

非心 Wishart 行列の Wishart 行列に関する固有値の 漸近展開と MANOVA モデルにおけるある種の検定統 計量の漸近展開

阪大・基工 磯貝 恭史

序. 多変量解析においてはよく次の固有値問題

$$(I) \quad \dots \quad |S_h - d S_e| = 0$$

の固有値の分布を調べることが必要になる (MANOVA ~~等~~).

但し、 S_e, S_h をそれぞれ互いに独立な $p \times p$ 行列とし、 S_e は central Wishart $W_p(n_e, \Sigma)$, S_h は non-central Wishart $W_p(n_h, \Sigma, \Omega)$ に従うものとする。ここで、 Ω は non centrality parameter である。

問題 (I) は簡単な変換によって

$$S_e \sim W_p(n_e, I_p), \quad S_h \sim W_p(n_h, I_p, \Omega)$$

Ω : diagonal matrix

に従う問題に帰着出来る。

問題 (I) の固有値の漸近展開を導出するのに、ここでは次の仮定を置く。仮定: $n_e : n_h = e : h$ ($e > 0, h > 0, e + h = 1$) 且 $n = n_e + n_h \rightarrow \infty$ とする。かつ、 $\Omega = n \Phi$ とおく。

この仮定の situation が時に必要であることが Fujihashi (1974) の論文で numerical example を伴って報告されている。この報告においては、固有値の漸近展開と MANOVA モデルでの次元の検定のための幾つかの統計量の漸近展開を示す。

§1. $S_h S_e^{-1}$ の固有値の分布の漸近展開

問題 (I) の固有値を大きさの順に並べて $\lambda_1 > \dots > \lambda_p$ とする。目標は $(\lambda_1, \dots, \lambda_p)$ の同時分布及び周辺分布の漸近展開を行うことである。 $S_h S_e^{-1}$ は $n \rightarrow \infty$ の時 $(h/e) I_p + (2/e) \textcircled{4}$ に確率収束するので λ_α の $(1/n) S_e = e I_p$, $(1/n) S_h = h I_p + 2 \textcircled{4}$ の周りでの展開を求めたい。

$T_e = \sqrt{n} (S_e/n - e I_p)$, $T_h = \sqrt{n} \{S_h/n - (h I_p + 2 \textcircled{4})\}$ と置けば、 $n \rightarrow \infty$ の時、 T_e と T_h は $p(p+1)/2$ 変量の平均 0 の正規分布に法則収束する。

$$\begin{aligned} (1.1) \quad S_e^{-1/2} S_h S_e^{-1/2} &= \{e I_p + (1/\sqrt{n}) T_e\}^{-1/2} \{ (h I_p + 2 \textcircled{4}) + (1/\sqrt{n}) T_h \} \\ &\quad \times \{e I_p + (1/\sqrt{n}) T_e\}^{-1/2} \\ &= \Lambda + (1/\sqrt{n}) V^{(0)} + (1/n) V^{(1)} + \dots \end{aligned}$$

$$\text{但し, } \Lambda = (h/e) I_p + (2/e) \textcircled{4}$$

$$V^{(0)} = e^{-1} \{ T_h - (1/2) (T_e \Lambda + \Lambda T_e) \}$$

$$V^{(1)} = e^{-2} \{ (3/8) (T_e^2 \Lambda + \Lambda T_e^2) - (1/2) (T_e T_h + T_h T_e) \}$$

$$+ T_n T_e) + (1/4) T_e \wedge T_e \}$$

と展開出来る。ここで次の補題を利用する。

[補題] 実行列 S ($p \times p$) が絶対値が充分小なる ε に対して

$$S = \Lambda + \varepsilon V^{(1)} + \varepsilon^2 V^{(2)} + \dots$$

に展開されている。但し $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ $\lambda_1 > \dots > \lambda_p$
 $V^{(j)}$ ($p \times p$) は実行列、 λ_α が単根の時、 S の α 番目の固有値 d_α
 $(\lambda_1 > \dots > \lambda_p)$ の擾動は

