非正規母集団における Behrens-Fisher 問題について

広島大・理　松本智恵子
広島大・原医研　冨田　哲治
筑波大・システム情報工　柳原　宏和

概要：共分散行列が等しくない二つの母集団について，それらの平均が等しいかどうかを検定する Behrens-Fisher 問題を考える．この問題を検定する検定統計量の帰無分布は，母集団の正規性の仮定が満たされていないときでも，カイ２乗分布に収束することが知られているが，標本数が小さいときにはその近似の精度は良くない．本発表では，非正規性のもとで，この検定統計量の帰無分布の漸近展開の導出をおこない，その方法について議論する．

1. 導入

平均 $\boldsymbol{\mu}_j$，共分散行列 $\boldsymbol{\Sigma}_j$ をもつ 2 つの母集団 Π_j $(j=1,2)$ があり，j 番目の母集団に属する i 番目の個体に対して p 個の変量が観測され，そのような個体の観測値ベクトル $\boldsymbol{y}_{ij}=(y_{ij1},\ldots,y_{ijp})'$ が n_j 個観測されたとする $(i=1,\cdots,n_j)$．このとき，2 つの母集団の平均が等しいかどうか，すなわち，仮説

$$H_0:\ \boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \text{ vs } H_1:\ \boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2 \tag{1.1}$$

を検定する問題である，Behrens-Fisher 問題を考える．この問題を検定する検定統計量は，

$$T_c=(\bar{\boldsymbol{y}}_1-\bar{\boldsymbol{y}}_2)'\left(\frac{\boldsymbol{S}_1}{n_1}+\frac{\boldsymbol{S}_2}{n_2}\right)^{-1}(\bar{\boldsymbol{y}}_1-\bar{\boldsymbol{y}}_2), \tag{1.2}$$

という形で表される． 但し，$n=n_1+n_2$ であり，

$$\bar{\boldsymbol{y}}_j=\frac{1}{n_j}\sum_{i=1}^{n_j}\boldsymbol{y}_{ij},\quad \boldsymbol{S}_j=\frac{1}{n_j-1}\sum_{i=1}^{n_j}\left(\boldsymbol{y}_{ij}-\bar{\boldsymbol{y}}_j\right)\left(\boldsymbol{y}_{ij}-\bar{\boldsymbol{y}}_j\right)',\ (j=1,2), \tag{1.3}$$

であるとする．なお，Π_1 と Π_2 は同一分布に従うとは限らないことに注意する．

$n_1=n_2$ のとき，この検定統計量は 2 母集団での Hotelling's T^2 検定統計量と同値であり，両方の観測値が正規分布に従いかつ $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2$ であれば，$(n-p-1)T_c/p(n-2)$ は自由度 p，$n-p-1$ の F 分布に従うことが知られている．より一般的な条件，すなわち，$n_1\neq n_2$ や $\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_2$ であるときには，Nel et al. (1990) により T_c の正確な帰無分布が正規性の仮定の下で求められている．しかしながら，この帰無分布は複雑で扱いづらいものであるため，より簡単な方法により帰無分布を近似する手法を考える必要がある．検定統計量 T_c の漸近帰無分布は，真の分布がどのような分布であったとしても自由度 p の χ^2-分布になるので，この χ^2 近似を用いて検定を行うことも可能であるが，標本数が小さい場合，この近似の精度は悪くなる．正規性の仮定の下では，Yanagihara and Yuan (2005) 等により，F 近似，バートレット補正，修正バートレット補正などを用いて小標本での近似の精度を改良する手法が提案されているが，非正規性の下での手法はまだ提案されていない．本発表では，Wakaki, Yanagihara and Fujikoshi (2002) にて用いられている手法を利用して，非正規性のもとで検定統計量 T_c の帰無分布の漸近展開の導出をおこない，その方法について議論する．

2005 年 8 月 6～9 日 統計 SummerSeminar2005 発表資料
Mail Address : g052010@hiroshima-u.ac.jp

2. 統計量 T_c の漸近展開とその応用

検定統計量 T_c の帰無分布の漸近展開式は，Wakaki, Yanagihara and Fujikoshi (2002) と同様の手法を用いて導出することができる．すなわち，まず，p 次元ベクトル $\boldsymbol{u}$ を

$$\boldsymbol{u} = \left(\frac{\boldsymbol{S}_1}{n_1} + \frac{\boldsymbol{S}_2}{n_2}\right)^{-1/2}(\bar{\boldsymbol{y}}_1 - \bar{\boldsymbol{y}}_2),$$

とおき，これを用いて $T_c = \boldsymbol{u}'\boldsymbol{u}$ と分解する．次に，観測値ベクトル $\boldsymbol{y}_{ij}$ を $\boldsymbol{\varepsilon}_{ij} = \boldsymbol{\Sigma}_j^{-1/2}(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_j)$ と標準化し，漸近正規性を持つ 4 つの統計量

$$\boldsymbol{z}_j = \frac{\rho_j}{\sqrt{n}}\sum_{i=1}^{n_j}\boldsymbol{\varepsilon}_{ij}, \quad \boldsymbol{V}_j = \frac{\rho_j}{\sqrt{n}}\sum_{i=1}^{n_j}(\boldsymbol{\varepsilon}_{ij}\boldsymbol{\varepsilon}'_{ij} - \boldsymbol{I}_p), \quad (j = 1, 2), \tag{2.1}$$

