + \tau_{3\cdot}h_2(x), \\
p_2(x) &= \tau_{1\cdot}^2 h_1(x) + (\tau_{4\cdot} + 2\tau_{1\cdot}\tau_{3\cdot} - 2\tau_{(13)\cdot})h_3(x) + \tau_{3\cdot}^2 h_5(x), \\
\tau_{\alpha\cdot} &= \frac{1}{d^\alpha}\sum_{j=1}^{m} b_j^\alpha \tau_{\alpha j}, \quad \tau_{(13)\cdot} = \frac{1}{d^4}\sum_{j=1}^{m} b_j^4 \tau_{1j}\tau_{3j}.
\end{aligned}$$

この漸近展開近似は、p が n に近づくと Box の漸近展開近似を相当に改良し、また、n が p より相当大きい場合においても Box の漸近展開近似と同程度な近似になっている。

4. 判別関数の分布近似と誤差限界

4.1 線形判別関数

p次元変数$\boldsymbol{x}$の観測値に基づいて、その観測値が２つの正規母集団$N(\boldsymbol{\mu}^{(1)}, \Sigma)$, $N(\boldsymbol{\mu}^{(2)}, \Sigma)$のいずれに属するかを判定する判別問題を考える。パラメータ$\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}, \Sigma$は未知であるが、各母集団$\Pi^{(i)}$からの大きさ$n_i$の標本が与えられているとする。従って、$\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)}$はそれぞれ標本平均ベクトル$\bar{\boldsymbol{x}}^{(1)}$, $\bar{\boldsymbol{x}}^{(2)}$で推定され、$\Sigma$は２つの標本分散行列を合併した標本分散行列$S$で推定される。$\Delta$を母集団マハラノビス距離、$D$を標本マハラノビス距離とする。このとき、線形判別関数

$$W = (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} \{x - \frac{1}{2}(\bar{x}^{(1)} + \bar{x}^{(2)})\}$$

を用いて、$W > 0 \ \Rightarrow x \in \Pi^{(1)}$, $W \leq 0 \ \Rightarrow x \in \Pi^{(2)}$ と判別するときの誤判別確率近似の誤差評価を問題にする。

一般に、誤判別確率を正確に評価することは困難であるが、標本数$n_1, \ n_2$が大きいときの漸近展開公式が求められている (Okamoto (1963), Siotani (1982)、等)。また、誤判別確率の高次漸近不偏推定量が構成されている (McLachlan (1974, 1976))。しかし、当然のことであろうが、これらの近似の精度は次元が大きくなるにつれて悪くなる。標本数に加え次元数も大きい場合の研究としては、ロシアの研究者 Deev (1970, 1972)、Rudys (1972) 等による線形判別関数に関する結果がある。この場合の漸近的枠組みとして

$$\lim_{p\to\infty} n_i/p = \lambda_i (> 0), \ i = 1, \ 2, \ \lim_{p\to\infty} (n - p) = \infty, \ \lim_{p\to\infty} \Delta^2 = d_0^2 \tag{4.1}$$

を仮定する。ここに$n = n_1 + n_2$。$n_1 = n_2$の場合に、彼らによって提案された漸近的近似公式、およびその拡張の精度は、Wyman et al. (1990)、Fujikoshi and Seo (1997) によって数値的に調べられ、その精度は著しく良く、また、pが小さいときにも有効げあることは注目されてよい。ここでは、高次元の枠組みでの近似のみならずその近似の誤差限界にいての結果を述べる。簡単のため、漸近分布に対する誤差限界について述べるが、その結果は漸近展開近似の場合に対しても拡張される (Fujikoshi(2000))。

以下では、$\boldsymbol{x}$が$\Pi^{(1)}$に属するとき誤判別確率の近似とその誤差限界を問題にする。このときWを

$$W = V^{-\frac{1}{2}} Z - U \tag{4.2}$$

と表すことができる。ここに

$$\begin{aligned} V &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{-1} \Sigma S^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2), \\ Z &= V^{-\frac{1}{2}} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1), \\ U &= (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{-1} (\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) - \frac{1}{2} D^2, \end{aligned}$$