$$(1.2) \quad d_\alpha = \lambda_\alpha + \varepsilon v_{\alpha\alpha}^{(1)} + \varepsilon^2 \left(v_{\alpha\alpha}^{(2)} - \sum_{j \neq \alpha} v_{\alpha j}^{(1)} \lambda_{j\alpha} v_{j\alpha}^{(1)} \right) + \varepsilon^3 \left\{ v_{\alpha\alpha}^{(3)} \right. \\
- \sum_{j \neq \alpha} \lambda_{j\alpha} (v_{\alpha j}^{(1)} v_{j\alpha}^{(1)} + v_{\alpha j}^{(2)} v_{j\alpha}^{(1)}) - v_{\alpha\alpha}^{(1)} \sum_{j \neq \alpha} \lambda_{j\alpha}^2 v_{\alpha j}^{(1)} v_{j\alpha}^{(1)} \\
\left. + \sum_{j \neq \alpha} \sum_{k \neq \alpha} \lambda_{j\alpha} \lambda_{k\alpha} v_{\alpha k}^{(1)} v_{k\alpha}^{(1)} v_{j\alpha}^{(1)} \right\} + o(\varepsilon^4)$$

但し、 $\lambda_{j\alpha} = (\lambda_j - \lambda_\alpha)^{-1}$, $v_{ij}^{(k)}$ は $V^{(k)}$ の (i, j) 要素。

と表わされる。

この補題を $S_n S_e^{-1}$ の展開式 (1.1) に利用して次の定理を得る。

[定理 1.1] θ_α が単根の時、 $S_n S_e^{-1}$ の α 番目の固有値 d_α の周辺分布の漸近展開は

$$(1.3) \quad P \left((e k_\alpha^2 n / 2)^{\frac{1}{2}} \{ d_\alpha - (k/e + (2/e) \theta_\alpha) \} \leq x \right) \\
= \Phi(x) + (1/\sqrt{n}) \{ a_1 \Phi'(x) + a_3 \Phi'''(x) \} + o(n^{-1})$$

但し、 $k_\alpha = e \{ (1 + 2\theta_\alpha)^2 - e \}^{-1/2}$

$$a_1 = (-1/2\sqrt{2e}) k_\alpha R_\alpha$$

$$a_3 = (-2/3\sqrt{2e}) k_\alpha S_\alpha$$

$$R_\alpha = 2(p+1)\{(1+2\theta_\alpha)\bar{e}^2 - 1\} + \sum_{j \neq \alpha}^p \frac{(1+2\theta_\alpha)(1+2\theta_j) - e}{e(\theta_\alpha - \theta_j)}$$

$$S_\alpha = k_\alpha^2 \bar{e}^{-3} \{2(1+2\theta_\alpha)^3 - 3e(1+2\theta_\alpha) + e^2\}$$

である。

[定理 1.2] $\theta_1 > \dots > \theta_p$ の時、 $S_n S_e^{-1}$ の固有値の結合分布の漸近展開は

$$(1.4) \quad P\left(\bigcap_{\alpha=1}^p (ek_\alpha^2 n/2)^{\frac{1}{2}} \{d_\alpha - (k/e + (2/e)\theta_\alpha)\} \leq \lambda_\alpha\right) \\ = \prod_{\alpha=1}^p \Phi(\lambda_\alpha) (1 + (1/\sqrt{n})K_1 + O(n^{-2}))$$

$$\text{但し } \psi^{(j)}(x) = \Phi^{(j)}(x) / \Phi(x)$$

$$K_1 = (-1/\sqrt{2e}) \sum_{\alpha=1}^p k_\alpha \left\{ \frac{1}{2} R_\alpha \psi^{(1)}(x_\alpha) + \frac{2}{3} S_\alpha \psi^{(3)}(x_\alpha) \right\}$$

