

多変量解析のフロンティア

杉山 高一 (中央大学名誉教授 創価大学客員教授)*

1. 固有ベクトルの分布

1.1. 固有ベクトルの分布

主成分分析は、要因分析の一つとして、最もよく使われている分析法である。主成分分析において、無作為標本 $\mathbf{X}_1, \dots, \mathbf{X}_N$ から計算した第1主成分 y_1 を

$$y_1 = h_{11}x_1 + \dots + h_{p1}x_p$$

と書く。ここで、 $\mathbf{X} = (x_1, \dots, x_p)$ 。この係数 (h_{11}, \dots, h_{p1}) は、 x_1, \dots, x_p のあらゆる1次式の中で y_1 の分散が最大になるように決める。これは、 y_1 の分散である固有値 l_1 に対応する固有ベクトルであり、 N 個の標本の組を抽出する度に異なった値をとる。次に、「第1主成分 y_1 とは無相関」という条件の下で分散が最大になるような1次式

$$y_2 = h_{12}x_1 + \dots + h_{p2}x_p$$

で第2主成分 y_2 を決め、以下同様にして、第3主成分 y_3 、第4主成分 y_4 、 \dots を決める。 \mathbf{X} が中学あるいは高校の成績データであれば、第1主成分は合計点に近いものになる。身体計測のデータであれば、第1主成分は、大きさの因子、第2主成分に体型の因子などが出てくる。主成分がどのような意味を持つかは、 x_i の係数 h_i ($i = 1, \dots, p$) の大きさを見て判断する。このように、主成分分析においては、主成分の係数 h_i の大きさを見て、その主成分がどのような意味を持つかを推測するが、もし、この係数が標本抽出の度に大きく変わるならば、推測した主成分の意味づけは無意味になる。このような観点から、主成分の係数 h_1, h_2, \dots, h_p の信頼性は重要な問題である。

まず始めに、2変量正規分布の場合について述べる。 $\mathbf{X}_1, \dots, \mathbf{X}_N$ を2変量正規分布 $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ からの無作為標本とする。このとき

$$\mathbf{S} = \sum_{\alpha=1}^N (\mathbf{X}_\alpha - \bar{\mathbf{X}})(\mathbf{X}_\alpha - \bar{\mathbf{X}})'$$

は、確率密度関数

$$g(s_{ij}) = \begin{cases} \text{Const. } |\mathbf{S}|^{\frac{n-3}{2}} \exp\left[-\frac{1}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{S}\right] & \mathbf{S} \text{ は正定値} \\ 0 & \text{その他} \end{cases}$$

を持つ。ここで、 $n = N - 1$ 、 $\boldsymbol{\Sigma} = \|\sigma_{ij}\|$ は正定値、 $\boldsymbol{\Sigma}^{-1} = \|\sigma^{ij}\|$ 、 $\mathbf{S} = \|s_{ij}\|$,

$$\text{Const} = \frac{|\boldsymbol{\Sigma}^{-1}|^{\frac{n}{2}}}{2^n \pi^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n-1}{2}\right)}.$$

* 〒192-8577 東京都八王子市丹木町 1-236 創価大学
e-mail: tsugi@soka.ac.jp

この $g(s_{ij})$ は、2次元のウィッシュャート分布の確率密度関数である。標本から計算した第1主成分 y_1 と第2主成分 y_2 をそれぞれ

$$\begin{aligned} y_1 &= h_{11}x_1 + h_{12}x_2 \\ y_2 &= h_{21}x_1 + h_{22}x_2 \end{aligned}$$

と書くと、固有ベクトル、つまり、係数 (h_{11}, h_{12}) , (h_{21}, h_{22}) は

$$\begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

と表される。これが意味することは、第1主成分と第2主成分は座標軸が回転移動している、つまり、真の主成分の位置からズレているということである。この座標軸の回転の角度 θ は確率変数であり、その分布は母集団での角度 γ と、母集団での固有値、つまり第1主成分の母分散 λ_1 と第2主成分の母分散 λ_2 に依存して定まる。

母集団での固有値を $\lambda_1, \lambda_2 (\lambda_1 \geq \lambda_2 > 0)$ とすると、確率変数 θ の確率密度関数は、Sugiyama(1965) によって

$$\frac{1}{\pi(n+1)} \left(\frac{\lambda_1 \lambda_2}{\bar{\lambda}^2} \right)^{\frac{n}{2}} {}_2F_1 \left(2, n; \frac{n+3}{2}; x \right)$$

と与えられている。ただし

$$x = (\sigma_{11} \cos^2 \theta + 2\sigma_{12} \sin \theta \cos \theta + \sigma_{22} \sin^2 \theta) / (\sigma_{11} + \sigma_{22})$$

$$F(\alpha, \beta, \gamma; z) = \sum_{i=0}^{\infty} \left[\frac{\Gamma(\alpha+i)}{\Gamma(\alpha)} \right] \left[\frac{\Gamma(\beta+i)}{\Gamma(\beta)} \right] \left[\frac{\Gamma(\gamma)}{\Gamma(\gamma+i)} \right] \left[\frac{z^i}{i!} \right]$$

$$\bar{\lambda} = (\lambda_1 + \lambda_2) / 2.$$

簡単のため、母集団の座標軸の回転の角度 γ を0とする。つまり、母集団での第1主成分 Y_1 と第2主成分 Y_2 は、それぞれ

$$\begin{aligned} Y_1 &= x_1 + 0 \cdot x_2 \\ Y_2 &= 0 \cdot x_1 + x_2 \end{aligned}$$

であるとする。母集団における係数が0のとき、 N 個の標本から計算した主成分はの係数 $h_{12} = \sin \theta$ はどのような値をとるかを調べる。固有ベクトルの成分から、 $h_{12} = -h_{21}$ 故、 h_{12} の変動の様子を見ることは、 h_{21} の変動の様子を見ることと同じである。図で表すと以下のようなになる。

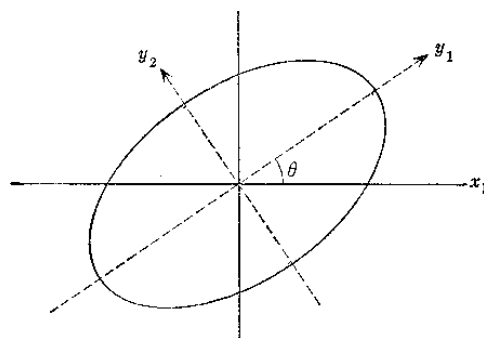


図 変数 x_1, x_2 と主成分 y_1, y_2 との関係

角度 θ の分布は Y_1 の分散 λ_1 と Y_2 の分散 λ_2 に依存すると述べたが、正確には λ_1 と λ_2 の比に依存する。そこで、 $\lambda_1 + \lambda_2$ を100として、係数 $h_{12} = \sin \theta$ の両側5%点を計算したのが表1.1である。

表1.1 第1主成分の x_2 の母係数が0のとき、標本から計算した係数
 $h_{12} = \sin \theta$ の両側5%点(±符号を省略), $n = N - 1$

n	$\lambda_1 + \lambda_2 = 100$ とした場合の λ_1 の値(()内はその時の θ の値)			
	90	80	70	60
10	.275($\theta=16.0$)	.528($\theta=31.9$)	.864($\theta=59.8$)	.982($\theta=79.1$)
20	.178($\theta=10.3$)	.329($\theta=19.2$)	.619($\theta=38.2$)	.958($\theta=73.3$)
50	.107($\theta=6.1$)	.193($\theta=11.1$)	.346($\theta=20.2$)	.807($\theta=53.8$)
100	.075($\theta=4.3$)	.133($\theta=7.6$)	.234($\theta=13.5$)	.559($\theta=34.0$)

1.2. 固有ベクトルの分布と検定

正規母集団 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ からの大きさ N の無作為標本 $\mathbf{X}_1, \dots, \mathbf{X}_N$ から計算された標本共分散 \mathbf{S} の固有ベクトルについて考える。いま、母共分散行列は

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma} \boldsymbol{\Lambda} \boldsymbol{\Gamma}', \quad \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p) \quad (\lambda_1 > \dots > \lambda_p), \quad \boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$$