で、Z は $N(0,1)$ に従い、Z と (U,V) は独立である。従って、$\boldsymbol{x}$ が $\Pi^{(1)}$ に属するときの誤判別確率は

$$\begin{aligned} e(2|1) &= \mathrm{P}(W \leq 0 | \boldsymbol{x} \in \Pi^{(1)}) \\ &= \mathrm{E}_{(U,V)} \left\{ \Phi(V^{-\frac{1}{2}} U) \right\} \end{aligned} \tag{4.3}$$

と表せる。この表示において、U, V をそれぞれの平均

$$\begin{aligned} \mathrm{E}(U) &= -\frac{n-2}{2(m-1)} \left\{ \Delta^2 + \frac{(n_1 - n_2)p}{n_1 n_2} \right\} = u_0 \quad (m > 1), \\ \mathrm{E}(V) &= \frac{(n-2)^2 (n-3)}{m(m-1)(m-3)} \left\{ \Delta^2 + \frac{np}{n_1 n_2} \right\} = v_0 \quad (m > 3), \end{aligned}$$

で置換えた近似

$$\Phi\left(\{\mathrm{E}(V)\}^{-\frac{1}{2}} \mathrm{E}(U) \right) = \Phi(\gamma) \tag{4.4}$$

が提案される。ここに、$m = n - p - 2$、

$$\begin{aligned} \gamma &= \{m(m-3)\}^{1/2} \{(n-3)(m-1)\}^{-1/2} \gamma_0, \\ \gamma_0 &= -\frac{1}{2} \{\Delta^2 + (N_1 - N_2) p (N_1 N_2)^{-1}\} \{\Delta^2 + Np(N_1 N_2)^{-1}\}^{-\frac{1}{2}} \end{aligned}$$

である。

この近似は、上記のような確率構造を用いることなく、Lachenbruch (1968) によっても提案されているものである。この近似について次が成り立つ (Fujikoshi (2000))。

Theorem 4.1. 高次元漸近的枠組み (4.1) のもとで、

$$|e(2|1) - \Phi(\gamma)| \leq B \tag{4.5}$$

が成り立つ。ここに、

$$\begin{aligned} B &= \boldsymbol{\beta}_{2,0} v_0^{-1} \mathrm{Var}(U) + \boldsymbol{\beta}_{2,2} v_0^{-2} \mathrm{Var}(V) \\ &\quad + \boldsymbol{\beta}_{2,1} v_0^{-\frac{3}{2}} \{\mathrm{Var}(U) \cdot \mathrm{Var}(V)\}^{\frac{1}{2}}, \end{aligned}$$

定数 $\beta_{2,j}$ は、

$$\beta_{2,0} = \frac{1}{2}h_1, \quad \beta_{2,1} = \frac{1}{2}h_2, \quad \beta_{2,2} = \frac{1}{2}\left\{\sqrt{1+h_1} + \frac{1}{2}\sqrt{h_3}\right\}^2$$

で与えられる。ただし、$h_j = \sup|h_j(x)\phi(x)|$ ($h_j(x)$ はエルミート多項式)。

定数項のより良い評価として

$$\beta_{2,0} = 0.121, \quad \beta_{2,1} = 0.2, \quad \beta_{2,2} = 0.5$$

を用いることができる。上界 B のオーダーは O_1^* を満たしているが、より具体的には、U と V の分散を評価する必要がある。ここに、O_j^* は高次元枠組みでのオーダー表示であって、$(n_1^{-1}, n_2^{-1}, p^{-1})$ に関するオーダー j の項を表す。通常の大標本でのオーダーは単に O_j と表す。U と V の分散は、次のように与えられる。

$$\begin{aligned}
\mathrm{Var}(U) &= \frac{(n-2)^2}{2m(m-1)(m-3)}\left[\frac{1}{m-1}\Delta^4\right. \\
&\quad + \frac{2(n-3)}{mn_2}\Delta^2\left\{1 + \frac{n_1 - n2}{(m-1)n_1}\right\} \\
&\quad \left. + \frac{2(n-3)p}{n_1 n_2}\left\{\frac{1}{m} + \frac{(n_1-n_2)^2}{2(m-1)n_1 n_2}\right\}\right], (m-3>0).
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(V) &= \frac{2(n-3)(n-2)^4}{m(m-1)^2(m-3)^2}\left[\frac{1}{m}\left\{1 + \frac{8(m-4)}{(m-5)(m-7)}\right\}\right. \\
&\times \left\{\frac{p-1}{m}\left(\Delta^2 + \frac{pn}{n_1 n_2}\right)^2 + \frac{n(n-5)}{n_1 n_2}\left(2\Delta^2 + \frac{pn}{n_1 n_2}\right)\right\} \\
&+ \left.\frac{4(n-3)(m-4)}{m(m-5)(m-7)}\left(\Delta^2 + \frac{pn}{n_1 n2}\right)^2\right], (m-7>0).
\end{aligned}$$