を用いて $\boldsymbol{u}$ の摂動展開を求める．そして，その摂動展開式と，$\boldsymbol{z}_j$ と $\boldsymbol{V}_j$ の同時特性関数から $\boldsymbol{u}$ の特性関数の展開式を導出し，それを反転することにより $\boldsymbol{u}$ の密度関数を求める．最終的に，$\boldsymbol{u}$ の密度関数を用いて T_c の特性関数の展開式を求め，それを反転することにより，統計量 T_c の帰無分布の漸近展開を得ることができる．

なお，漸近展開を行う際の仮定として，n_j と $\boldsymbol{\varepsilon}_j = \boldsymbol{\varepsilon}_{1j}$（$j = 1, 2$）について次の条件が成り立っているものとする．

A1. $\rho_j = \sqrt{n_j/n} = O(1)$, $(j = 1, 2)$.

B1. $\mathrm{E}(||\boldsymbol{\varepsilon}_j||^8) < \infty$．但し，$||\cdot||$ はユークリッドノルムを表しているものとする．

B2. $\boldsymbol{\varepsilon}_j$ と $\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j$ の同時分布について Cramér 条件が成り立つ，すなわち，任意の $b > 0$ に対し，

$$\sup_{||\boldsymbol{t}'||+||T||>b} |\mathrm{E}[\exp\{i\boldsymbol{t}'\boldsymbol{\varepsilon}_j + itr(\boldsymbol{\varepsilon}'_j T\boldsymbol{\varepsilon}_j)\}]| < 1,$$

が成り立つ．但し，$\boldsymbol{t}$ は p 次元ベクトル，$T = [t_{ab}(1 + \delta_{ab})/2]$ は $p \times p$ 対称行列であり，δ_{ab} は Kronecker のデルタを表しているものとする．

また，$\boldsymbol{\varepsilon}_{ij}$ はそれぞれ独立であるが，$\boldsymbol{\varepsilon}_{11}, \ldots, \boldsymbol{\varepsilon}_{n_1 1}$ と $\boldsymbol{\varepsilon}_{12}, \ldots, \boldsymbol{\varepsilon}_{n_2 2}$ は同一分布に従うとは限らない（j が同じ時には同一分布に従う）．そこで，$\boldsymbol{\varepsilon}_j = (\varepsilon_1^{(j)}, \ldots, \varepsilon_p^{(j)})'$ のモーメントを $\mu_{i_1,\cdots,i_k}^{(j)} = \mathrm{E}(\varepsilon_{i_1}^{(j)} \ldots \varepsilon_{i_k}^{(j)})$ と上添え字で区別する．なお，モーメント $\mu_{i_1,\cdots,i_k}^{(j)}$ とキュミュラント $\kappa_{i_1,\cdots,i_k}^{(j)}$ の関係は

$$\begin{aligned}
\mu_{ab}^{(j)} &= \kappa_{ab}^{(j)} = \delta_{ab}, \\
\mu_{abc}^{(j)} &= \kappa_{abc}^{(j)}, \\
\mu_{abcd}^{(j)} &= \kappa_{abcd}^{(j)} + \sum_{[3]}\delta_{ab}\delta_{cd}, \\
\mu_{abcde}^{(j)} &= \kappa_{abcde}^{(j)} + \sum_{[10]}\kappa_{abc}^{(j)}\delta_{de}, \\
\mu_{abcdef}^{(j)} &= \kappa_{abcdef}^{(j)} + \sum_{[15]}\kappa_{abcd}^{(j)}\delta_{ef} + \sum_{[10]}\kappa_{abc}^{(j)}\kappa_{def}^{(j)} + \sum_{[15]}\delta_{ab}\delta_{cd}\delta_{ef},
\end{aligned}$$

となっている．

2.1. $\boldsymbol{u}$ の摂動展開

(1.3) 式の $\bar{\boldsymbol{y}}_j$ と $\boldsymbol{S}_j$ をそれぞれ $\boldsymbol{z}_j$ と $\boldsymbol{V}_j$ で表すと，

$$
\begin{aligned}
\bar{\boldsymbol{y}}_j &= \frac{\rho_j}{\sqrt{n}}\boldsymbol{\Sigma}_j^{1/2}\boldsymbol{z}_j, && (2.2)\\
\boldsymbol{S}_j &= \boldsymbol{\Sigma}_j^{1/2}\left\{I_p + \frac{\rho_j}{\sqrt{n}}\boldsymbol{V}_j + \frac{\rho_j^2}{n}\left(I_p - \boldsymbol{z}_j\boldsymbol{z}_j'\right)\right\}\boldsymbol{\Sigma}_j^{1/2} + \mathrm{O}_p(n^{-3/2}),
\end{aligned}
$$