と表わされる。

§ 2. MANOVA モデルにおける種の検定統計量の漸近展開

次の多変量線型モデル (MANOVA) を考える。

$$X (N \times p) = A (N \times m) \Omega (m \times p) + E (N \times p)$$

但し、 A は既知の行列で $\text{rank}(A) = m$ 、 Ω は未知行列、誤差行列 E の各行は互いに独立で $N_p(0, \Sigma)$ に従うものとする。

仮設検定問題

$$H_{01}: \text{rank}(B\Omega) = k, \quad H_{11}: \text{rank}(B\Omega) > k$$

$$(B: k \times m \text{ の既知行列で } \text{rank}(B) = k)$$

に対する代表的統計量

ii) 尤度比統計量 : $\Lambda_k = \prod_{\alpha=k+1}^p (1+d_\alpha)^{-\frac{z}{2}}$

iii) Hotelling 統計量 : $T_k = \sum_{\alpha=k+1}^p d_\alpha$

iv) Pillai 統計量 : $V_k = \sum_{\alpha=k+1}^p d_\alpha / (1+d_\alpha)$

の分布を調べる。但し、 $d_1 > \dots > d_p$ は $S_n S_e^{-1}$ の固有値で、 S_e は $W_p(\gamma, I_p)$, S_n は $W_p(\alpha, I, \Omega)$ に従うとしてよい。 Ω は diagonal, $\gamma = N - m$ 。 α を固定し $\gamma \rightarrow \infty$ とした時の上記の統計量の漸近展開は Fujikoshi (1974) によって $\gamma^{-\frac{1}{2}}$ の項まで与られている。ここでは、 $\gamma = n_e$, $\alpha = n_n$, $\Omega = n \textcircled{4}$ ($n = n_e + n_n$, $\textcircled{4} = \text{diag}(\alpha_1, \dots, \alpha_p)$) とする。

(Nonnull case) $\alpha_{k+1}, \dots, \alpha_p$ を単根とする。尤度比統計量 Λ_k について考える。 $\alpha_{k+1}, \dots, \alpha_p$ が単根であるから §1 の [補題] が利用出来て、 Λ_k は次の様に展開出来る。

$$\begin{aligned} \tilde{\Lambda}_k &= \sqrt{n} \left(- (2/n) \log \Lambda_k - \sum_{\alpha=k+1}^p \log(1+\lambda_\alpha) \right) \\ &= \sum_{\alpha=k+1}^p \frac{\lambda_\alpha^{(1)}}{1+\lambda_\alpha} + \frac{1}{\sqrt{n}} \sum_{\alpha=k+1}^p \left(\frac{\lambda_\alpha^{(2)}}{1+\lambda_\alpha} - \frac{\lambda_\alpha^{(1)2}}{2(1+\lambda_\alpha)^2} \right) + O(\bar{n}^{-1}). \end{aligned}$$

従って $\tilde{\Lambda}_k$ の特性関数は

$$\begin{aligned} E \exp it \sum_{\alpha=k+1}^p \frac{\lambda_\alpha^{(1)}}{1+\lambda_\alpha} &\left[1 + \frac{1}{\sqrt{n}} \sum_{\alpha=k+1}^p \left(\frac{\lambda_\alpha^{(2)}}{1+\lambda_\alpha} - \frac{\lambda_\alpha^{(1)2}}{2(1+\lambda_\alpha)^2} \right) \right. \\ &\left. + O(\bar{n}^{-1}) \right]. \end{aligned}$$