4.2 ２次判別関数

２つの正規母集団 $\Pi_1 : N_p(\boldsymbol{\mu}_1, \Sigma_1), \Pi_2 : N_p(\boldsymbol{\mu}_2, \Sigma_2)$ に関する判別問題を扱う。母数 $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2)$ が既知の場合，最適な判別ルールは

$$Q(\boldsymbol{x}; \boldsymbol{\theta}) = (\boldsymbol{x} - \boldsymbol{\mu}_1)'\Sigma_1^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_1) - (\boldsymbol{x} - \boldsymbol{\mu}_2)'\Sigma_2^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_2) + \log\det\Sigma_1\Sigma_2^{-1}$$

の大小に基づいて行われる。$\boldsymbol{\theta}$ が未知の場合，Π_1, Π_2 からの，それぞれ大きさ n_1, n_2 の標本に基づく推定量

$$\widehat{\boldsymbol{\theta}} = (\bar{X}_1, \bar{X}_2, S_1, S_2)$$

を用いた判別関数 $Q(\boldsymbol{x};\widehat{\boldsymbol{\theta}})$ が用いられることが多い。ただし，X_i, S_i はそれぞれ，Π_i からの標本平均ベクトル，不偏分散共分散行列である。$Q(\boldsymbol{x};\widehat{\boldsymbol{\theta}})$ を標本２次判別関数と呼ぶこととする。

$X \sim \Pi_j$ ($j=1$ または $j=2$) として，$Q = Q(X;\widehat{\boldsymbol{\theta}})$ の確率分布を考える。p を固定して，$n_1 \to \infty, n_2 \to \infty$ としたときの Q の極限分布は，p 個の独立な非心カイ２乗変数の１次結合の分布として表現できる。$\Sigma_1^{-1} - \Sigma_2^{-1}$ が正定値行列または負定値行列ならば，１次結合の係数は同符号となるので，極限分布関数は自由度との異なるカイ２乗分布関数の無限級数として表される。この場合，Q の特性関数を $1/n_1, 1/n_2$ について展開し，反転することで分布関数の漸近展開も同様の表現が可能と思われるが p が大きい場合その近似精度は期待できない。また，$\Sigma_1^{-1} - \Sigma_2^{-1}$ が正定値でも負定値でもない場合には，筆者の知るかぎりでは，Q の極限分布についについて扱い易い（数値積分を必要としない）表現は得られていない。

一方，p も大きくなる場合には，上記の１次結合の係数列および非心率の列が適当な条件をみたすならば Q の極限分布は正規分布となる。この小節では $p \to \infty$ に対して

$$\nu_i := p/(n_i - 1) < 1 \ \liminf_{p\to\infty} \nu_i > 0, \ i = 1, 2$$

と仮定したときの Q の分布の漸近展開導出について述べる。

$j = 1$ すなわち $X \sim \Pi_1$ とする。Q の分布のアフィン変換に対する不変性から，$\Sigma_1 = I_p, \Sigma_2 = \mathrm{diag}(\lambda_1, \cdots, \lambda_p), \boldsymbol{\mu}_1 = 0, \boldsymbol{\mu}_2 = (\eta_1, \cdots, \eta_p)'$ と仮定して一般性を失わない。ウィシャート行列に関する定理（例えば，Muirhead(1982), Theorem 3.2.10）から，

$$Q = \sum_{k=1}^{p} \left\{ \frac{n_1 - 1}{v_{11}}(x_k - y_{1k})^2 - \frac{n_2 - 1}{v_{21}}\lambda_k^{-1}(x_k - y_{2k})^2 + \log(v_{1k}/(n_1 - 1)) - \log(v_{2k}/(n_2 - 1)) \right\},$$