となる．これを利用して，$\boldsymbol{u}$ を $\boldsymbol{z}_j$ と $\boldsymbol{V}_j$ で表すと，下の式を得る．

$$
\begin{aligned}
\boldsymbol{u} &= \rho_2^{-1}\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1^{-1}\boldsymbol{K}_2\boldsymbol{z}_2 && (2.3)\\
&-\frac{1}{2\sqrt{n}}\left(\rho_1\rho_2^{-3}\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_2^{-2}\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_2\boldsymbol{z}_2\right.\\
&\qquad\left.+\rho_1^{-2}\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1^{-3}\rho_2\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_2\boldsymbol{z}_2\right)\\
&+\frac{1}{n}\left\{\frac{3}{8}\left(\rho_1^2\rho_2^{-5}\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1\rho_2^{-4}\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_2\boldsymbol{z}_2\right.\right.\\
&\qquad+\rho_1^{-1}\rho_2^{-2}\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_1\boldsymbol{z}_1 + \rho_1^{-1}\rho_2^{-2}\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_1\boldsymbol{z}_1\\
&\qquad-\rho_1^{-2}\rho_2^{-1}\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_2\boldsymbol{z}_2 - \rho_1^{-2}\rho_2^{-1}\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_1\boldsymbol{V}_1\boldsymbol{K}_1\boldsymbol{K}_2\boldsymbol{z}_2\\
&\qquad\left.+\rho_1^{-4}\rho_2\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1^{-5}\rho_2^2\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_2\boldsymbol{V}_2\boldsymbol{K}_2\boldsymbol{K}_2\boldsymbol{z}_2\right)\\
&\quad-\frac{1}{2}\left(\rho_1^2\rho_2^{-3}\boldsymbol{K}_1(I_p-\boldsymbol{z}_1\boldsymbol{z}_1')\boldsymbol{K}_1\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1\rho_2^{-2}\boldsymbol{K}_1(I_p-\boldsymbol{z}_1\boldsymbol{z}_1')\boldsymbol{K}_1\boldsymbol{K}_2\boldsymbol{z}_2\right.\\
&\qquad\left.\left.+\rho_1^{-2}\rho_2\boldsymbol{K}_2(I_p-\boldsymbol{z}_2\boldsymbol{z}_2')\boldsymbol{K}_2\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1^{-3}\rho_2^2\boldsymbol{K}_2(I_p-\boldsymbol{z}_2\boldsymbol{z}_2')\boldsymbol{K}_2\boldsymbol{K}_2\boldsymbol{z}_2\right)\right\} + \mathrm{O}_p(n^{-3/2})\\
&= f_0(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{V}_1,\boldsymbol{V}_2) + \frac{1}{\sqrt{n}}f_1(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{V}_1,\boldsymbol{V}_2) + \frac{1}{n}f_2(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{V}_1,\boldsymbol{V}_2),
\end{aligned}
$$

但し，

$$
\bar{\boldsymbol{\Sigma}}_c = \rho_2^{-2}\boldsymbol{\Sigma}_1 + \rho_1^{-2}\boldsymbol{\Sigma}_2, \quad \boldsymbol{K}_j = \bar{\boldsymbol{\Sigma}}_c^{-1/2}\boldsymbol{\Sigma}_j^{1/2}, \ \ (j=1,2), \tag{2.4}
$$

であるとする．($\boldsymbol{K}_j$ が対称行列であることに注意する)

2.2. $\boldsymbol{u}$ の密度関数

(2.3) 式を利用すると，$\boldsymbol{u}$ の特性関数は以下のように表される．

$$
C_{\boldsymbol{u}}(\boldsymbol{t}) = C_0(\boldsymbol{t}) + \frac{1}{\sqrt{n}}C_1(\boldsymbol{t}) + \frac{1}{n}C_2(\boldsymbol{t}), \tag{2.5}
$$

但し，$\boldsymbol{t}' = (t_1,\dots,t_p)'$ であり，

$$
\begin{aligned}
C_0(\boldsymbol{t}) &= \mathrm{E}\left[\exp\left\{i\boldsymbol{t}'(\rho_2^{-1}\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1^{-1}\boldsymbol{K}_2\boldsymbol{z}_2)\right\}\right],\\
C_1(\boldsymbol{t}) &= i\mathrm{E}_{\boldsymbol{z},\boldsymbol{V}}\left[\boldsymbol{t}'f_1(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{V}_1,\boldsymbol{V}_2)\cdot\exp\left\{i\boldsymbol{t}'(\rho_2^{-1}\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1^{-1}\boldsymbol{K}_2\boldsymbol{z}_2)\right\}\right],\\
C_2(\boldsymbol{t}) &= \mathrm{E}\left[\left\{i\boldsymbol{t}'f_2(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{V}_1,\boldsymbol{V}_2) + \frac{i^2}{2}(\boldsymbol{t}'f_1(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{V}_1,\boldsymbol{V}_2))^2\right\}\cdot\exp\left\{i\boldsymbol{t}'(\rho_2^{-1}\boldsymbol{K}_1\boldsymbol{z}_1 - \rho_1^{-1}\boldsymbol{K}_2\boldsymbol{z}_2)\right\}\right],
\end{aligned}
$$

であるとする．ここで，$\boldsymbol{t}_1 = \rho_2^{-1}\boldsymbol{K}_1\boldsymbol{t}$，$\boldsymbol{t}_2 = -\rho_1^{-1}\boldsymbol{K}_2\boldsymbol{t}$ とおき，$(\boldsymbol{z}_1,\boldsymbol{V}_1)$ と $(\boldsymbol{z}_2,\boldsymbol{V}_2)$ が互いに独立であることと，$(\boldsymbol{z}_j,\boldsymbol{V}_j)$ の同時特性関数

$$
\begin{aligned}
\Psi_j(\boldsymbol{t}_j,T) &= \mathrm{E}\left[\exp\left\{i\mathrm{tr}\left(\boldsymbol{t}_1'\boldsymbol{z}_j + n_j^{-1/2}T\boldsymbol{V}_j\right)\right\}\right], && (2.6)\\
\boldsymbol{t}_j' &= (t_1^{(j)},\dots,t_p^{(j)})', \quad T = \left\{\frac{1}{2}\left(1+\delta_{ab}\right)t_{ab}\right\}, \quad t_{ab}=t_{ba},
\end{aligned}
$$