と分解でき、固有値 λ_α に対応する固有ベクトルを $\boldsymbol{\gamma}_\alpha$ とする。また、標本共分散行列も同様に

$$\mathbf{S} = \mathbf{H} \mathbf{L} \mathbf{H}', \quad \mathbf{L} = \text{diag}(l_1, \dots, l_p) \quad (l_1 > \dots > l_p), \quad \mathbf{H} = (\mathbf{h}_1, \dots, \mathbf{h}_p)$$

と分解でき、固有値 l_α に対応する固有ベクトルを \mathbf{h}_α とする。

Sugiyama (1965) による $p = 2$ の場合の最大固有値に対応する固有ベクトルの精密分布については、前節で詳しく述べた。一般的な p に対する精密分布はSugiyama(1967)で得られており、

$$\begin{aligned} f(\mathbf{h}_1) d\mathbf{h}_1 &= \left[|\boldsymbol{\Sigma}|^{n/2} \Gamma\left(\frac{n}{2}\right) \right]^{-1} \Gamma_p\left(\frac{pn}{2}\right) B_{p-1}\left(\frac{p+2}{2}, \frac{n-1}{2}\right) \\ &\quad \times \sum_{k=0}^{\infty} \frac{\binom{pn}{2}_k}{k!} \sum_{\kappa} \frac{\binom{n-1}{2}_\kappa}{\binom{n+p+1}{2}_\kappa} C_\kappa\left(-\frac{\boldsymbol{\Sigma}_{p-1}}{\mathbf{h}_1' \boldsymbol{\Sigma}^{-1} \mathbf{h}_1}\right)^{pn/2} d\mathbf{h}_1 \end{aligned}$$

と表される。ただし、 $n = N - 1$ である。また、分布の漸近展開はSugiura(1976), Kollo and von Rosen (2005) で与えられている。

次に固有ベクトルに関する検定問題

$$H_0 : \boldsymbol{\gamma}_\alpha = \boldsymbol{\gamma}_0, \quad H_1 : \boldsymbol{\gamma}_\alpha \neq \boldsymbol{\gamma}_0 \quad (\boldsymbol{\gamma}_0 : \text{既知ベクトル})$$

を考える。Anderson による検定問題の検定統計量の1つはAnderson(1963)による

$$\Lambda_1 = n \left(l_\alpha \boldsymbol{\gamma}_0' \mathbf{S}^{-1} \boldsymbol{\gamma}_0 + \frac{1}{l_\alpha} \boldsymbol{\gamma}_0' \mathbf{S} \boldsymbol{\gamma}_0 - 2 \right)$$

がある。この統計量の分布の漸近展開はHayakawa(1978)が求めている。この検定統計量 Λ_1 に匹敵する他の検定統計量として

$$\begin{aligned} \Lambda_2 &= n \left(\boldsymbol{\gamma}_0' \mathbf{S}^2 \boldsymbol{\gamma}_0 - 2 l_\alpha \boldsymbol{\gamma}_0' \mathbf{S} \boldsymbol{\gamma}_0 + l_\alpha^2 \right) \\ \Lambda_3 &= n \left(\frac{1}{l_\alpha} \boldsymbol{\gamma}_0' \mathbf{S}^2 \boldsymbol{\gamma}_0 - 2 \boldsymbol{\gamma}_0' \mathbf{S} \boldsymbol{\gamma}_0 + l_\alpha \right) \end{aligned}$$

が塚田・杉山(1997), Tsukada(1998)で提案され, 帰無分布や検出力の漸近展開が得られ, この統計量を用いた Bootstrap 検定も与えられている. 検定統計量 Λ_2, Λ_3 の上側 5%点については 2次元及び 3次元の場合に固有値を変えた 3種類が得られている (Tsukada(1999)).

有意水準 5%, $n = 100, p = 3$, 母共分散行列を $\Sigma = \text{diag}(1.8, 0.9, 0.3)$ とし, 帰無仮説 $H_0 : \gamma_1 = (1, 0, 0)'$, 対立仮説 $H_1 : \gamma_1 = (\cos \theta_1, \sin \theta_1 \cos \theta_2, \sin \theta_1 \sin \theta_2)'$ としたとき, それぞれの検出力は下表のようである.

θ_2	θ_1	3°	6°	9°	12°	15°	18°	21°	24°	27°
1.5°	Λ_1	.081	.116	.177	.263	.368	.482	.592	.690	.771
	Λ_2	.070	.119	.199	.304	.425	.546	.656	.747	.818
	Λ_3	.070	.119	.202	.312	.440	.567	.682	.776	.846
3°	Λ_1	.148	.187	.250	.336	.436	.542	.642	.729	.800
	Λ_2	.090	.142	.225	.332	.452	.569	.675	.762	.828
	Λ_3	.090	.142	.228	.341	.468	.592	.703	.791	.857
4.5°	Λ_1	.272	.311	.372	.451	.540	.631	.714	.784	.841
	Λ_2	.133	.188	.273	.381	.497	.608	.706	.785	.845
	Λ_3	.133	.190	.280	.393	.516	.634	.736	.815	.874

第1固有ベクトルに関しては, θ_1 が θ_2 より大きい範囲で統計量 Λ_3 を用いた検定の検出力が大きくなり, 逆に θ_2 が θ_1 より大きい範囲で統計量 Λ_1 を用いた検定の検出力が大きい. また, 第1主成分の寄与率が大きいほど θ_1 と θ_2 ($\theta_1 > \theta_2$) の差が小さい場合でも Λ_3 が最良となっている.

また, 第2固有ベクトルに関する検出力は 帰無仮説 $H_0 : \gamma_2 = (0, 1, 0)'$, 対立仮説 $H_1 : \gamma_2 = (-\sin \theta_1, \cos \theta_1 \cos \theta_2, \cos \theta_1 \sin \theta_2)'$ として, 下表のようである.

θ_2	θ_1	4°	8°	12°	16°	20°	24°	28°	32°	36°
2°	Λ_1	.081	.143	.252	.395	.547	.682	.788	.861	.908
	Λ_2	.079	.166	.304	.468	.626	.753	.842	.899	.935
	Λ_3	.078	.163	.296	.455	.607	.732	.822	.882	.921
4°	Λ_1	.119	.184	.295	.436	.583	.712	.810	.876	.919
	Λ_2	.082	.170	.309	.475	.631	.757	.845	.902	.937
	Λ_3	.080	.166	.300	.460	.612	.736	.826	.885	.923
6°	Λ_1	.187	.256	.367	.503	.640	.756	.842	.898	.935
	Λ_2	.087	.178	.319	.486	.642	.766	.852	.906	.941
	Λ_3	.085	.173	.309	.469	.621	.744	.832	.890	.927

第2固有ベクトルに関しても, 第1固有ベクトルと同じ傾向があるが Λ_3 よりも Λ_2 の方が大きくなる. 3次元の場合, θ_2 が大きくなる対立仮説に対して Λ_1 がよく, θ_1 が大きくなる対立仮説に対して Λ_2 や Λ_3 が良いことが分かる. このような傾向は 5次元の場合でも見られる.

複数母集団における固有ベクトルで張られる部分空間の比較は Kzanowski(1979) で提案され, 部分空間の同等性検定として Flury(1988) は Common principal component analysis(CPCA) を提案した. CPCA では各群の母共分散行列 Σ_i に対して

$$H_{CPC} : \Sigma_g = \Gamma \Lambda_g \Gamma', (g = 1, \dots, k)$$

を検定する. この検定に対して尤度比検定統計量が提案されたが, 計算機による繰り返し計算が必要であるため極限分布のみ得られている. また, 固有ベクトルの一部だけが

等しい Partial Common principal component

$$H_{PCPC} : \Sigma_g = (\gamma_1, \dots, \gamma_k, \gamma_{k+1}^{(g)}, \dots, \gamma_p^{(g)}) \Lambda_i (\gamma_1, \dots, \gamma_k, \gamma_{k+1}^{(g)}, \dots, \gamma_p^{(g)})',$$

$$(g = 1, \dots, k)$$

も提案されている。また, Schott(1988) では2母集団における固有値を用いた検定統計量が提案されている。

母集団分布が4次モーメントまでが存在する一般的な母集団分布の場合, Kollo and Neudecker (1993) は共分散行列に対する固有ベクトルの極限分布を次のように得ている。

$$\sqrt{n}(\mathbf{h}_\alpha - \gamma_\alpha) \xrightarrow{D} N(\mathbf{0}, T_\alpha)$$

