数理解析研究所講究録
第 247 巻 1975 年 1-11

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) ・・・・・・ $|S_h - \lambda S_e| = 0$

の固有値の分布を調べることが必要になる（MANOVA ~~etc~~）。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

$$S_e \sim W_p(n_e, I_p) \quad , S_h \sim W_p(n_h, I_p, \Omega)$$
$$\Omega : \text{diagonal matrix}$$

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： $n_e : n_h = e : h$ （$e>0, h>0, e+h=1$）で $n = n_e + n_h \to \infty$ とする。かつ、$\Omega = n\textcircled{\scriptsize ω}$とおく。

この仮定のsituationが時に必要であることがFujikoshi(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$l_1 > \cdots > l_p$とする。目標は$(l_1, \cdots, l_p)$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$n \to \infty$の時 $(h/e) I_p + (2/e)\textcircled{4}$ に確率収束するのでl_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。

$$T_e = \sqrt{n}(S_e/n - e I_p), \quad T_h = \sqrt{n}\{S_h/n - (h I_p + 2\textcircled{4})\}$$

と置けば、$n \to \infty$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。

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

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

$$V^{(0)} = e^{-1}\{T_h - (1/2)(T_e \Lambda + \Lambda T_e)\}$$
$$V^{(1)} = e^{-2}\{(3/8)(T_e^2 \Lambda + \Lambda T_e^2) - (1/2)(T_e T_h$$

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

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 $S = \Lambda + \varepsilon V^{(0)} + \varepsilon^2 V^{(1)} + \cdots$

に展開されている。但し $\Lambda = diag(\lambda_1, \cdots, \lambda_p)$ $\lambda_1 \geqq \cdots \geqq \lambda_p$ $V^{(j)}$（$p \times p$）は実行列。λ_α が単根の時、S の α 番目の固有値 d_α（$d_1 \geqq \cdots \geqq d_p$）の摂動は

$$(1.2)\quad d_\alpha = \lambda_\alpha + \varepsilon v_{\alpha\alpha}^{(0)} + \varepsilon^2 \Big(v_{\alpha\alpha}^{(1)} - \sum_{j \neq \alpha}^{p} v_{\alpha j}^{(0)} \lambda_{j\alpha} v_{j\alpha}^{(0)} \Big) + \varepsilon^3 \Big\{ v_{\alpha\alpha}^{(2)} - \sum_{j \neq \alpha}^{p} \lambda_{j\alpha} \big(v_{\alpha j}^{(1)} v_{j\alpha}^{(0)} + v_{\alpha j}^{(0)} v_{j\alpha}^{(1)} \big) - v_{\alpha\alpha}^{(0)} \sum_{j \neq \alpha}^{p} \lambda_{j\alpha}^2 v_{\alpha j}^{(0)} v_{j\alpha}^{(0)} + \sum_{j \neq \alpha}^{p} \sum_{k \neq \alpha}^{p} \lambda_{j\alpha} \lambda_{k\alpha} v_{\alpha k}^{(0)} v_{kj}^{(0)} v_{j\alpha}^{(0)} \Big\} + O(\varepsilon^4)$$

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 $S_h S_e^{-1}$ 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、$S_h S_e^{-1}$ の α 番目の固有値 d_α の周辺分布の漸近展開は

$$(1.3)\quad P\Big((e k_\alpha^2 n/2)^{\frac{1}{2}} \{ d_\alpha - (h/e + (2/e)\theta_\alpha) \} \leqq x \Big) = \Phi(x) + (1/\sqrt{n}) \{ a_1 \Phi^{(1)}(x) + a_3 \Phi^{(3)}(x) \} + O(n^{-1})$$

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

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

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

$$R_\alpha = 2(p+1)\{(1+2\theta_\alpha)e^{-2}-1\} + \sum_{j\neq\alpha}^{p} \frac{(1+2\theta_\alpha)(1+2\theta_j)-e}{e(\theta_\alpha-\theta_j)}$$
$$S_\alpha = k_\alpha^2 e^{-3}\{2(1+2\theta_\alpha)^3 - 3e(1+2\theta_\alpha) + e^2\}$$

である。

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

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

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3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。

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

但し、Aは既知の行列で $rank(A)=m$, Ξ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

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

(B: $b\times m$ の既知行列で $rank(B) = b$)