と表すことができる。ただし，$v_{11}, v_{12}, \cdots, v_{1p}$; $v_{21}, v_{22}, \cdots, v_{2p}$; $z_1, \cdots, z_p$; $y_{11}, y_{12}, \cdots, y_{1p}$; $y_{21}, y_{22}, \cdots, y_{2p}$ は，互いに独立な１次元確率変数で，

$$z_k \sim N(0, 1); \quad y_{1k} \sim N(0, n_1^{-1}); \quad y_{2k} \sim N(\eta_k, n_2^{-1}\lambda_k); \quad v_{ik} \sim \chi^2_{n_i - 1 - p + k}$$

$(k = 1, \cdots, p; i = 1, 2)$ である。

したがって，Q の $v_{i,k}, (i = 1, 2; k = 1, \cdots, p)$ を与えたときの条件付き分布は，独立な確率変数の和の分布となりキュムラント計算によりエッジワース展開可能となる。得られた展開公式の，$v_{i,k}$ の分布に関する期待値をとれば Q の分布の漸近展開公式が得られる。この導出方法の妥当性のために次の補題を準備する。この補題は，Bhattacharya and Ranga Rao (1976) の Corollary 20.4 と本質的に同等で，i.i.d. の枠

組みを，Bhattacharya and Ranga Rao の Theorem 20.6 よりほんのわずか広げただけのものである。

Lemma 4.2 $n = 1, 2, \cdots,$ に対して，$\{Z_{1,n}, Z_{2,n}, \cdots, Z_{n,n}\}$ を独立な n 個の k-次元確率ベクトルとする。$\mathrm{E}(Z_{j,n}) = 0, \mathrm{Cov}(Z_{j,n})$ はすべて正則，ある $s \geq 3$ に対して，$\mathrm{E}(||Z_{j,n}||^s) < \infty$ であり，次の 3 つの条件 (i)〜(iii) を満たすとする。

(i) λ_n を $V_n = n^{-1}\sum_{j=1}^n \mathrm{Cov}(Z_{j,n})$ の最小固有値とすると，$\liminf_{n\to\inf} \lambda_n > 0,$

(ii) $\limsup_{n\to\infty} n^{-1}\sum_{j=1}^n \mathrm{E}(||Z_{j,n}||^s) < \infty,$ かつ，任意の正数 ε に対して

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^n \mathrm{E}(1_{\{z;||z||>\varepsilon n^{1/2}\}}(Z_{j,n})||Z_{j,n}||^s) = 0$$

(iii) $g_{j,n}(t)$ を $Z_{j,n}$ の特性関数とすると，任意の正数 b に対して

$$\limsup_{n\to\infty} \max_{1\leq j\leq n} \sup_{||t||>b} |g_{j,n}(t)| < 1,$$

ただし，$||\cdot||$ はユークリッドノルム，R^k の部分集合 A に対して，$1_A(z)$ は A の定義関数，すなわち，$z \in A$ のとき 1，$z \notin A$ のとき 0 をとる z の関数を表す。このとき，$\mathcal{C}$ を R^k のボレル可測な凸部分集合の全体とし，$\Psi_s(z)$ をキュムラントに基づく $U_n = n^{-1/2}V_n^{-1}\sum_{j=1}^n Z_{j,n}$ の分布関数の $O(n^{-(s-2)/2}$ の項までのエッジワース展開とすると，

$$\sup_{C\in\mathcal{C}} \left| \mathrm{Prob}(U_n \in C) - \int_C d\Psi_s \right| = o(n^{-(s-2)/2})$$

が成り立つ。

Lemma 4.2. の $Z_{j,n}$ として，

$$Z_{k,p} = \begin{pmatrix} (x_k - y_{1k})^2 - \mathrm{E}\{(x_k - y_{1k})^2\} \\ \lambda_k^{-1}(x_k - y_{2k})^2 - \mathrm{E}\{\lambda_k^{-1}(x_k - y_{2k})^2\} \end{pmatrix},\ k = 1, \cdots, p$$

$$U_p = \left\{ p^{-1}\sum_{k=1}^p \mathrm{Cov}(Z_{k,p}) \right\}^{-1/2} \sum_{k=1}^p Z_{k,p}$$