ただし, $T_\alpha = \{\gamma_\alpha \otimes \Gamma(\lambda_\alpha I - \Lambda)^+ \Gamma'\} M_4(\mathbf{X}) \{\gamma_\alpha \otimes \Gamma(\lambda_\alpha I - \Lambda)^+ \Gamma'\}$ であり, $M_4(\mathbf{X})$ は平均まわりの4次モーメントで, A^+ は行列 A のムーア・ペンローズ逆行列を表している。

母集団分布が正規分布でない場合, 前述の検定問題

$$H_0 : \gamma_\alpha = \gamma_0, \quad H_1 : \gamma_\alpha \neq \gamma_0$$

に対して, Anderson(1963) による統計量は適切でないことをシミュレーションにより示し, 4つのノンパラメトリック検定法を牛澤(1998) が提案している。

ここでは主成分分析に関する固有ベクトルについて詳しく述べたが, 固有ベクトルに関する問題は, 多変量解析の他の分析法でも存在し, 多くがこれからの研究に期待されている。

2. 固有値の分布

2.1. 固有値の分布と検定

ゾーナル多項式やその級数として定義される超幾何関数は, 多変量正規母集団下で共分散行列の分布, 特に固有値分布を記述するために重要である。ゾーナル多項式は, James (1960) によって, 一般線型群と直交群に関する表現論を利用して導出された。James (1960) 以降, 共分散行列の固有値に関する多くの統計量が, ゾーナル多項式を使って求められている (James (1964), Constantine (1963), 早川 (1971, 1972)), Sugiyama (1967) は, 最大固有値分布をゾーナル多項式の無限級数として求め, Khatri (1972) は, 分布がゾーナル多項式の有限級数になる場合の統計量の考察を行った。しかしゾーナル多項式の数値計算の困難性や, 収束の遅さ故から, 漸近論の重要性が強調されるようになった (早川 [?])。固有値分布の漸近展開は Sugiura (1972), Fujikoshi (1977) で行われている。

1960, 1970年代での高次ゾーナル多項式の計算実現は, Sugiyama (1979) の2変量で200次の場合に限られる。この Sugiyama (1979) の計算方法を利用して, Sugiura (1990) では, 2変量の場合に最大固有値, 最小固有値の分布を数値計算し, 基本統計量など分布の様相を調べている。3変量以上のゾーナル多項式の数値計算は, Hashiguchi and Niki (2006), Hashiguchi, Nakagawa and Niki (2000) で実装され, これを利用した最大, 最小固有値の分布の数値計算が Hashiguchi and Niki (2006) で行われている。

p 変量正規分布 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ から N 個の標本 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ を得たとき、標本分散共分散行列 \mathbf{S} の固有値 l_1, \dots, l_p の同時分布は James (1960, 1964) によってゾーナル多項式 $C_\kappa(\mathbf{Y})$ を用いて導出された。

$$\frac{\pi^{\frac{1}{2}p^2}}{2^{\frac{1}{2}pn} \Gamma_p(\frac{1}{2}n) \Gamma_p(\frac{1}{2}p)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}n} {}_0F_0(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Lambda}) |\boldsymbol{\Lambda}|^{\frac{1}{2}(n-p-1)} \prod_{i < j} (l_i - l_j) \prod_i dl_i$$

ただし $\boldsymbol{\Lambda}$ と超幾何関数 ${}_0F_0(\mathbf{S}, \mathbf{T})$ は次のように与えられる。

$$\boldsymbol{\Lambda} = \text{diag}(l_1, \dots, l_p), \quad {}_0F_0(\mathbf{S}, \mathbf{T}) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{C_\kappa(\mathbf{S}) C_\kappa(\mathbf{T})}{C_\kappa(\mathbf{I}_p) k!}$$

最大固有値の分布については Sugiyama (1967) によって次のように導出された

$$F(l_1 < nx) = \text{Const.} \cdot \exp\left(-\frac{nx}{2} \text{tr} \boldsymbol{\Sigma}^{-1}\right) (nx)^{\frac{pn}{2}} \cdot {}_1F_1\left(\frac{p+1}{2}; \frac{n+p+1}{2}; \frac{nx}{2} \boldsymbol{\Sigma}^{-1}\right),$$

ただし定数 Const. と超幾何関数 ${}_1F_1(a; c; \mathbf{Y})$ は次のように与えられる。

$$\text{Const.} = |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \Gamma_p\left(\frac{p+1}{2}\right) / 2^{\frac{np}{2}} \Gamma_p\left(\frac{n+p+1}{2}\right), \quad {}_1F_1(a; c; \mathbf{Y}) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a)_\kappa C_\kappa(\mathbf{Y})}{(c)_\kappa k!}.$$

この分布を用いて、母集団分布が 2 次元正規分布の場合の固有値に関する精密分布の計算と検出力が、Sugiyama (1972) によって得られている。固有値の検定に際してパーセント点の値を求めるためには固有値の分布の計算は必要不可欠であるが、ゾーナル多項式 $C_\kappa(\mathbf{Y})$ の計算は特別な場合 ($\boldsymbol{\Sigma} = \mathbf{I}$) を除き計算が困難であるため、一般の場合の固有値の分布を計算することも困難であった。Sugiyama (1979) によって 2 次元の場合のゾーナル多項式の係数計算の結果が求められたが、その 3 次元以上の計算は不可能であった。その後 Sugiyama, Fukuda and Takeda (1999) によって超幾何関数 ${}_2F_1(a, b; c; \mathbf{Y})$, ${}_1F_1(a; c; \mathbf{Y})$, ${}_0F_1(c; \mathbf{Y})$ を求める漸化式が 7 次元まで導出された。例えば、超幾何関数 $F = {}_1F_1(a, b; c; \mathbf{Y})$ の満たす偏微分方程式は、Muirhead (1982) にあるように $i = 1, \dots, p$ に対して

$$y_i \frac{\partial^2 F}{\partial y_i^2} + \left\{ c - \frac{p-1}{2} - y_i + \frac{1}{2} \sum_{j=1, j \neq i}^p \frac{y_i}{y_i - y_j} \right\} \frac{\partial F}{\partial y_i} - \frac{1}{2} \sum_{j=1, j \neq i}^m \frac{y_j}{y_i - y_j} \frac{\partial F}{\partial y_j} = aF$$

であり、基本対称式 r_1, \dots, r_p で変換した場合、 $i = 1, \dots, p$ に対して

$$\sum_{k,l=1}^p a_{k,l}^{(i)} \frac{\partial^2 F}{\partial r_k \partial r_l} + \left[c - \frac{1}{2} \right] \frac{\partial F}{\partial r_i} + \frac{\partial F}{\partial r_{i-1}} - \delta_{1i} \left(\sum_{k=1}^p r_k \frac{\partial F}{\partial r_k} + aF \right) = 0$$

となる。これをもとに、超幾何関数 ${}_1F_1(a; c; \mathbf{Y})$ を基本対称式で展開したときの係数 $\gamma(j)$ の漸化式が得られる。3 次元までの結果は次の通りである。

$$\begin{aligned} \gamma(j_1, 0, 0) &= \{(j_1 + a - 1) \gamma(j_1 - 1, 0, 0, 0)\} \times \{j_1(c + j_1 - 1)\}^{-1} \\ \gamma(j_1, j_2, 0) &= \{(j_1 + 1)(j_1 + 2) \gamma(j_1 + 2, j_2 - 1, 0, 0) - (j_1 + 1) \gamma(j_1 + 1, j_2 - 1, 0, 0)\} \\ &\quad \times \{j_2(j_2 + c - \frac{3}{2})\}^{-1} \\ \gamma(j_1, j_2, j_3) &= \{2(j_1 + 1)(j_2 + 1) \gamma(j_1 + 1, j_2 + 1, j_3 - 1, 0) \\ &\quad + (j_2 + 1)(j_2 + 2) \gamma(j_1 - 1, j_2 + 2, j_3 - 1, 0) \\ &\quad - (j_2 + 1) \gamma(j_1, j_2 + 1, j_3 - 1, 0)\} \times \{j_3(j_3 + c - 2)\}^{-1} \end{aligned}$$