を用いて期待値の計算を行うと，$\boldsymbol{u}$ の特性関数の漸近展開式を求めることができる．まず，$C_0(\boldsymbol{t})$ については，

$$
\begin{aligned}
C_0(t) &= \Psi_1(\boldsymbol{t}_1,0)\Psi_2(\boldsymbol{t}_2,0) \qquad (2.7)\\
&= \exp\left\{\frac{i^2}{2}\boldsymbol{t}_1'\boldsymbol{t}_1 + \frac{i^3}{6\sqrt{n}}\sum_{abc}^{p} t_a^{(1)}t_b^{(1)}t_c^{(1)}\rho_1^3\kappa_{abc}^{(1)} + \frac{i^4}{24n}\sum_{abcd}^{p} t_a^{(1)}t_b^{(1)}t_c^{(1)}t_d^{(1)}\rho_1^4\kappa_{abcd}^{(1)} + o(n^{-1})\right\}\\
&\quad\cdot\exp\left\{\frac{i^2}{2}\boldsymbol{t}_2'\boldsymbol{t}_2 + \frac{i^3}{6\sqrt{n}}\sum_{abc}^{p} t_a^{(2)}t_b^{(2)}t_c^{(2)}\rho_2^3\kappa_{abc}^{(2)} + \frac{i^4}{24n}\sum_{abcd}^{p} t_a^{(2)}t_b^{(2)}t_c^{(2)}t_d^{(2)}\rho_2^4\kappa_{abcd}^{(2)} + o(n^{-1})\right\}\\
&= \exp\left[\frac{i^2}{2}(\boldsymbol{t}_1'\boldsymbol{t}_1+\boldsymbol{t}_2'\boldsymbol{t}_2) + \frac{i^3}{6\sqrt{n}}\sum_{abc}^{p}\left\{t_a^{(1)}t_b^{(1)}t_c^{(1)}\rho_1^3\kappa_{abc}^{(1)} + t_a^{(2)}t_b^{(2)}t_c^{(2)}\rho_2^3\kappa_{abc}^{(2)}\right\}\right.\\
&\quad\left.+\frac{i^4}{24n}\sum_{abcd}^{p}\left\{t_a^{(1)}t_b^{(1)}t_c^{(1)}t_d^{(1)}\rho_1^4\kappa_{abcd}^{(1)} + t_a^{(2)}t_b^{(2)}t_c^{(2)}t_d^{(2)}\rho_2^4\kappa_{abcd}^{(2)}\right\}\right] + o(n^{-1})\\
&= \exp\left(\frac{i^2}{2}\boldsymbol{t}'\boldsymbol{t}\right)\left[1 + \frac{i^3}{6\sqrt{n}}\sum_{abca'b'c'}^{p} t_{a'}t_{b'}t_{c'}\left\{k_{aa'}^{(1)}k_{bb'}^{(1)}k_{cc'}^{(1)}\rho_1^3\rho_2^{-3}\kappa_{abc}^{(1)} + k_{aa'}^{(2)}k_{bb'}^{(2)}k_{cc'}^{(2)}\rho_1^{-3}\rho_2^3\kappa_{abc}^{(2)}\right\}\right.\\
&\quad\left.+\frac{i^4}{24n}\sum_{abcda'b'c'd'}^{p} t_{a'}t_{b'}t_{c'}t_{d'}\left\{k_{aa'}^{(1)}k_{bb'}^{(1)}k_{cc'}^{(1)}k_{dd'}^{(1)}\rho_1^4\rho_2^{-4}\kappa_{abcd}^{(1)} + k_{aa'}^{(2)}k_{bb'}^{(2)}k_{cc'}^{(2)}k_{dd'}^{(2)}\rho_1^{-4}\rho_2^4\kappa_{abcd}^{(2)}\right\}\right] + o(n^{-1}),
\end{aligned}
$$

と計算することができる．但し，$k_{aa'}^{(j)}$ は (2.4) 式で定義された行列 $\boldsymbol{K}_j$ の (a,a') 成分であるとする．$C_1(\boldsymbol{t})$ と $C_2(\boldsymbol{t})$ については，$(\boldsymbol{z}_j,\boldsymbol{V}_j)$ の同時特性関数の微分を用いた期待値計算