ととると，

$$\begin{aligned} &\mathrm{P}(Q < c | v_{11}, \cdots, v_{2p}) \\ =&\mathrm{P}\left\{ \left(\frac{n_1 - 1}{v_{11}}, -\frac{n_2 - 1}{v_{21}} \right)' \left(p^{-1}\sum_{k=1}^p \mathrm{Cov}(Z_{k,p}) \right)^{1/2} U_p < constant \,\middle|\, v_{11}, \cdots, v_{2p} \right\} \end{aligned}$$

と表され，これは，U_p が平面で仕切られた領域内にある確率であるから，$Z_{k,p}$ が Lemma 4.2 の条件を満たせば，Q の条件付き分布関数を展開したときの誤差のオーダーは，$v_{11}, \cdots, v_{2p}$ に関して一様であることがわかる。したがって，未知母数に関して次の条件を仮定すれば $O(p^{-(s-2)/2})$ の項まで漸近展開可能となる。

$$\frac{1}{p}\sum_{j=1}^{p}\lambda_j^{-k}|\eta_j|^2 = O(1), \quad (k = 1, \cdots, s; l = 0, \cdots, s)$$

$v_{11}, \cdots, v_{2p}$ に関する期待値計算が残っているが，自由度が大きくなるカイ２乗確率変数の有界な C^∞ 級関数の期待値を展開すれば良いので，各カイ２乗変数をその自由度で割って 1 の周りでテーラー展開し，剰余項のオーダーに注意して項別に期待値をとれば Q の分布関数の漸近展開が得られる。

現在，Matsumoto and Wakaki により，$O(p^{-1})$ のまでの漸近展開公式を導出しており，その精度を数値計算によって調べている所である。

5. 判別分析における変数選択基準の構成

5.1 AIC 型基準

変数 $\boldsymbol{y} = (y_1, ..., y_p)'$ が q 個の母集団 $\Pi_1, ..., \Pi_q$ で観測され、全観測データを

$$Y = [\boldsymbol{y}_1^{(1)}, ..., \boldsymbol{y}_{n_1}^{(1)}, ..., \boldsymbol{y}_1^{(q)}, ..., \boldsymbol{y}_{n_q}^{(q)}]'$$

とする。ここに、 $\boldsymbol{y}_1^{(i)}, ..., \boldsymbol{y}_{n_i}^{(i)}$ は、Π_i からの大きさ n_i の無作為標本である。 $n \times p$ の観測行列 Y は、真のモデル M_* のもとで確率密度関数 $g(Y)$ をもち、

$$M_*: \ \mathrm{E}^*(\boldsymbol{y}|\Pi_i) = \boldsymbol{\mu}^{(i)*}, \qquad Var^*(\boldsymbol{y}|\Pi_i) = \Sigma^*, \tag{5.1}$$

とする。E^* および Var^* は、真のモデル M_* のもとでの平均、共分散行列である。

添え字の集合 $P = \{1, 2, \ldots, p\}$ の任意の部分集合を K とし、添え字の集合が K である $\boldsymbol{y}$ の部分ベクトルを $\boldsymbol{y}_K$ とする。各部分集合 K に対して、変数選択モデル M_K を定め、その AIC 型リスク R_K の推定量 AIC_K を構成することにより、変数選択基準が構成される。関心のある部分集合の集まりを $\mathcal{K}$ とするとき、変数選択基準は

$$\min_{K\in\mathcal{K}} AIC_K = AIC_{\hat{\mathcal{K}}}$$

となる $\hat{K}$ を選択する。記号を簡単にするため、以下では，$K = k = \{1, 2, \ldots, k\}$ とする。モデル M_k のリスク（悪さ）を AIC 基準の考えに沿って

$$R_k = \mathrm{E}_Y^*\mathrm{E}_X^*\{-2\log f(X; \hat{\Theta}_k)\}$$

で定義する。ここに、$n \times p$ 確率行列 X は Y と同じ分布をもち、Y とは独立であるとする。また、$\hat{\Theta}_k$ はモデル M_k のもとでの Θ の最尤推定量である。変数選択基準の構成においては、R_k の推定量を構成することが本質的であるが、Akaike (1973) に沿って R_k を

$$\hat{R}_k = -2 \log f(Y; \hat{\Theta}_k) + \hat{b}_k$$

として推定する。ここに、$\hat{b}_k$ は

$$b_k = \mathrm{E}_Y^* \mathrm{E}_X^* \{-2 \log[f(X; \hat{\Theta}_k)/f(Y; \hat{\Theta}_k)]\}$$