この漸化式を用いて3次元以上の場合の固有値の分布の計算が可能になったが、コンピュータのメモリの制約や引き算による丸め誤差の影響で、精密に計算できる範囲は3次元の場合でも標本数 N が小さい部分に限られた。実際に $\Sigma = I$ と $\Sigma = \text{diag}(2.0, 1.2, 0.8)$ の場合の最大固有値の上側5%点の値を計算した結果は次の通りである。ここでは値の正確性を確認するため、 $\Sigma = I$ の場合 Sugiyama (1972) の結果、 $\Sigma = \text{diag}(2.0, 1.2, 0.8)$ の場合シミュレーション結果を載せている。

Table 2.1 $\Sigma = I$ のとき

n	漸化式の結果	Sugiyama の結果
2	5.370173	5.370173
4	3.810174	3.810174
6	3.181457	3.181457
8	2.828000	2.828000
10	2.596608	2.596608
12	2.431132	2.431132
14	2.305742	2.305741
16	2.206759	2.206759
18	2.126207	2.126207
20	2.059093	2.059093
22	2.002116	2.002116

Table 2.2 $\Sigma = \text{diag}(2.0, 1.2, 0.8)$ のとき

n	漸化式の結果	シミュレーションの結果
2	7.646561	7.65
4	5.602895	5.61
6	4.779797	4.78
8	4.318502	4.32
10	4.017546	4.02
12	3.803020	3.81
14	3.640940	3.64
16	3.514184	3.52

$$n = N - 1$$

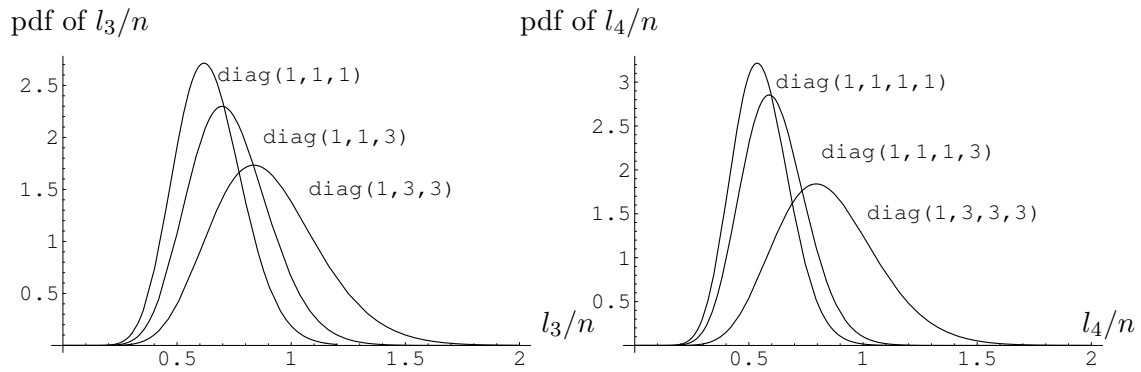
また Sugiyama, Takeda and Fukuda (1998) による漸化式で用いての数値計算に対して、ゾーナル多項式そのものを求めて数値計算することが考えられよう。ゾーナル多項式を使うことの利点として、前述の漸化式を用いた超幾何関数の計算では a, b, c の値が変化すると漸化式の再計算を行う必要があったが、ゾーナル多項式の係数は一度だけ計算すればよい。さらにゾーナル多項式を monomial symmetric function を用いて表わしているため正項級数となり、引き算による丸め誤差の影響を受けないなどの利点もある。

Khatri (1972) では、ゾーナル多項式の有限級数となる分布について考察が行われている。もし $j = (n - p - 1)/2$ が非負整数なら、最小固有値 l_p の上側確率は、

$$\Pr[l_p > x] = \exp\left(-\frac{x}{2}\text{tr}\Sigma^{-1}\right) \sum_{k=0}^{jp} \frac{x^k}{2^k k!} \sum_{\kappa}^* C_{\kappa}(\Sigma^{-1}) \quad (1)$$

で与えられる。ただし、和の記号 \sum_{κ}^* は、 κ の第一要素 κ_1 が $\kappa_1 \leq j$ を満たす分割にわたって和を取ることを意味する。

Hashiguchi et. al. (2000) は、ゾーナル多項式 $C_{\kappa}(Y)$ を基本対称式で展開するアルゴリズムを提案し、Hashiguchi and Niki (2006) で数値計算を行った。下図に $p = 3, 4$ での密度のグラフを示す。 $p = 2$ の場合は、杉山・牛澤 (1980) で数値計算結果が与えられている。



(a) 次元 $m = 3$, 自由度 $n = 30$ のとき (b) 次元 $m = 4$, 自由度 $n = 31$ のとき

固有値の検定については漸近分布を用いてさまざまな検定法が提案されているが、2標本間の j 番目の固有値 λ_j に関する同等性に関する検定法として Takeda(2001) では並べ替え検定を用いた方法が次のように提案している。

2つの母集団から標本 $\mathbf{x}_1^{(1)}, \dots, \mathbf{x}_{N_1}^{(1)}, \mathbf{x}_1^{(2)}, \dots, \mathbf{x}_{N_2}^{(2)}$ が得られたとき、標本分散共分散行列 $S^{(i)}$ をそれぞれ計算する。さらに $S^{(i)}$ の j 番目に大きな固有値 l_j に対する固有ベクトルを $\mathbf{b}_j^{(i) \prime}$ としたとき、下記の変換を考える。

$$z_{jk}^{(i)} = \mathbf{b}_j^{(i) \prime} \left(\mathbf{x}_k^{(i)} - \bar{\mathbf{x}}^{(i)} \right), \quad k = 1, 2, \dots, N_i, \quad i = 1, 2.$$

ここで $z_{jk}^{(i)}$ は主成分スコアであるから、固有値の同等性を主成分スコアの分散の同等性に置き換えて並べ替え検定を行うことが可能となる。

その後 Sakaori (2002), Hino, Murakami and Sugiyama (2009) などの論文において相関行列や級内相関係数に関する同等性検定などに応用された。Wishart 分布の場合の固有値の分布について詳述したが、Non-central Wishart 分布の問題 (Sugiyama(1972) 等) や、多変量解析の他の分析法でも同様の問題があり、研究されている。

3. 正準相関分析

3.1. 正準相関分析における検定統計量の検出力について

確率ベクトル $\mathbf{x}(p \times 1)$ と $\mathbf{y}(q \times 1)$ 間の関連性を調べる方法として、正準相関分析が知られている。正準相関分析に関する仮説検定について、最大根や Lawley-Hotelling 型の統計量の検出力の比較を杉山・牛澤 (1992), 牛澤・杉山 (1994) で行っている。

3.2. 正準相関分析における変数選択基準

正準相関分析における変数選択問題として、有効な変数の組あるいは冗長な変数の組を選ぶ問題がある。正準相関分析における変数選択基準は、藤越 (1985), 市川・小西 (1999), 藤越・倉田 (2008), 小西・北川 (2008) などでたくさんの提案がされている。それらの先行研究の変数選択基準を元に、藤越, 櫻井, 神田, 杉山 (2008) は、ブートストラップを用いて、正準相関分析における変数選択基準を提案した。

正準相関分析におけるブートストラップ変数選択規準による1組の冗長性モデル

$$BDIC = -n \log \left\{ \frac{|S_{(2y)(2y) \cdot 1}|}{|S_{22 \cdot 1}| |S_{yy \cdot 1}|} \right\} + \frac{1}{B} \sum_{j=1}^B B_k^{(j)}$$

正準相関分析におけるブートストラップ変数選択規準による2組の冗長性モデル

$$BDIC = -n \log \left\{ \frac{|\mathbf{S}_{(24)(24) \cdot 13}|}{|\mathbf{S}_{22 \cdot 1}| |\mathbf{S}_{44 \cdot 3}|} \right\} + \frac{1}{B} \sum_{j=1}^B B_k^{(j)}$$

3.3. 主成分正準相関分析

3.3.1. 主成分正準相関分析について

正準相関分析から得られた結果がはっきりしない場合，母集団の情報をどの程度把握しているか分からない場合，意味づけの仕方が難しい場合がある．それらの問題の対応方法の一つとして，Sugiyama and Takeda (1999) は主成分正準相関分析を提案した．主成分正準相関分析とは，各確率ベクトルに対して，それぞれ主成分分析を行い，それぞれの主成分スコアを用いて正準相関分析を行うことである．