期待値を求めて反転すれば次の結果を得る。

[定理 2.1] $\alpha_{k+1}, \dots, \alpha_p$ が単根の時、尤度比統計量 Λ_k の分布

の漸近展開は

$$(2.1) \quad P\left(\frac{\sqrt{n}}{\tau_1} \left\{ -\frac{2}{n} \log \Lambda_k - \sum_{\alpha=k+1}^p \log \frac{1+2\theta_\alpha}{e} \right\} \leq x\right) \\ = \Phi(x) - \frac{1}{\sqrt{n}} \left\{ \frac{a_1}{\tau_1} \Phi'(x) + \frac{a_3}{\tau_1^3} \Phi^{(3)}(x) \right\} + O(n^{-1}).$$

$$\text{但し. } \tau_1^2 = (2/e) \{ (p-k) - e\Lambda_2 \}$$

$$a_1 = (1/2) \{ (p-k)(p+k+1)\bar{e}^{-1} - 2(p+1)\Lambda_1 + \Lambda_1^2 \\ + \Lambda_2 + \sum_{\alpha=k+1}^p \sum_{j=1}^k \frac{(1+2\theta_\alpha)(1+2\theta_j) - e}{e(\theta_\alpha - \theta_j)(1+2\theta_\alpha)} \}$$

$$a_3 = (2/3e^2) \{ (p-k) + 2e^2\Lambda_3 - 3e^2\Lambda_4 \}$$

$$\Lambda_i = \sum_{\alpha=k+1}^p (1+2\theta_\alpha)^{-i}.$$

同様の方法で, Hotelling 統計量 T_k と Pillai 統計量 V_k の漸近展開が得られる。

[定理 2.2] $\theta_{k+1}, \dots, \theta_p$ が単根の時, Hotelling 統計量 T_k と Pillai 統計量 V_k の分布の漸近展開は次の様になる。

$$(2.2) \quad P\left(\frac{\sqrt{n}}{\tau_2} \left\{ T_k - \sum_{\alpha=k+1}^p (k+2\theta_\alpha)\bar{e}^{-1} \right\} \leq x\right) \\ = \Phi(x) - \frac{1}{\sqrt{n}} \left\{ \frac{b_1}{\tau_2} \Phi'(x) + \frac{b_3}{\tau_2^3} \Phi^{(3)}(x) \right\} + O(n^{-1}).$$

$$\text{但し. } \tau_2^2 = (2/e^3) \{ t_2 - e(p-k) \}$$

$$b_1 = (1/2e^2) \{ 2(p+1)t_1 - 2(p+1)e(p-k) \}$$

$$+ \sum_{\alpha=k+1}^p \sum_{j=1}^k \frac{(1+2\theta_\alpha)(1+2\theta_j) - e}{\theta_\alpha - \theta_j} \}$$

$$t_2 = (4/3e^5) \{ e^2(p-k) - 3et_1 + 2t_3 \}$$

$$t_i = \sum_{\alpha=k+1}^p (1+2\theta_\alpha)^i.$$

そして

$$(2.3) \quad P\left(\frac{\sqrt{n}}{t_3} \left\{ V_k - \sum_{\alpha=k+1}^p \frac{k+2\theta_\alpha}{1+2\theta_\alpha} \right\} \leq x\right) \\ = \Phi(x) - \frac{1}{\sqrt{n}} \left\{ \frac{C_2}{t_3} \Phi''(x) + \frac{C_3}{t_3} \Phi'''(x) \right\} + O(n^{-1}).$$

$$\text{但し, } t_3^2 = 2e(\lambda_2 - e\lambda_4)$$

$$C_1 = e \{ k\bar{e}^{-1}\lambda_1 - (p+1)\lambda_2 + \lambda_2\lambda_2 + \lambda_3 \\ + \frac{1}{2e} \sum_{\alpha=k+1}^p \sum_{j=1}^k \frac{(1+2\theta_\alpha)(1+2\theta_j) - e}{(\theta_\alpha - \theta_j)(1+2\theta_\alpha)^2} \}$$

$$C_3 = (4e/3) \{ -\lambda_3 + 3e\lambda_5 + e^2\lambda_6 - 3e^2\lambda_7 \}.$$

(null case). $\theta_1 \geq \dots \geq \theta_k > \theta_{k+1} = \dots = \theta_p = 0$ とする。重根のある場合の擾動公式は既に Lawley (1956) によって得られているが、ここでは、Lawley の手法によって Fujikoshi (1974) が得た擾動公式を利用して、統計量の漸近展開を求めた。