$$
\begin{aligned}
&\mathrm{E}\left[\{i\mathrm{tr}(\boldsymbol{t}_j'K_i^2\boldsymbol{V}_i\boldsymbol{z}_j)\}\exp(i\boldsymbol{t}_i'\boldsymbol{z}_i)\exp(i\boldsymbol{t}_j'\boldsymbol{z}_j)\right] \qquad (2.8)\\
&\quad= -\sum_{abcd}^{p} t_a^{(j)}k_{ab}^{(i)}k_{bc}^{(i)}\left(\frac{\partial}{\partial t_{cd}}\Psi_i(\boldsymbol{t}_i,T)\bigg|_{T=0}\right)\left(\frac{\partial}{\partial t_d^{(j)}}\Psi_j(\boldsymbol{t}_j,T)\bigg|_{T=0}\right),\\
&\mathrm{E}\left[\{i\mathrm{tr}(\boldsymbol{t}_i'K_i^2K_j^2\boldsymbol{V}_i\boldsymbol{V}_j\boldsymbol{z}_i)\}\exp(i\boldsymbol{t}_i'\boldsymbol{z}_i)\exp(i\boldsymbol{t}_j'\boldsymbol{z}_j)\right] \qquad (2.9)\\
&\quad= i\sum_{abcdefg}^{p} t_a^{(i)}k_{ab}^{(i)}k_{bc}^{(i)}k_{cd}^{(j)}k_{de}^{(j)}\left(\frac{\partial^2}{\partial t_g^{(i)}\partial t_{ef}}\Psi_i(\boldsymbol{t}_i,T)\bigg|_{T=0}\right)\left(\frac{\partial}{\partial t_{fg}}\Psi_j(\boldsymbol{t}_j,T)\bigg|_{T=0}\right),\\
&\mathrm{E}\left[\{i\mathrm{tr}(\boldsymbol{t}_j'\boldsymbol{K}_i^2\boldsymbol{z}_j)\}\exp(i\boldsymbol{t}_i'\boldsymbol{z}_i)\exp(i\boldsymbol{t}_j'\boldsymbol{z}_j)\right] \qquad (2.10)\\
&\quad= -i\sum_{abc}^{p} t_a^{(j)}k_{ab}^{(i)}k_{bc}^{(i)}\left(\frac{\partial}{\partial t_c^{(j)}}\Psi_j(\boldsymbol{t}_j,T)\bigg|_{T=0}\right)\Psi_i(\boldsymbol{t}_i,0),\\
&\mathrm{E}\left[\{i\mathrm{tr}(\boldsymbol{t}_j'\boldsymbol{K}_i^2\boldsymbol{z}_i\boldsymbol{z}_i'\boldsymbol{z}_j)\}\exp(i\boldsymbol{t}_i'\boldsymbol{z}_i)\exp(i\boldsymbol{t}_j'\boldsymbol{z}_j)\right] \qquad (2.11)\\
&\quad= i\sum_{abcd}^{p} t_a^{(j)}k_{ab}^{(i)}k_{bc}^{(i)}\left(\frac{\partial^2}{\partial t_c^{(i)}\partial t_d^{(i)}}\Psi_i(\boldsymbol{t}_i,T)\bigg|_{T=0}\right)\left(\frac{\partial}{\partial t_d^{(j)}}\Psi_j(\boldsymbol{t}_j,T)\bigg|_{T=0}\right),\\
&\mathrm{E}\left[\{i\mathrm{tr}((\boldsymbol{t}_j'\boldsymbol{K}_i^2\boldsymbol{V}_i\boldsymbol{z}_j)(\boldsymbol{t}_i'\boldsymbol{K}_j^2\boldsymbol{V}_j\boldsymbol{z}_i))\}\exp(i\boldsymbol{t}_i'\boldsymbol{z}_i)\exp(i\boldsymbol{t}_j'\boldsymbol{z}_j)\right] \qquad (2.12)\\
&\quad= \sum_{abcda'b'c'd'}^{p} t_a^{(j)}t_{a'}^{(i)}k_{ab}^{(j)}k_{bc}^{(j)}k_{a'b'}^{(i)}k_{b'c'}^{(i)}\left(\frac{\partial^2}{\partial t_{d'}^{(i)}\partial t_{c'd'}}\Psi_i(\boldsymbol{t}_i,T)\bigg|_{T=0}\right)\left(\frac{\partial^2}{\partial t_d^{(j)}\partial t_{cd}}\Psi_j(\boldsymbol{t}_j,T)\bigg|_{T=0}\right),
\end{aligned}
$$

を用いて期待値を計算することができる．その結果，$C_1(\boldsymbol{t})$ については

$$
\begin{aligned}
C_1(\boldsymbol{t}) &= \exp\left(\frac{i^2}{2}\boldsymbol{t}'\boldsymbol{t}\right)\left[1 + \frac{i^3}{6\sqrt{n}}\sum_{abca'b'c'}^{p} t_{a'}t_{b'}t_{c'}\left\{k_{aa'}^{(1)}k_{bb'}^{(1)}k_{cc'}^{(1)}\rho_1^3\rho_2^{-3}\kappa_{abc}^{(1)} + k_{aa'}^{(2)}k_{bb'}^{(2)}k_{cc'}^{(2)}\rho_1^{-3}\rho_2^3\kappa_{abc}^{(2)}\right\}\right]\\
&\cdot\frac{1}{\sqrt{n}}\cdot\frac{i}{2}\cdot\left[i^2\sum t_{a'}\left(k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(1)}\rho_1^2\rho_2^{-3}\kappa_{cdd}^{(1)} - k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(2)}\rho_1^{-3}\rho_2^2\kappa_{cdd}^{(2)}\right)\right.\\
&\quad+i^4\sum t_{a'}t_{d'}t_{e'}\left(k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(1)}k_{dd'}^{(1)}k_{ee'}^{(1)}\rho_1^2\rho_2^{-5}\kappa_{cde}^{(1)} - k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(2)}k_{dd'}^{(1)}k_{ee'}^{(1)}\rho_1^{-5}\rho_2^2\kappa_{cde}^{(2)}\right.\\
&\qquad\left.+k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(2)}k_{dd'}^{(2)}k_{ee'}^{(1)}\rho_2^{-3}\kappa_{cde}^{(1)} - k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(1)}k_{dd'}^{(1)}k_{ee'}^{(2)}\rho_1^{-3}\kappa_{cde}^{(2)}\right)
\end{aligned}
$$