主成分正準相関分析は主成分 $\mathbf{u} = \Gamma'_x \mathbf{x}$, $\mathbf{v} = \Gamma'_y \mathbf{y}$ の共分散行列

$$\text{Cov} \left[\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \right] = \Sigma = \begin{pmatrix} \Sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{pmatrix} = \begin{pmatrix} \Gamma'_x \Psi_{xx} \Gamma_x & \Gamma'_x \Psi_{xy} \Gamma_y \\ \Gamma'_y \Psi_{yx} \Gamma_x & \Gamma'_y \Psi_{yy} \Gamma_y \end{pmatrix}$$

を用いて正準相関分析を適用する方法である．

主成分正準相関分析を用いる利点は次のような事柄がある．(1) 正準相関分析結果は解釈が難しいことがあるが，主成分スコアの正準相関分析結果は解釈が容易になる場合がある．(2) 主成分は情報量の大きい順番に並ぶ ($\text{Var}(u_1) > \text{Var}(u_2) > \dots$) ．(3) 主成分は互いに無相関になる ($\text{Cov}(u_i, u_j) = 0$ ($i \neq j$)) ．

3.3.2. 極限分布

主成分正準相関分析の結果の信頼性・安定性を調べるために，推定量の分布の考察を行った．Yamamoto, Sugiyama, Murakami and Sakaori (2007) は \mathbf{x} の第 i 主成分と \mathbf{y} 第 j 主成分による正準相関係数の分布について，Sugiyama, Ogura, Sakaori, Yamada (2007) は， \mathbf{x} の $1 \sim p_1$ 主成分と \mathbf{y} の $1 \sim q_1$ 主成分による正準相関係数について，共分散行列の固有値と固有ベクトルを摂動展開して，その推定量の極限分布を求め，それが正規分布になることを示した．

3.3.3. 変数選択

Ogura (2010) 及び Ogura, Fujikoshi, Sugiyama (2011) は，藤越・倉田 (2008) の正準相関分析における変数選択基準 DIC を主成分正準相関分析に適用した．主成分が情報の大きい順番に並ぶことを利用して，上位数個の主成分のみを用いての主成分正準相関分析の考察を行った．

主成分正準相関分析における1組の変数選択基準は

$$DIC = -n \log \left\{ \frac{|\mathbf{S}_{(2y)(2y) \cdot 1}|}{(|\mathbf{S}_{22}| |\mathbf{S}_{yy \cdot 1}|)} \right\} + 2 \left\{ (p+q)(p+q+1)/2 - (p-p_1)q \right\}.$$

であり，主成分正準相関分析における2組の変数選択基準

$$DIC = -n \log \left\{ \frac{|\mathbf{S}_{(24)(24) \cdot 13}|}{(|\mathbf{S}_{22}| |\mathbf{S}_{44}|)} \right\} + 2 \left\{ \frac{1}{2} (p+q)(p+q+1) - (p-p_1)q - p_1(q-q_1) \right\}.$$

で与えられている。

小椋透(2012)は、この考え方を対応分析に応用して、研究集会「多変量解析の新展開」で、「Variable Selection based on DIC in Correspondence Analysis」という題名で、研究発表を行った。その後、この問題は小椋透、菅民郎、杉山高一、藤越康祝により、研究が進められている。正準相関分析は多変量解析の中でも数学的に難しい方法論で、他に比べると論文数は少ない。最近では、高次元漸近的枠組で、Fujikoshi and Sakurai(2008)等が興味深い結果を導いている。

4. 平均ベクトル間の多変量多重比較法

多変量正規母集団における平均ベクトル間の多重比較法についての同時信頼区間を考える。平均ベクトル間の多変量多重比較法は、通常、同時信頼区間の形を用いて取り扱われることが多く、「同時信頼区間の構成」が重要な問題となる。しかし、一般に正確な棄却限界値を求めることが困難なため、いくつかの保守的な近似同時信頼区間の構成法が提案されている。特に対比較の場合では、ボンフェロニの不等式を利用した修正2次近似法がSiotani(1959), Seo and Siotani(1992), Seo(1995)などで提案されている。また、より簡便な近似法として、多変量Tukey-Kramer法がSeo, Mano and Fujikoshi(1994)で提案されている。

一般に k 個の p 変量正規母集団において、 μ_i を i 番目の母集団からの平均ベクトルとし、 μ_i を並べた $p \times k$ 行列を $M = [\mu_1, \dots, \mu_k]$ とする。さらに、 $\widehat{M} = [\widehat{\mu}_1, \dots, \widehat{\mu}_k]$ を M の推定量とし、 $\text{vec}(\mathbf{X}) \sim N_{kp}(\mathbf{0}, \mathbf{V} \otimes \Sigma)$ とする。ここで、 $\mathbf{X} = [x_1 \dots, x_k] = \widehat{M} - M$ 、 $\mathbf{V} = [v_{ij}]$ は $k \times k$ 既知行列、 Σ は $p \times p$ 未知行列とする。また、 S を Σ の不偏推定量とし、 νS は \widehat{M} と独立で $W_p(\Sigma, \nu)$ に従うものとする。このとき、全ての対比較に対する多変量Tukey-Kramer法による同時信頼区間は、

$$\mathbf{a}'(\mu_i - \mu_j) \in \left[\mathbf{a}'(\widehat{\mu}_i - \widehat{\mu}_j) \pm t_{p,I} \sqrt{d_{ij} \mathbf{a}' \mathbf{S} \mathbf{a}} \right], \quad \forall \mathbf{a} \in \mathbb{R}^p - \{\mathbf{0}\}, 1 \leq i < j \leq k \quad (2)$$

として与えられる。ここで $d_{ij} = v_{ii} - 2v_{ij} + v_{jj}$ であり、 $t_{p,I}^2$ は、 $\mathbf{V} = \mathbf{I}$ (すなわち、 $d_{ij} = d$)のときの次のような $T_{\max \cdot p}^2$ 統計量の上側 $100\alpha\%$ 点である。

$$T_{\max \cdot p}^2 = \max_{i < j} \{(\mathbf{x}_i - \mathbf{x}_j)'(d_{ij} \mathbf{S})^{-1}(\mathbf{x}_i - \mathbf{x}_j)\}.$$

この手法の重要な性質の一つとして、「(2)の同時信頼区間は常に保守的である」という多変量一般化Tukey予想が知られており、 $k = 3, 4$ のときに成り立つことが理論的にSeo, Mano and Fujikoshi(1994), Nishiyama and Seo(2008)によってそれぞれ示されている。さらに、この手法の保守性の程度も議論されており、特に $k = 4$ の場合において以下の定理が得られている。

定理 1. (Nishiyama and Seo(2008))

任意の固定された正定数 t に対して、確率 $Q(t, \mathbf{V}, \mathbb{C}) = \Pr\{(\mathbf{X}\mathbf{c})'(\nu\mathbf{S})^{-1}(\mathbf{X}\mathbf{c}) \leq t(\mathbf{c}'\mathbf{V}\mathbf{c}) \text{ for any } \mathbf{c} \in \mathbb{C}\}$ とする。ただし、 $\mathbb{C} = \{\mathbf{c} \in \mathbf{R}^k : \mathbf{c} = \mathbf{e}_i - \mathbf{e}_j, 1 \leq i < j \leq k\}$ であり、 \mathbf{e}_i は i 番目の要素が1である k 次元単位ベクトルである。 $k = 4$ のとき、任意の正定値行列 \mathbf{V} について次の不等式が成り立つ。

$$1 - \alpha = Q(t_p^*, \mathbf{I}, \mathbb{C}) \leq Q(t_p^*, \mathbf{V}, \mathbb{C}) < Q(t_p^*, \mathbf{V}_0, \mathbb{C}).$$

ここで $t_p^* = t_{p,I}^2/\nu$ であり, V_0 はある i, j, ℓ (i, j, ℓ はすべて異なる) に対して “ $\sqrt{d_{ij}} = \sqrt{d_{i\ell}} + \sqrt{d_{j\ell}}$ かつ $\sqrt{d_{ij}} = \sqrt{d_{im}} + \sqrt{d_{jm}}$ ” の一つを満たす行列である.