の推定量である。一般に、AIC 基準 (Akaike (1973)) では、通常の大標本理論と候補のモデルが真のモデルを含むという仮定のもとで、b_k をモデル M_k に含まれる独立パラメータ数の 2 倍として推定する。リスクにより忠実な基準を構成するためには、より不偏性の高い b_k の推定量を構成する必要があるが、これがいわゆる AIC の改良問題である。一方、変数選択に直結するモデル M_k の導入も重要である。より具体的には、最初の k 変数 $\boldsymbol{y}_1 = (y_1, ..., y_k)$ が十分で、残りの変数 $\boldsymbol{y}_2 = (y_{k+1}, ..., y_p)$ は冗長である、すなわち、変数 $\boldsymbol{y}_2$ は、正準判別分析において追加情報をもたないというモデルを考える。このようなモデル M_k を定式化するため、

$$\boldsymbol{\mu}^{(i)} = \begin{pmatrix} \boldsymbol{\mu}_1^{(i)} \\ \boldsymbol{\mu}_2^{(i)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

と分割し、

$$\boldsymbol{\mu}_{2\cdot1}^{(i)} = \boldsymbol{\mu}_2^{(i)} - \Gamma \boldsymbol{\mu}_1^{(i)}, \quad i = 1, ..., q, \ \Gamma = \Sigma_{21} \Sigma_{11}^{-1}.$$

とおく。このとき、M_k を、

$$M_k : \boldsymbol{\mu}_{2\cdot1}^{(1)} = \ldots = \boldsymbol{\mu}_{2\cdot1}^{(q)}. \tag{5.2}$$

と定める (Rao (1948,1970))。$\boldsymbol{y}|\Pi_i \sim N(\boldsymbol{\mu}^{(i)}, \Sigma), i = 1, ..., q$ のときの Y の確率密度関数を $f(Y; \Theta)$ とかく。ただし、$\Theta = \{\boldsymbol{\mu}^{(1)}, ..., \boldsymbol{\mu}^{(q)}, \Sigma\}$。第 i 群および全標本の平均ベクトルをそれぞれ、$\bar{\boldsymbol{y}}^{(i)}$ 、 $\bar{\boldsymbol{y}}$ とする。また、群内平方和積和行列を W 、群間平方和積和行列を B とし、$T = W + B$ とする。このとき、最尤推定量は

$$\begin{aligned} \bar{\boldsymbol{\mu}}_1^{(i)} = \bar{\boldsymbol{y}}_1^{(i)}, \bar{\boldsymbol{\mu}}_{2\cdot1}^{(i)} = \bar{\boldsymbol{y}}_2 - \hat{\Gamma} \bar{\boldsymbol{y}}_1, \\ \hat{\Gamma} = T_{21} T_{11}^{-1}, \ n\hat{\Sigma}_{11} = W_{11}, \\ n\hat{\Sigma}_{22\cdot1} = T_{22\cdot1} = T_{22} - T_{21} T_{11}^{-1} T_{12}. \end{aligned}$$

で与えられる。従って

$$\begin{aligned} -2\log f(Y;\hat{\Theta}_k) &= -n\log|n^{-1}W_{11}| \\ &\quad + n\log|n^{-1}T_{22\cdot1}| + np(1+\log 2\pi) \\ &= -n\log\{|W_{22\cdot1}|/|T_{22\cdot1}|\} \\ &\quad + n\log|n^{-1}W| + np(1+\log 2\pi). \end{aligned} \tag{5.3}$$

バイヤス項 b_k は次のように表せる (Fujikoshi (1985, 2000))。

$$\begin{aligned} b_k &= -np + \frac{nk(n+q)}{n-k-q-1} \\ &+ E_Y^*[n^2\mathrm{tr}T^{-1}\{(1+n^{-1})\Sigma+\Omega\} \\ &- n^2\mathrm{tr}T_{11}^{-1}\{(1+n^{-1})\Sigma_{11}+\Omega_{11}\}]. \end{aligned} \tag{5.4}$$