$$+\frac{1}{\sqrt{n}}\left\{i^3\sum t_{a'}t_{e'}\left(k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(1)}k_{ee'}^{(1)}\rho_1\rho_2^{-4}(\mu_{cdde}^{(1)}-\delta_{cd}\delta_{de})+k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(2)}k_{ee'}^{(2)}\rho_1^{-4}\rho_2(\mu_{cdde}^{(2)}-\delta_{cd}\delta_{de})\right)\right.$$
$$+\frac{i^5}{2}t_{a'}t_{d'}t_{e'}t_{f'}\left(k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(1)}k_{dd'}^{(1)}k_{ee'}^{(1)}k_{ff'}^{(1)}\rho_1\rho_2^{-6}(\mu_{cdef}^{(1)}-\delta_{cd}\delta_{ef})\right.$$
$$+k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(2)}k_{dd'}^{(2)}k_{ee'}^{(2)}k_{ff'}^{(2)}\rho_1^{-6}\rho_2(\mu_{cdef}^{(2)}-\delta_{cd}\delta_{ef})$$
$$+k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(2)}k_{dd'}^{(2)}k_{ee'}^{(1)}k_{ff'}^{(1)}\rho_1^{-1}\rho_2^{-4}(\mu_{cdef}^{(1)}-\delta_{cd}\delta_{ef})$$
$$\left.+k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(1)}k_{dd'}^{(1)}k_{ee'}^{(2)}k_{ff'}^{(2)}\rho_1^{-4}\rho_2^{-1}(\mu_{cdef}^{(2)}-\delta_{cd}\delta_{ef})\right)$$
$$+\frac{i^5}{2}t_{a'}t_{e'}t_{f'}t_{g'}\left(k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(1)}k_{ee'}^{(1)}k_{ff'}^{(1)}k_{gg'}^{(1)}\rho_1^5\rho_2^{-6}\kappa_{cde}^{(1)}\kappa_{def}^{(1)}+k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(2)}k_{ee'}^{(2)}k_{ff'}^{(2)}k_{gg'}^{(2)}\rho_1^{-6}\rho_2^5\kappa_{cde}^{(2)}\kappa_{def}^{(2)}\right.$$
$$\left.\left.\left.-k_{ab}^{(1)}k_{bc}^{(1)}k_{aa'}^{(1)}k_{ee'}^{(2)}k_{ff'}^{(1)}k_{gg'}^{(1)}\rho_1^{-1}\kappa_{cde}^{(2)}\kappa_{def}^{(1)}-k_{ab}^{(2)}k_{bc}^{(2)}k_{aa'}^{(2)}k_{ee'}^{(1)}k_{ff'}^{(2)}k_{gg'}^{(2)}\rho_1^{-1}\kappa_{cde}^{(1)}\kappa_{def}^{(2)}\right)\right\}\right], \quad (2.13)$$

という結果を得，$C_2(\boldsymbol{t})$ については，その結果は，t_a の 2 次と 4 次と 6 次の項からなる多項式の形になる．

これらを整理し，反転することにより，次の定理を得る．

定理 2.1

条件 *A1*，*B1*，*B2* のもとで，$\boldsymbol{u}$ の密度関数は，漸近的に

$$f(\boldsymbol{u})=\phi_p(\boldsymbol{u})\left\{1+\frac{1}{\sqrt{n}}R_1(\boldsymbol{u})+\frac{1}{n}R_2(\boldsymbol{u})\right\},$$

と表すことができる．但し，

$$R_1(\boldsymbol{u})=\sum H_{a'}(\boldsymbol{u})q_1+\sum H_{a'b'c'}(\boldsymbol{u})q_3,$$
$$R_2(\boldsymbol{u})=\sum H_{a'b'}(\boldsymbol{u})q_2+\sum H_{a'b'c'd'}(\boldsymbol{u})q_4+\sum H_{a'b'c'd'e'f'}(\boldsymbol{u})q_6,$$

であり，q_k は ρ_j，行列 $\boldsymbol{K}_j$ の成分 $k_{ab}^{(j)}$，$\boldsymbol{\varepsilon}_j=\boldsymbol{\varepsilon}_{ij}$ $(j=1,2)$ の *3* 次と *4* 次のキュミュラントから成る定数，そして $H_{a_1\ldots a_l}(\boldsymbol{u})$ は

$$H_{a_1\ldots a_l}(\boldsymbol{u})=(-1)^l\frac{\partial^l}{\partial a_1\ldots\partial a_l}\phi_p(\boldsymbol{u}),$$

で定義される Hermite 多項式であるとする．

2.3. 統計量 T_c の帰無分布の漸近展開

$\boldsymbol{u}$ の密度関数の漸近展開式を用いると，統計量 T_c の特性関数は

$$\begin{aligned}C_{T_c}(t)\ &=\mathrm{E}\left(e^{itT_c}\right) &(2.14)\\ &=\int\exp\left(it\boldsymbol{u}'\boldsymbol{u}\right)\exp\left(-\frac{1}{2}\boldsymbol{u}'\boldsymbol{u}\right)\left\{1+\frac{1}{\sqrt{n}}R_1(\boldsymbol{u})+\frac{1}{n}R_2(\boldsymbol{u})\right\}d\boldsymbol{u},\end{aligned}$$