また, 対照比較の場合においても対比較の場合と同様, Seo (1995) によって修正 2 次近似法が提案され, さらに多変量 Tukey-Kramer 法に関連して次の同時信頼区間が常に保守的であると予想されている.

$$\mathbf{a}'(\boldsymbol{\mu}_i - \boldsymbol{\mu}_k) \in \left[\mathbf{a}'(\hat{\boldsymbol{\mu}}_i - \hat{\boldsymbol{\mu}}_k) \pm t_{c,V_1} \sqrt{d_{ik} \mathbf{a}' \mathbf{S} \mathbf{a}} \right],$$

$$\forall \mathbf{a} \in \mathbb{R}^p - \{0\}, 1 \leq i \leq k-1. \quad (3)$$

ただし, t_{c,V_1}^2 は, V がすべての i, j ($1 \leq i < j \leq k-1$) に対して $d_{ij} = d_{ik} + d_{jk}$ を満たす行列 V_1 のときの, 次のような $T_{\max-c}^2$ 統計量の上側 $100\alpha\%$ 点である.

$$T_{\max-c}^2 = \max_{i=1, \dots, k-1} \{(\mathbf{x}_i - \mathbf{x}_k)'(d_{ik} \mathbf{S})^{-1}(\mathbf{x}_i - \mathbf{x}_k)\}.$$

この予想の理論的な証明, および保守性の程度が $k = 3, 4, 5$ の場合で議論されており, 特に $k = 5$ の場合において以下の定理が得られている.

定理 2. (Nishiyama (2011))

任意の固定された正定数 t に対して, 確率 $Q(t, V, \mathbb{D}) = \Pr\{(\mathbf{X}\mathbf{d})'(\nu\mathbf{S})^{-1}(\mathbf{X}\mathbf{d}) \leq t(\mathbf{d}'\mathbf{V}\mathbf{d}) \text{ for any } \mathbf{d} \in \mathbb{D}\}$ とする. ただし, $\mathbb{D} = \{\mathbf{d} \in \mathbb{R}^k : \mathbf{d} = \mathbf{e}_i - \mathbf{e}_k, 1 \leq i \leq k-1\}$ である. $k = 5$ のとき, 任意の正定値行列 V について次の不等式が成り立つ.

$$1 - \alpha = Q(t_c^*, V_1, \mathbb{D}) \leq Q(t_c^*, V, \mathbb{D}) < Q(t_c^*, V_2, \mathbb{D}).$$

ここで $t_c^* = t_{c,V_1}^2/\nu$ であり, V_1 はすべての i, j ($1 \leq i < j \leq 4$) に対して $d_{ij} = d_{i5} + d_{j5}$ を満たす行列, V_2 はすべての i, j ($1 \leq i < j \leq 4$) に対して $\sqrt{d_{ij}} = |\sqrt{d_{i5}} - \sqrt{d_{j5}}|$ を満たす行列である.

確率不等式を扱う難しい問題であり, これらの研究成果は高く評価される. 多重比較法の国際会議は, 数年毎に開催され, 2 年ほど前には東京理科大学で開催された. 医学, 薬学等の分野では, 重要な統計的方法であり, その分野に関連して研究の発展が続くと考えられる.

5. 多元分割表における対称性のモデルと分解

多元 r^T 分割表 ($T \geq 3$) において, X_k を k 番目の確率変数とし, $\mathbf{i} = (i_1, \dots, i_T)$ セル確率を $p_{\mathbf{i}}$ とする ($i_k = 1, \dots, r; k = 1, \dots, T$). 対称 (S^T) モデルは次のように定義される: 任意の \mathbf{i} に対して

$$p_{\mathbf{i}} = p_{\mathbf{j}}.$$

ただし, $\mathbf{j} = (j_1, \dots, j_T)$ は $\mathbf{i} = (i_1, \dots, i_T)$ の任意の並べ替えである.

任意に固定した h ($h = 1, \dots, T-1$) に対して, h 次準対称 (QS_h^T) モデルは次のように定義される (Bhapkar and Darroch, 1990): 任意の \mathbf{i} に対して

$$\log p_{\mathbf{i}} = u + \sum_{k=1}^T u_{k(i_k)} + \sum_{1 \leq k_1 < k_2 \leq T} u_{k_1 k_2(i_{k_1}, i_{k_2})}$$

$$+ \dots + \sum_{1 \leq k_1 < \dots < k_h \leq T} u_{k_1 \dots k_h(i_{k_1}, \dots, i_{k_h})} + u_{12 \dots T}(\mathbf{i}).$$

ただし $u_{12\dots T}(i) = u_{12\dots T}(j)$ であり, $j = (j_1, \dots, j_T)$ は $i = (i_1, \dots, i_T)$ の任意の並べ替えである. h 次周辺確率 $P(X_{s_1} = i_1, \dots, X_{s_h} = i_h)$ を p_i^s とする ($h = 1, \dots, T-1$). ここに, $s = (s_1, \dots, s_h)$ ($1 \leq s_1 < \dots < s_h \leq T$), $i = (i_1, \dots, i_h)$ ($i_k = 1, \dots, r; k = 1, \dots, h$) である. h 次周辺対称 (MH_h^T) モデルは次のように定義される: 任意の $s = (s_1, \dots, s_h)$, $t = (t_1, \dots, t_h)$ に対して,

$$p_i^s = p_j^s = p_i^t = p_j^t.$$

ただし, $j = (j_1, \dots, j_h)$ は $i = (i_1, \dots, i_h)$ の任意の並べ替えである (Bhaskar and Darroch, 1990).

モデル M に対する尤度比カイ二乗統計量を $G^2(M)$ と記す. Tomizawa and Tahata (2007) は次の定理を得た.

定理1 任意の h ($h = 1, \dots, T-1$) に対して, $G^2(S^T)$ は $G^2(QS_h^T)$ と $G^2(MH_h^T)$ の和に漸近的に同等である.

また, 点対称 (P^T) モデルは次のように定義される: 任意の i に対して

$$p_i = p_{i^*}.$$

ただし $i^* = (i_1^*, \dots, i_T^*)$, $i_k^* = r + 1 - i_k$ である. Tahata and Tomizawa (2008) は, 任意に固定した h ($h = 1, \dots, T-1$) に対して, h 次準点対称 (QP_h^T) モデルと h 次周辺点対称 (MP_h^T) モデルを導入し (詳細は略), 次の定理を与えた.

定理2 任意の h ($h = 1, \dots, T-1$) に対して, $G^2(P^T)$ は $G^2(QP_h^T)$ と $G^2(MP_h^T)$ の和に漸近的に同等である.

対称性のモデルは、いろいろとある。例えば、歯学の口腔外科では、顔の歪んでいる人に手術を行うことにより、顔の歪みを治す。その際、顔が左右対称であるということ、どのように捉えるかは重要になってくる。また、年をとると緑内障になる人が増えてくるが、片方の眼だけが緑内障ということはほとんどなく、左右共に緑内障にかかる。病状によっては手術をして治すことになる。術後、何時点かでの治り具合は、左右でどうか等を調べ、手術が成功であったか否かを判断する。この分野での富澤 貞男氏、田畑 耕治氏を中心とした理論的な研究成果は高く評価される。これらの研究成果を、いろいろなデータに適用することにより、さらなる研究の深化と進展が期待される。

6. 判別分析および正準相関分析における固有値の高次元漸近分布

判別分析および正準相関分析における固有値・固有ベクトルの分布の大標本での漸近分布は次元が大きくなるにつれて近似が悪くなる。一般に、このような問題点を克服する方法として、次元と標本数が共に大としたときの高次元漸近分布が導出されている。母集団固有値が単根である場合には, Fujikoshi, Himeno and Wakaki (JSPI, 2008), Fujikoshi and Sakurai (JMA, 2008) によって求められている。また, 母集団固有値に重複性がある場合への拡張も試みられている (藤越 (2011)).