に対する代表的統計量

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

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

(iii) 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、$d_1 > \cdots > d_p$ は $S_h S_e^{-1}$ の固有値で、S_e は $W_p(g, I_p)$, S_h は $W_p(b, I, \Omega)$ に従うとしてよい。Ω は diagonal, $g = N - m$。b を固定し $g \to \infty$ とした時の上記の統計量の漸近展開は Fujikoshi (1974) によって $g^{-\frac{1}{2}}$ の項まで得られている。ここでは、$g = n_e$, $b = n_h$, $\Omega = n\Theta$ ($n = n_e + n_h$, $\Theta = \mathrm{diag}(\theta_1, \cdots, \theta_p)$) とする。

(Nonnull case) $\theta_{k+1}, \cdots, \theta_p$ を単根とする。尤度比統計量 Λ_k について考える。$\theta_{k+1}, \cdots, \theta_p$ 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。

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

従って $\tilde{\Lambda}_k$ の特性関数は

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

期待値を求めて反転すれば次の結果を得る。

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

の漸近展開は

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

但し．$\tau_1^2=(2/e)\{(p-k)-eA_2\}$

$$a_1=(1/2)\{(p-k)(p+k+1)e^{-1}-2(p+1)A_1+A_1^2$$
$$+A_2+\sum_{\alpha=k+1}^{p}\sum_{j=1}^{k}\frac{(1+2\theta_\alpha)(1+2\theta_j)-e}{e(\theta_\alpha-\theta_j)(1+2\theta_\alpha)}\}$$
$$a_3=(2/3e^2)\{(p-k)+2e^2A_3-3e^2A_4\}$$
$$A_i=\sum_{\alpha=k+1}^{p}(1+2\theta_\alpha)^{-i}.$$

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］ $\theta_{k+1},\cdots,\theta_p$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) $$P\left(\frac{\sqrt{n}}{\tau_2}\left\{T_k-\sum_{\alpha=k+1}^{p}(h+2\theta_\alpha)e^{-1}\right\}\leq x\right)$$
$$=\Phi(x)-\frac{1}{\sqrt{n}}\left\{\frac{b_1}{\tau_2}\Phi^{(1)}(x)+\frac{b_3}{\tau_2^3}\Phi^{(3)}(x)\right\}+O(n^{-1}).$$

但し．$\tau_2^2=(2/e^3)\{t_2-e(p-k)\}$

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

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

$$b_2=(4/3e^5)\{e^2(p-k)-3et_1+2t_3\}$$

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

そして

$$(2.3)\quad P\Big(\frac{\sqrt{n}}{\tau_3}\Big\{V_k-\sum_{\alpha=k+1}^{p}\frac{k+2\theta_\alpha}{1+2\theta_\alpha}\Big\}\le x\Big)$$
$$=\Phi(x)-\frac{1}{\sqrt{n}}\Big\{\frac{c_2}{\tau_3}\Phi^{(1)}(x)+\frac{c_3}{\tau_3^3}\Phi^{(3)}(x)\Big\}+O(n^{-1})$$

但し、$\tau_3^2=2e(\Delta_2-e\Delta_4)$

$$c_1=e\Big\{ke^{-1}\Delta_1-(p+1)\Delta_2+\Delta_1\Delta_2+\Delta_3$$
$$+\frac{1}{2e}\sum_{\alpha=k+1}^{p}\sum_{j=1}^{k}\frac{(1+2\theta_\alpha)(1+2\theta_j)-e}{(\theta_\alpha-\theta_j)(1+2\theta_\alpha)^2}\Big\}$$

$$c_3=(4e/3)\{-\Delta_3+3e\Delta_5+e^2\Delta_6-3e^2\Delta_7\}。$$

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

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

$$\Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \ddots & & \\ & & \lambda_{\alpha-1} & \\ 0 & & & \lambda_\alpha I_{(p-\alpha+1)} \end{pmatrix} \quad \begin{pmatrix} \lambda_1 > \cdots > \lambda_\alpha \\ I_{(p-\alpha+1)} : \text{identity} \end{pmatrix}$$