と表される．ここで，$\varphi=(1-2it)^{-1/2}$，$\boldsymbol{x}=\varphi^{-1}\boldsymbol{u}$ とおくと，上の特性関数は

$$\begin{aligned}C_{T_c}(t)\ &=\int\varphi^p\exp\left(it\boldsymbol{x}'\boldsymbol{x}\right)\left\{1+\frac{1}{\sqrt{n}}R_1(\varphi\boldsymbol{x})+\frac{1}{n}R_2(\varphi\boldsymbol{x})\right\}d\boldsymbol{x}\\ &=\varphi^p\mathrm{E}\boldsymbol{x}\left[1+\frac{1}{\sqrt{n}}R_1(\varphi\boldsymbol{x})+\frac{1}{n}R_2(\varphi\boldsymbol{x})\right], &(2.15)\end{aligned}$$

と表すことができるので，$\boldsymbol{x}\sim N_p(\boldsymbol{0},I_p)$ とみなして期待値計算をすることにより，T_c の特性関数を計算することができる．その結果は，

$$C_{T_c}(t)\ =\varphi^p\left[1+\frac{1}{n}\left\{(\varphi^2-1)a_1+(\varphi^2-1)^2a_2+(\varphi^2-1)^3a_3,\right\}\right], \quad (2.16)$$

となる．但し，

$$a_1 = \sum \delta_{a'b'} q_2, \quad a_2 = \sum_{[3]} \delta_{a'b'} \delta_{c'd'} q_4, \quad a_3 = \sum_{[15]} \delta_{a'b'} \delta_{c'd'} \delta_{e'f'} q_6, \tag{2.17}$$

であるとする．この特性関数を反転することにより，次の結果を得る．

定理 2.2
条件 *A1*，*B1*，*B2* のもとで，T_c の分布関数は漸近的に

$$\mathrm{P}(T_c \geq x) = G_p(x) + \frac{1}{n} \sum_{k=1}^{3} b_k G_{p+2j}(x) + o(n^{-1}), \tag{2.18}$$

と表すことができる．但し，$G_m(x)$ は自由度 m の χ^2-分布の分布関数であり，

$$\begin{aligned} b_0 &= -a_1 + a_2 - a_3, \quad b_1 = a_1 - 2a_2 + 3a_3, \\ b_2 &= a_2 - 3a_3, \quad b_3 = a_3, \end{aligned} \tag{2.19}$$

であるとする．

A. 付録・漸近展開について

p 次元のサンプル $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ が存在するとき，推定や検定などの統計的推測はこれらのサンプルより計算される統計量 $T_n = T(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ に基づいて行われる．この時，推定における推定誤差の評価や検定における有意点の計算等，推定方式を評価する為には T_n の（標本）分布を知る必要がある．しかしながら，統計量 T_n 自体が複雑な場合には，T_n の分布を正確に求めることが困難であることも少なくない．ところが，このような場合でも，サンプル数 n を大きくしていくと，統計量の分布が正規分布や χ^2-分布などといった標準的な分布に近づいていく場合がしばしばある．この場合，サンプル数 n が十分大きいときには，統計量 T_n の分布を $n \to \infty$ とした極限の分布で近似することができる．これが大標本論の基礎であるが，大標本論で与えられる第 1 次近似，すなわち極限の分布のみでは近似が満足のいかないものであることも多い．そこで，近似を良くする為に更に高次元の近似を求めようとするのが漸近展開の考え方である．

漸近展開において注意すべきことは，漸近展開は理論的には標本数 n が十分大きなときに収束を改善する手法であり，n が小さいところでは漸近展開によって近似が改良されるとは限らないという事である．また，モーメントやキュミュラントの計算など，漸近展開式を導出する計算が面倒であることや，導出した漸近展開式が理論的に複雑で使いづらいものになる可能性があること，そして，近似の妥当性について条件等を考える必要があるといった難点も存在する．しかしながら経験的に，漸近展開はさまざまな状況下で有効であることが分かっている．

最近では，近似を考える際に，Bootstrap 法などといった計算機を用いてノンパラメトリック法の枠組みの中で近似を計算するといった方法もとられているが，漸近展開の利点には以下のことが挙げられる．まず，漸近展開は式の形で表されているので，一度計算しておくと様々に応用が利く点が挙げられる．又，式を見て傾向等が把握できる点も魅力的であろう．

漸近展開の導出方法は様々であるが，ここでは，本編にて用いられている Edgeworth 展開について，1 次元の場合を例に述べる．

$Y, Y_1, \cdots, Y_n$ を平均が μ, 分散が σ^2 である独立同一分布に従う確率変数で, r 次キュミュラント κ_r を持っているものと仮定する．又，標準化されたキュミュラントを $\rho_r := \kappa_r/\sigma^r$ と表す事にする．この確率変数の和 $Y_1+\cdots+Y_n$ を S_n と定義し，その標準化を

$$S_n^* := \frac{S_n - n\mu}{\sigma\sqrt{n}}, \tag{A.1}$$

と表すと，S_n^* のキュミュラント母関数 $K(S_n^*;t)$ は,

$$\begin{aligned} K(S_n^*;t) &= -\sqrt{n}\mu t/\sigma + nK(Y;t/\sigma\sqrt{n}) \\ &= \frac{1}{2}t^2 + \frac{\rho_3 t^3}{6\sqrt{n}} + \frac{\rho_4 t^4}{24n} + O(n^{-3/2}), \end{aligned} \tag{A.2}$$