まず $(q+1)$ 個の p 次元母集団の正準判別分析における固有値について考える。群間平方和積和行列を S_b , 群内平方和積和行列を S_w とする。このとき, 正準判別分析においては, $S_w^{-1}S_b$ の第 i 固有値 ℓ_i と対応する固有ベクトル h_i , すなわち, 固有方程式

$$S_b h_i = \ell_i S_w h_i, \quad h_i' S_w h_j = n \delta_{ij}, \quad i, j = 1, \dots, p,$$

の解が重要な役割を演じる. ここに, $\ell_1 > \dots > \ell_t > \ell_{t+1} = \dots = \ell_p = 0$, $t = \min(p, q)$ で n は全標本数である. 正規性のもとで, S_b は非心ウィシャート分布 $W_p(q, \Sigma; \Sigma^{1/2}\Omega\Sigma^{1/2})$ に従い, S_w はウィシャート分布 $W_p(n - q - 1, \Sigma)$ に従い, これらは互いに独立である. ここに, 非心行列 Ω は

$$\Sigma^{1/2}\Omega\Sigma^{1/2} = \sum_{i=1}^{q+1} n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})', \quad \bar{\boldsymbol{\mu}} = \sum_{i=1}^{q+1} (n_i/n)\boldsymbol{\mu}_i$$

として定義される. Ω の固有値 $\omega_1 \geq \dots \geq \omega_p \geq 0$ について

$$\begin{aligned} \omega_1 = \dots = \omega_{q_1} = n\lambda_1, \quad \omega_{q_1+1} = \dots = \omega_{q_1+q_2} = n\lambda_2, \quad \dots, \\ \omega_{q-q_r+1} = \dots = \omega_q = n\lambda_r = 0, \quad \lambda_1 > \lambda_2 > \dots > \lambda_r = 0 \end{aligned}$$

とする. 標本固有値を

$$Z_i = (\sqrt{m}/\sigma_\alpha)(\ell_i - \mu_\alpha), \quad i \in J_\alpha = \{q_1 + \dots + q_{\alpha-1} + 1, \dots, q_1 + \dots + q_\alpha\},$$

$\alpha = 1, \dots, r$ と基準化する. ここに, $m = n - p + q$, $\mu_\alpha = c + (n/m)\lambda_\alpha$, $c = p/n$, $\sigma_\alpha^2 = 2c(1+c) + 4(1+c)(n/m)\lambda_\alpha + 2(n/m)^2\lambda_\alpha^2$. このとき, 高次元漸近的枠組

$$c = p/n \rightarrow c_0 \in [0, 1)$$

のもとで, (Z_1, \dots, Z_q) の漸近確率密度関数は次のように与えられる.

$$\prod_{\alpha=1}^r \frac{\pi^{q_\alpha(q_\alpha-1)/4}}{2^{q_\alpha/2}\Gamma_{q_\alpha}(\frac{1}{2}q_\alpha)} \exp\left(-\frac{1}{2} \sum_{i \in J_\alpha} z_i^2\right) \prod_{i < j; i, j \in J_\alpha} (z_i - z_j).$$

とくに, 第 i 母集団固有値 $\omega_i = n\lambda_i$ が単根であるとして

$$Z_i = \frac{\sqrt{n}}{\sigma_i} (\ell_i - \lambda_i), \quad \sigma_i^2 = 4\lambda_i + 2\lambda_i^2$$

とおく. このとき, Z_i の漸近分布は標準正規分布である.

さらに, これらの結果において, $c = 0$ とすると大標本の場合の漸近的結果 (Siotani et al. (1985) を参照) が得られる.

次に 2 組の変数 $\boldsymbol{x} = (x_1, \dots, x_p)'$ と $\boldsymbol{y} = (y_1, \dots, y_q)'$, $p \geq q$ 間の正準相関係数の分布を考える. 母集団共分散行列 Σ , および, 大きさ $N = n + 1$ の標本に基づく標本共分散行列 S を

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

とする. 母集団正準相関係数 $\rho_1 \geq \rho_2 \geq \dots \geq \rho_q \geq 0$ の 2 乗は固有方程式

$$|\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \rho_i^2\Sigma_{11}| = 0$$

の解である. 標本正準相関係数 $r_1 > r_2 > \dots > r_q > 0$ の 2 乗は固有方程式

$$|S_{12}S_{22}^{-1}S_{21} - r_i^2S_{11}| = 0$$

の解である.

第 i 番目の母正準相関係数 ρ_i が単根であると, 高次元漸近的枠組

$$q : \text{固定}, c = p/n \rightarrow c_0 \in (0, 1)$$

のもとで次が成り立つ (Fujikoshi and Sakurai (2008)).

$$\sqrt{n}(r_i - \tilde{\rho}_i) / (\sigma_i \tilde{\rho}_i^{-1} / 2) \xrightarrow{d} N(0, 1).$$

ここに

$$\tilde{\rho}_i = \{\rho_i^2 + c(1 - \rho_i^2)\}^{1/2}, \quad \sigma_i^2 = 2(1 - c)(1 - \rho_i^2)^2 \{2\rho_i^2 + c(1 - 2\rho_i^2)\}.$$

とくに, $p/n \rightarrow 0$ とき, 大標本の結果

$$\sqrt{n}(r_i - \rho_i) / (1 - \rho_i^2) \xrightarrow{d} N(0, 1)$$

が得られる.

上述の高次元の漸近分布の結果は漸近展開に拡張され (Fujikoshi and Sakurai (2008)), さらに, 母集団固有値が重根をもつ場合に拡張されている (藤越 (2011)). 高次元近似と大標本近似を数値的比較すると, 次の数値的精度から多くの場合において高次元近似が優れていることがわかる.

数値的精度: 最大正準相関係数の 95 % 点近似

$$q = 3, \rho_1 = 0.9, \rho_2 = 0.5, \rho_3 = 0.3$$

N	p	a_{L0}	a_{L1}	a_{H0}	a_{H1}
50	3	0.058	0.042	0.032	0.042
50	7	0.120	0.033	0.035	0.044
50	17	0.462	0.008	0.040	0.044
50	27	0.886	0.000	0.047	0.041

(a_{L0} : 大標本極限分布近似, a_{L1} : 大標本漸近展開近似)

(a_{H0} : 高次元極限分布近似, a_{H1} : 高次元漸近展開近似)

7. 高次元漸近展開近似の誤差限界

統計量の標本分布などについて, 標本数 n を無限大にしたときの極限分布を第 1 次近似とする漸近展開近似の典型的な形は

$$P(T_n \leq x) = G(x) + \sum_{j=1}^{k-1} a_j(x)g(x)n^{-j} + R_{k,n}(x),$$

または, n を \sqrt{n} で置き換えたものである. ここに, g は極限分布関数 G の確率分布関数である. このような漸近展開はエッジワース Edgeworth 展開と呼ばれる. 誤差の評価については, 通常オーダー評価が与えられる. 誤差項のオーダー評価は, 例えば, $R_{k,n}(x) = O(n^{-k})$ と表せるが, これは

$$|R_{k,n}(x)| \leq B_k n^{-k}, \quad n \geq n_0$$

を満たす十分大きな定数 B_k, n_0 の存在を保証するものである。しかし、通常、 B_k, n_0 の存在性が主張されるだけであって、それらの値は与えられていない。従って、この種の誤差項の評価では、与えられた n に対して誤差がどの程度であるか依然として未知のままである。そこで、「与えられた n に対して、できるだけ小さい B_k を与える」誤差限界を求めることが重要になる。分布がパラメータに依存する場合の誤差限界はパラメータの既知関数として求めることになる。このような結果の代表的なものとして、独立和の分布の正規近似に対する誤差限界である Berry-Esseen 限界がある。また、尺度混合変数、あるいは、その関数と表せるものについて、ある種の漸近展開近似とその誤差限界を与える方法が発展してきている。これによって、MANOVA 統計量の漸近展開近似等に関して誤差限界が求められるようになってきている。このような結果に関する総合報告としては、例えば、Fujikoshi (2004) を参照されたい。

線形判別関数の高次元近似に対しては、位置尺度混合変数の近似とその誤差限界に関する結果が利用されている。2群の判別問題における初期データ

$$\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i} \sim \text{i.i.d. } \Pi_i; N_p(\boldsymbol{\mu}_i, \Sigma_i), \quad i = 1, 2$$

が与えられているとする。 \mathbf{x} は $\mathbf{x} \in \Pi_1$ あるいは $\mathbf{x} \in \Pi_2$ のいずれかであるが所属不明であって線形判別関数を用いて通常

$$W \equiv (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} \left[\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \right] \geq 0 \Leftrightarrow \mathbf{x} \in \Pi_1$$

と判別する。ここに、 $\bar{\mathbf{x}}_i = (1/n_i) \sum_{j=1}^{n_i} \mathbf{x}_{ij}$, $i = 1, 2$, $S_i = \{1/(n_i - 1)\} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'$, $i = 1, 2$, $S = \{1/(n - 2)\} \sum_{i=1}^2 (n_i - 1)S_i$, $n = n_1 + n_2$ である。このとき、平均誤判別確率

$$P(2|1) = P(W < 0 | \mathbf{x} \in \Pi_1), \quad P(1|2) = P(W > 0 | \mathbf{x} \in \Pi_2)$$