[補題] $S = \Lambda + \varepsilon V^{(0)} + \varepsilon^2 V^{(1)} + \dots$ と展開され、さらに

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_{\alpha-1} & \\ & & & \cup \\ & & & & \lambda_{\alpha} I_{(p-\alpha+1)} \end{pmatrix} \begin{pmatrix} \lambda_1 > \dots > \lambda_{\alpha} \\ I_{(p-\alpha+1)}: \text{identity} \end{pmatrix}$$

で且つ、 $V^{(j)}$ ($j=0, 1, \dots$) は $p \times p$ の実対称行列で、 Λ と同じ order で

$$V^{(j)} = \begin{pmatrix} V_{11} & \dots & V_{1,\alpha} \\ \vdots & & \vdots \\ V_{\alpha,1} & \dots & V_{\alpha,\alpha} \end{pmatrix}$$

と分割されている。この時 S の i 番目 ($i=\alpha, \dots, p$) の固有値は以下の様に定義された Z の $(i-\alpha+1)$ 番目の固有値である。但し、

$$\begin{aligned} (2.4) \quad Z = & \lambda_{\alpha} I_{(p-\alpha+1)} + \varepsilon V_{\alpha,\alpha}^{(0)} + \varepsilon^2 \left[V_{\alpha,\alpha}^{(1)} + \sum_{j=1}^{\alpha-1} \lambda_{\alpha j} V_{\alpha j}^{(0)} V_{j\alpha}^{(0)} \right] \\ & + \varepsilon^3 \left[V_{\alpha,\alpha}^{(2)} + \sum_{j=1}^{\alpha-1} \lambda_{\alpha j} \{ V_{\alpha j}^{(0)} V_{j\alpha}^{(1)} + V_{\alpha j}^{(1)} V_{j\alpha}^{(0)} \} \right] \\ & + \sum_{j=1}^{\alpha-1} \sum_{r=1}^{\alpha-1} \lambda_{\alpha j} \lambda_{\alpha r} V_{\alpha j}^{(0)} V_{jr}^{(0)} V_{r,\alpha}^{(0)} \\ & - (1/2) \left\{ \sum_{j=1}^{\alpha-1} \lambda_{\alpha j}^2 V_{\alpha j}^{(0)} V_{j\alpha}^{(0)} \right\} V_{\alpha,\alpha}^{(0)} \\ & - (1/2) V_{\alpha,\alpha}^{(0)} \left\{ \sum_{j=1}^{\alpha-1} \lambda_{\alpha j}^2 V_{\alpha j}^{(0)} V_{j\alpha}^{(0)} \right\} \right] + O(\varepsilon^4). \end{aligned}$$

この補題を利用して尤度比統計量 Λ_k の展開は、 $\alpha = k+1$ と置いて

$$\begin{aligned} \hat{\Lambda}_k &= \sqrt{n} \left\{ -\frac{2}{n} \log \Lambda_k - (p-k) \log (1 + \lambda_{k+1}) \right\} \\ &= \frac{\text{tr} Z^{(1)}}{1 + \lambda_{k+1}} + \frac{1}{\sqrt{n}} \left\{ \frac{\text{tr} Z^{(2)}}{1 + \lambda_{k+1}} - \frac{\text{tr} Z^{(1)2}}{2(1 + \lambda_{k+1})^2} \right\} + O(n^{-1}) \end{aligned}$$

$$\text{但し, } Z^{(1)} = V_{k+1, k+1}^{(0)} \\ Z^{(2)} = V_{k+1, k+1}^{(2)} + \sum_{j=1}^k \lambda_{k+1, j} V_{k+1, j}^{(0)} V_{j, k+1}^{(0)}$$

従って $\tilde{\Lambda}_k$ の特性関数は

$$E \exp \frac{it \operatorname{tr} Z^{(1)}}{1 + \lambda_{k+1}} \left[1 + \frac{it}{\sqrt{n}} \left\{ \frac{\operatorname{tr} Z^{(2)}}{1 + \lambda_{k+1}} - \frac{\operatorname{tr} Z^{(1)2}}{2(1 + \lambda_{k+1})^2} \right\} + o(\bar{n}) \right]$$