と表される．S_n^* のモーメント母関数は $M(S_n^*;t) = \exp\{K(S_n^*;t)\}$ となるので,

$$M(S_n^*;t) = \exp(\frac{1}{2}t^2)\{1 + \frac{\rho_3 t^3}{6\sqrt{n}} + \frac{\rho_4 t^4}{24n} + \frac{\rho_3^2 t^6}{72n} + O(n^{-3/2})\}, \tag{A.3}$$

と表される．ここで，Y の特性関数 $\varphi(t)$ が Cramér 条件

$$\limsup_{|t|\to\infty} |\varphi(t)| < 1,$$

を満たすと仮定する．この仮定により，(A.3) のモーメント母関数は validity を損なわずに反転を行う事ができる．よって，S_n^* の確率密度関数について以下の展開を得る．

$$f(S_n^*;x) = \phi(x)\{1 + \frac{\rho_3 h_3(x)}{6\sqrt{n}} + \frac{\rho_4 h_4(x)}{24n} + \frac{\rho_3^2 h_6(x)}{72n}\} + O(n^{-3/2}), \tag{A.4}$$

但し，$\phi(x)$ は標準正規分布の確率密度関数であり，$h_r(x)$ は r 次の Hermite 多項式であるとする．このような展開を独立和の Edgeworth 展開という．なお，同じ仮定の下での分布関数についての展開は

$$F(S_n^*;x) = \Phi(x) - \phi(x)\{\frac{\rho_3 h_2(x)}{6\sqrt{n}} + \frac{\rho_4 h_3(x)}{24n} + \frac{\rho_3^2 h_5(x)}{72n}\} + O(n^{-3/2}), \tag{A.5}$$

と表される．

また，統計量の分布の漸近展開を利用して，その統計量のパーセント点を求めることを考える．例えば，統計量 T_n の分布関数 $F_n(x)$ が $n \to \infty$ で標準正規分布 $N(0,1)$ に従うとする．すなわち，全ての x に対して

$$\lim_{n\to\infty} F_n(x) = \Phi(x), \tag{A.6}$$

であるとする．このとき，分布関数 F，Φ の上側 α 点をそれぞれ x，u とすると

$$F_n(x) = \Phi(u) = 1 - \alpha, \tag{A.7}$$

となる．この方程式を満たす解 $x = x(u)$ が求まれば，T_n の上側 α 点が Ψ の上側 α 点 u を用いて求められたことになる．

統計量 T_n の特性関数 $C_n(t)$ が

$$C_n(t) = e^{(it)^2/2}\left[1 + \frac{1}{\sqrt{n}}\{a_1(it) + a_3(it)^3\} + \frac{1}{n}\{b_2(it)^2 + b_4(it)^4 + b_6(it)^6\} + o(n^{-1})\right], \tag{A.8}$$

と展開されるとする．このとき，適当な正則条件の下で

$$\begin{aligned} \mathrm{P}(T_n \le x) &= F_n(x) \\ &= \Phi(x) + \phi(x)\left\{\frac{1}{\sqrt{n}}p_1(x) + \frac{1}{n}p_2(x)\right\} + o(n^{-1}), \end{aligned} \tag{A.9}$$

と表すことができる．但し，

$$p_1(x) = -\{a_1 + a_3 h_2(x)\},$$
$$p_2(x) = -\{b_2 h_1(x) + b_4 h_3(x) + b_6 h_5(x)\},$$

であり，h_r は r 次の Hermite 多項式であるとする．この展開式に

$$x = x(u) = u + \frac{1}{\sqrt{n}} q_1(u) + \frac{1}{n} q_2(u) + o(n^{-1}),$$

を代入して，$F_n(x(u)) = \Phi(u)$ となる $q_1(u)$，$q_2(u)$ を決定することができ，

$$q_1(u) = -p_1(u), \tag{A.10}$$
$$q_2(u) = -p_2(u) - \frac{1}{2} u p_1(u) + p_1'(u) p_1(u),$$

となる．

参考文献

[1] Barndorff-Nielsen, O. E. and Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics*, Chapman and Hall.

[2] Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansions, *J.Multivariate Anal.* **27**, 68-79.

[3] Nel, D. G., Van der Merwe, C. A. and Moser, B. K. (1990). The exact distribution of the univariate and multivariate Behrens-Fisher statistics with a comparison of several solutions in the univariate case. *Comm. Statist. Theory Methods*, **19**, 279–298.

[4] Wakaki, H., Yanagihara, H. and Fujikoshi, Y. (2002). Asymptotic expansion of the null distributions of test statistics for multivariate linear hypothesis under nonnormality. *Hiroshima Math. J.*, **32**, 17–50.

[5] Yanagihara, H. and Yuan, K.-H. (2005). Three approximate solutions to the multivariate Behrens-Fisher problem. *Comm. Statist. Simulation Comput.*, **34** (in press).