ただし、ここでは真のモデルのもとでのパラメータ $\boldsymbol{\mu}^{(i)*}, \Sigma^*$ を、単に $\boldsymbol{\mu}^{(i)}, \Sigma$ と表し,

$$\bar{\boldsymbol{\mu}} = \sum_{i=1}^{q}(n_i/n)\boldsymbol{\mu}^{(i)}, \quad \Omega = \sum_{i=1}^{q}(n_i/n)(\boldsymbol{\mu}^{(i)}-\bar{\boldsymbol{\mu}})(\boldsymbol{\mu}^{(i)}-\bar{\boldsymbol{\mu}})'$$

である。

5.2 高次元・2群の場合

高次元の枠組みで、2群の場合には (5.4) 式は次のように評価される。

定理 5.1. 真のモデル M_* に関して正規性を仮定し、$q=2$ とする。このとき、バイヤス項 b_k は高次元枠組みのもとで次のように展開される。

$$b_k = b_{k1} + O_1^*, \tag{5.5}$$

ただし、

$$\begin{aligned} b_{k1} &= \frac{n}{n-k-3}\{-(p-k)n+pk+3p+2k\} \\ &+ n^2\{Q(n,p,\tau^2)-Q(n,k,\tau_1^2)\}. \end{aligned}$$

ここに、$\tau^2 = \frac{n_1n_2}{n^2}\Delta^2, \tau_1^2 = \frac{n_1n_2}{n^2}\Delta_1^2$, Δ, Δ_1 はそれぞれ $\boldsymbol{y}, \boldsymbol{y}_1$ に基づくマハラノビスの距離、

$$\begin{aligned} Q(n,p,\tau^2) &= \frac{p(n+1)}{n(n-p-2)} \\ &+ \tau^2\left\{\frac{1}{n-p-3} - \frac{n+1}{n(n-3)(n-p-2)}\right\} \\ &- \frac{\tau^2}{(n-p-1)(n+1)}\left\{\tau^2 + \frac{n-2}{n-p-2}\right\}. \end{aligned}$$

系 4.1. M_k が真のモデル M_* を含むとき、すなわち、$\tau^2 = \tau_1^2$ のとき、

$$\begin{aligned} b_{k1} &= b_{k0} + (p-k)\tau^2 + \mathrm{O}_1 \\ &= b_{k0} + \mathrm{O}(n^{-1}), \quad \text{if } \tau^2 = \mathrm{O}_1, \end{aligned} \tag{5.6}$$

ここに、$b_{k0} = 2\{2k + (p-k) + \frac{1}{2}p(p+1)\}$。

b_{k0} はモデル M_k のもとでの独立パラメータの 2 倍であることを注意したい。b_k の推定としては、上記評価式において、Δ^2, Δ_1^2 にそれぞれ不偏推定量

$$\hat{\Delta}^2 = \frac{N-p-3}{N-2}D^2 - \frac{Np}{N_1N_2}, \quad \hat{\Delta}_1^2 = \frac{N-k-3}{N-2}D_1^2 - \frac{Nk}{N_1N_2},$$

を代入したものが提案される。ここに、D, D_1 はそれぞれ $\boldsymbol{y}, \boldsymbol{y}_1$ に基づく標本マハラノビスの距離である。以上より AIC_k 基準の修正として

$$MAIC_k = -2\log f(Y; \hat{\Theta}_k) + \hat{b}_{k1}. \tag{5.7}$$

が提案される。最大尤度は

$$\begin{aligned} -2\log f(Y; \hat{\Theta}_k) &= -n\log\left\{1 + \frac{D_1^2 - D^2}{n(n-2)(n_1n_2)^{-1} + D^2}\right\} \\ &+ \log|n^{-1}W| + np(1 + \log 2\pi), \end{aligned} \tag{5.8}$$

と表せる。ここに、$(n-2)^{-1}W$ は通常の合併標本共分散行列である。

References

[1] Akaike, H.(1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Eds.B.N.Petrov and F.Csáki,pp.267-281. Budapest: Akadémia Kiado.