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

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

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

$$\begin{aligned}(2.4)\quad Z = {} & \lambda_\alpha I_{(p-\alpha+1)} + \varepsilon V^{(0)}_{\alpha,\alpha} + \varepsilon^2\Big[V^{(1)}_{\alpha,\alpha} + \sum_{j=1}^{\alpha-1}\lambda_{\alpha j}V^{(0)}_{\alpha j}V^{(0)}_{j\alpha}\Big] \\ & + \varepsilon^3\Big[V^{(2)}_{\alpha\alpha} + \sum_{j=1}^{\alpha-1}\lambda_{\alpha j}\big\{V^{(0)}_{\alpha j}V^{(1)}_{j\alpha} + V^{(1)}_{\alpha j}V^{(0)}_{j\alpha}\big\} \\ & + \sum_{j=1}^{\alpha-1}\sum_{r=1}^{\alpha-1}\lambda_{\alpha j}\lambda_{\alpha r}V^{(0)}_{\alpha j}V^{(0)}_{jr}V^{(0)}_{r,\alpha} \\ & - (1/2)\Big\{\sum_{j=1}^{\alpha-1}\lambda^2_{\alpha j}V^{(0)}_{\alpha j}V^{(0)}_{j\alpha}\Big\}V^{(0)}_{\alpha\alpha} \\ & - (1/2)V^{(0)}_{\alpha\alpha}\Big\{\sum_{j=1}^{\alpha-1}\lambda^2_{\alpha j}V^{(0)}_{\alpha j}V^{(0)}_{j\alpha}\Big\}\Big] + O(\varepsilon^4)。\end{aligned}$$

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

$$\begin{aligned}\tilde{\Lambda}_k &= \sqrt{n}\Big\{-\frac{2}{n}\log\Lambda_k - (p-k)\log(1+\lambda_{k+1})\Big\} \\ &= \frac{\mathrm{tr}\,Z^{(1)}}{1+\lambda_{k+1}} + \frac{1}{\sqrt{n}}\Big\{\frac{\mathrm{tr}\,Z^{(2)}}{1+\lambda_{k+1}} - \frac{\mathrm{tr}\,Z^{(1)2}}{2(1+\lambda_{k+1})^2}\Big\} + O(n^{-1})\end{aligned}$$

但し $Z^{(1)} = V^{(0)}_{k+1,k+1}$

$$Z^{(2)} = V^{(2)}_{k+1,k+1} + \sum_{j=1}^{k} \lambda_{k+1,j} V^{(0)}_{k+1,j} V^{(0)}_{j,k+1}$$

従って $\hat{\Lambda}_k$ の特性関数は

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

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] $\theta_1 \geq \cdots \geq \theta_k > \theta_{k+1} = \cdots = \theta_p = 0$ 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。

$$(2.5)\quad P\left(\frac{\sqrt{n}}{\tau_1}\left\{-\frac{2}{n}\log\Lambda_k+(p-k)\log e\right\}\leq x\right)$$
$$=\Phi(x)-\frac{1}{\sqrt{n}}\left\{\frac{a_1}{\tau_1}\Phi^{(1)}(x)+\frac{a_3}{\tau_1^3}\Phi^{(3)}(x)\right\}+O(n^{-1})$$

但し $\tau_1^2 = (2/e)(p-k)h$

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

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

同様の方法で

[定理2.4] $\theta_1 \geq \cdots \geq \theta_k > \theta_{k+1} = \cdots = \theta_p = 0$ 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。

$$(2.6)\quad P\left(\frac{\sqrt{n}}{\tau_2}\left\{T_k-\frac{(p-k)h}{e}\right\}\leq x\right)$$
$$=\Phi(x)-\frac{1}{\sqrt{n}}\left\{\frac{b_1}{\tau_2}\Phi^{(1)}(x)+\frac{b_3}{\tau_2^3}\Phi^{(3)}(x)\right\}+O(n^{-1}).$$

但し. $\tau_2^2 = (2/e^3)(p-k)h$

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

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

そして

$$(2.7)\quad P\left(\frac{\sqrt{n}}{\tau_3}\{V_k-(p-k)h\}\le x\right) = \Phi(x) - \frac{1}{\sqrt{n}}\left\{\frac{c_1}{\tau_3}\Phi^{(1)}(x)+\frac{c_3}{\tau_3^3}\Phi^{(3)}(x)\right\} + O(n^{-1})$$

但し. $\tau_3^2 = 2e(p-k)h$

$$c_1 = \frac{(p-k)}{2}\left\{2kh - 2k - h\sum_{j=1}^{k}\frac{1}{\theta_j}\right\}$$

$$c_3 = (4eh/3)(p-k)(e-h)。$$

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$\theta_{k+1}=\cdots=\theta_p=0$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 Berkley 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.