である。特性関数の期待値を求め反転して次の定理を得る。

[定理 2.3] $\theta_1 > \dots > \theta_k > \theta_{k+1} = \dots = \theta_p = 0$ の時、尤度比統計量 Λ_k の分布の漸近展開は次の様になる。

$$(2.5) \quad P\left(\frac{\sqrt{n}}{c_1} \left\{ -\frac{2}{n} \log \Lambda_k + (p-k) \log e \right\} \leq x\right) \\ = \Phi(x) - \frac{1}{\sqrt{n}} \left\{ \frac{a_2}{c_1} \Phi^{(1)}(x) + \frac{a_3}{c_1^3} \Phi^{(3)}(x) \right\} + o(\bar{n}^{-2})$$

$$\text{但し, } c_1^2 = (2/e)(p-k)h$$

$$a_1 = \frac{(p-k)}{2e} \left\{ (p-k+1)h - 2k - h \sum_{j=1}^k \frac{1}{\theta_j} \right\}$$

$$a_3 = (2/3e^2)(p-k)(1+e)h.$$

同様の手法で

[定理 2.4] $\theta_1 > \dots > \theta_k > \theta_{k+1} = \dots = \theta_p = 0$ の時、Hotelling 統計量 T_k 及び Pillai 統計量 V_k の分布の漸近展開は次の様になる。

$$(2.6) \quad P\left(\frac{\sqrt{n}}{c_2} \left\{ T_k - \frac{(p-k)h}{e} \right\} \leq x\right) \\ = \Phi(x) - \frac{1}{\sqrt{n}} \left\{ \frac{a_1}{c_2} \Phi^{(1)}(x) + \frac{a_3}{c_2^3} \Phi^{(3)}(x) \right\} + o(\bar{n}^{-2}).$$

$$\text{但し. } T_2^2 = (2/e^3)(p-k)h$$

$$b_1 = \frac{p-k}{2e^2} \left\{ 2(p+1)h - 2k - h \sum_{j=1}^k \frac{1}{\theta_j} \right\}$$

$$b_3 = (4/3e^5)(2-e)(p-k)h.$$

そして

$$\begin{aligned} (2.7) \quad P\left(\frac{\sqrt{n}}{T_3} \{V_k - (p-k)h\} \leq x\right) \\ = \Phi(x) - \frac{1}{\sqrt{n}} \left\{ \frac{C_1}{T_3} \Phi''(x) + \frac{C_3}{T_3^3} \Phi^{(3)}(x) \right\} + O(n^{-1}) \end{aligned}$$

$$\text{但し. } T_3^2 = 2e(p-k)h$$

$$C_1 = \frac{(p-k)}{2} \left\{ 2kh - 2k - h \sum_{j=1}^k \frac{1}{\theta_j} \right\}$$

$$C_3 = (4eh/3)(p-k)(e-k).$$

定理2.1 ~ 2.4 において $k=0$ と置けば 各 $n^{\frac{1}{2}}$ の係数は Fujikoshi (1974[2]) の論文と一致する。 n^1 の係数は修士論文で与えてある。また [定理2.1] 及び [定理2.2] で

$\theta_{k+1} = \dots = \theta_p = 0$ と置けば [定理2.3] 及び [定理2.4] の結果が得られることを注意して置く。

References.

- [1] Anderson, T.W. (1951) The asymptotic distribution of certain characteristic roots and vectors.
2-nd Berkeley Symposium
- [2] Fujikoshi, Y. (1974) Asymptotic formulae for the non null distributions of three statistics for multivariate linear hypothesis.
(to appear).
- [3] Fujikoshi, Y. (1974). The asymptotic expansions for certain test statistics in multivariate analysis. (数理研).
- [4] Lawley, D.N (1956) Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* 43. 128-136.