[2] Anderson, N. H., Hall, P. and Titterington, T. M. (1998). Edgeworth expansions in very-high-dimensional problems. *Journ. Stat. Plan. Inf.* **70**, 1-18.

[3] Anderson, T. W. (1984). *An Introduction to Multivariate Analysis* (2nd ed.). John Wiley & Sons.

[4] Bai, Z. D. (1999). Methodologies in spectoral analysis of large dimensional random marices, a review. *Statistica Sinica* **9**, 611-677.

[5] Bai, Z. D., Yin, Y. Q. and Krishnaiah, P. R. (1987). On the limiting empirical distributionfunction of the eigenvalues of a multivariate F matrix. *Theory Probab. Appl.*, **32**, 490-500.

[6] Bhattacharya,R.N. and Rao ,R.R.(1976) *Normal Approximation and Asymptotic Expansions.* Wiley, New York. Reprint with corrections and Supplemental material(1986). Krieger, Malabar, Florida.

[7] Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika.*, **36**, 317-346.

[8] Deev, A. D. (1970). Representation of statistics of discriminant analysis and asymptotic expansions when space dimensions are comparable with sample size. *Soviet Math. Dokl.* **11**, 1547-1550.

[9] Dempster, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.*, **29**, 995-1010.

[10] Dempster, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, **16**, 41-50.

[11] Fujikoshi, Y.(1985). Selection of variables in discriminant analysis and canonical correlation analysis. In Multivariate Analysis-VI, Ed. P.R. Krishnaian, pp. 219-236, Elsevier Science Publishers B.V.

[12] 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.

[13] Fujikoshi, Y. (2002). Selection of variables for discriminant analysis in a high-dimensional case. To appear in Sankhya.

[14] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC's of the linear and the quadratic discriminant functions when the samples sizes and the dimension are large. *Statist. Anal. Random Arrays* **6** , 269-280.

[15] Geman, S. (1980). A limit theorem for the norm of random matrices. *Ann. Probab.*, **8**, 252-261.

[16] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.*, **29**, 295-327.

[17] Jonsson, D. (1982). Some limit theorems for eigenvalues of a sample covariance matrix. *J. Multivairiate Anal.*, **12**, 1-38. Lachenbruch, P. A. (1968). On expected probabilities of misclassification in discriminant analysis, necessary sample size, and a relation with the multiple correlation coefficients. *Biometrics* **24**, 823-834.

[18] Marčenko, V. A. and Pastur, L. A. (1967). Distribution for some sets of random matrices. *Math. USSR-Sb*, **1**, 457-483.

[19] McLachlan, G. J. (1974). An asymptotic unbiased technique for estimating the error rates in discriminant analysis, *Biometrics* **30**, 239-249.

[20] McLachlan, G. J. (1976). The bias of the apparent error rate in discriminant analysis. *Biometrika* **63**, 239-244.

[21] Muirhead ,R. J.(1982) *Aspects of Multivariate Statistical Theory.* John Willey and Sons, Inc., New York.

[22] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.* **34**, 1286-1301.

[23] Owen, A. B. (2000). Assessing linearity in high dimensions. *Ann. Statist.* , **28**, 1-19.

[24] Portnoy, S. (1986). On central limit theorem in R^p when $p \to \infty$. *Wahrscheinlichkeitstheorie verw.Geb.* **73**, 571-583.

[25] Rao, C. R. (1948). Tests of significance in multivariate analysis. *Biometrika* **35**, 58-79.

[26] Rao, C. R. (1970). Inference on discriminant function coefficients. In *Essays in Prob. and Statist.* Ed. R. C. Bose, pp. 537-602, Univ. of North Carolina Press, Chapel Hill.

[27] Raudys, S. (1972). On the amount of priori information in designing the classification algorithm. *Tech, Cybern.* **4**,168-174(in Russian).

[28] Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistic. *Handbook of Statistics* **2** (P. R. Krishnaiah and L. N. Kanal, Eds.), North–Holland Publishing Company, 47-60.

[29] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*. American Sciences Press, Columbus, Ohio.

[30] Wyman, F. J., Young, D. M. and Turner, D. W. (1990). A comparison of asymptotic error rate expansions for the sample linear discriminant function. *Pattern Recognition* **23**, 775-783.