の高次元近似とその誤差限界に関心がある。ここでの高次元近似とは

$$p/n_i \rightarrow c_i \in (0, 1), \quad i = 1, 2$$

のもとでの近似である。

いま $\mathbf{x} \in \Pi_1$ とすると、 W を

$$W = V^{1/2}Z - U$$

と表すことができる。ここに Z, V, U は

$$V = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} \Sigma S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \quad Z = V^{-1/2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} (\mathbf{x} - \boldsymbol{\mu}_1)$$

$$U = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} (\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1) - \frac{1}{2} D^2, \quad D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

であり、

$$Z \sim N(0, 1), \quad V > 0, \quad Z \perp (U, V)$$

を満たす。このような W は位置尺度混合変数であると呼ばれる。このとき、任意の $u, v > 0$ に対して次が成り立つ (Fujikoshi (2000)).

$$|P(X \leq x) - \Phi(v^{-1/2}(x + u))| \leq B.$$

ここに

$$B = \frac{1}{2\sqrt{2\pi e}} v^{-1} E[(U - u)^2] + \frac{1}{2} v^{-2} E[(V - v)^2] \\ + \frac{1}{2\sqrt{2\pi}} v^{-3/2} \sqrt{E[(U - u)^2] E[(V - v)^2]}.$$

この結果において, $(u, v) = (E(U), E(V))$ と定めると, すなわち

$$u = -\frac{n-2}{2(m-1)} \left(\Delta^2 + \frac{(n_1 - n_2)p}{n_1 n_2} \right), \quad v = \frac{(n-2)^2(n-3)}{m(m-1)(m-3)} \left(\Delta^2 + \frac{np}{n_1 n_2} \right)$$

と定めると, 高次元近似が得られる. ここに, $m = n - p - 2$, $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. また, 誤差項 B (省略) も具体的に求められ, $B = O_1^*$ である. ここで, O_j^* は $(n_1^{-1}, n_2^{-1}, p^{-1})$ に関して j 次の項である. 位置尺度混合変数の漸近近似と誤差評価について述べたが, その精密化である漸近展開に対しても拡張されている.

次に大きさ N の標本に基づく多変量尤度比基準 λ で, $\Lambda = \lambda^{2/N}$ の h 次のモーメントが

$$E(\Lambda^h) = \left(\frac{\prod_{k=1}^b y_k^{y_k}}{\prod_{j=1}^a x_j^{x_j}} \right)^h \frac{\prod_{j=1}^a \Gamma[x_j + \xi_j + h] \prod_{k=1}^b \Gamma[y_k + \eta_k]}{\prod_{j=1}^a \Gamma[x_j + \xi_j] \prod_{k=1}^b \Gamma[y_k + \eta_k + h]}$$

と表される尤度比基準の分布を考える. ここに, $\sum_{j=1}^a x_j = \sum_{k=1}^b y_k$, $x_j = c_j N$, $i = 1, \dots, a$; $y_k = d_k N$, $k = 1, \dots, b$. このようなモーメントは Box クラスに属すると呼ばれる. 例えば, MANOVA における尤度比基準 λ の帰無分布は

$$\lambda^{1/N} = \Lambda = |S_e| / |S_e + S_h| \sim \Lambda_{p,q,n} \equiv \Lambda_{q,p,n-p+q}$$

と表せる. ここに, $S_e \sim W_p(n, \Sigma)$, $S_h \sim W_p(q, \Sigma)$, $S_e \perp S_h$, $n = N - q - 1$. Λ のモーメントは次のように表せる.

$$E[\Lambda^h] = \prod_{j=1}^q \frac{\Gamma[\frac{1}{2}(n-p+j) + h]}{\Gamma[\frac{1}{2}(n-p+j)]} \frac{\Gamma[\frac{1}{2}(n+j)]}{\Gamma[\frac{1}{2}(n+j) + h]}.$$

このようなモーメントをもつ統計量に対して, 大標本漸近的枠組のもとで, $-2 \log \lambda$ のカイ 2 乗分布を第 1 近似とする漸近展開が知られている. 一方, 最近, 高次元の枠組で, 正規分布を第 1 近似とする漸近展開が求められるようになり, さらにそれらの誤差限界が求められている. 誤差限界に関する研究には, Wakaki (2008) [Proceedings of the Conference in Kayeseri; " $\Sigma = \Sigma_0$ ", " $\Sigma = \sigma^2 I_p$ ", " $\Sigma_1 = \dots = \Sigma_k$ "] に始まり, Kato, Yamada and Fujikoshi (2010) [JMA; $\Sigma = \sigma^2 \{(1 - \rho) I_p + \rho \mathbf{1}_p \mathbf{1}'_p\}$, Akita, Jin and Wakaki (2010) [JMA; " $\Sigma = \text{diag}(\Sigma_{11}, \dots, \Sigma_{kk})$ "], Wakaki and Fujikoshi (2011) [personal communication; 正準相関分析における冗長性検定], 等の研究がある. 導出においては, $\log \Gamma(a + b)$ のテイラー展開

$$\log \Gamma(a + b) = \log \Gamma(a) + \sum_{k=1}^{\infty} \frac{1}{k!} \psi^{(k-1)}(a) b^k$$

を用いる. ここに, $\psi(z) = (d/dz) \log \Gamma(z)$, $\psi^{(s)}(a) = d^s \psi(z)/dz^s|_{z=a}$ であって

$$\psi^{(s)}(a) = \begin{cases} -C + \sum_{k=0}^{\infty} \left(\frac{1}{1+k} - \frac{1}{k+a} \right), & s = 0 \\ \sum_{k=0}^{\infty} \frac{(-1)^{s+1} s!}{(k+a)^{s+1}}, & s = 1, 2, \dots \end{cases}$$

と表され, C はオイラー定数である.

以下簡単のため, MANOVA における尤度比基準の帰無分布の場合で考えることにする. $V = -\log \Lambda$ の特性関数を $C(t)$ とする. 一般に, $V = -\log \Lambda$ のキュムラント $\kappa^{(s)}$ を用いて

$$\log C(t) = \sum_{s=1}^{\infty} \frac{\kappa^{(s)}}{s!} (it)^s$$

と展開されるが, いまの場合

$$\begin{aligned} \kappa^{(1)} &= \sum_{j=1}^a x_j \log x_j - \sum_{k=1}^b y_k \log y_k - \sum_{j=1}^a \psi^{(0)}(x_j + \xi_j) + \sum_{k=1}^b \psi^{(0)}(y_k + \eta_k), \\ \kappa^{(s)} &= (-1)^s \left\{ \sum_{j=1}^a \psi^{(s-1)}(x_j + \xi_j) - \sum_{k=1}^b \psi^{(s-1)}(y_k + \eta_k) \right\}, \quad s \geq 2 \end{aligned}$$

と陽に書ける. これより, 基準化された統計量 $Z = (-\log \Lambda - \kappa^{(1)})/(\kappa^{(2)})^{1/2}$ の特性関数は次のように展開される.

$$C_Z(t) = \exp\left(-\frac{t^2}{2}\right) \left\{ 1 + \sum_{k=1}^{\infty} \frac{1}{k!} (it)^{3k} \sum_{j=0}^{\infty} \gamma_{k,j} (it)^j \right\}.$$

ここに, $\tilde{\kappa}^{(s)} = \kappa^{(s)}/(\kappa^{(2)})^{s/2}$, $s = 0, 1, \dots$,

$$\gamma_{k,j} = \sum_{s_1 + \dots + s_k = j} \frac{\tilde{\kappa}^{(s_1+3)} \dots \tilde{\kappa}^{(s_k+3)}}{(s_1+3)! \dots (s_k+3)!}$$

したがって, s 次の項までの近似を用いた特性関数の近似

$$C_{Z,s}(t) = \exp\left(-\frac{t^2}{2}\right) \left\{ 1 + \sum_{k=1}^s \frac{1}{k!} (it)^{3k} \sum_{j=0}^{s-k} \gamma_{k,j} (it)^j \right\}$$

を考え, それを反転することによって高次元近似

$$P(Z \leq x) \approx Q_s(x) \equiv \Phi(x) - \phi(x) \left\{ \sum_{k=1}^s \frac{1}{k!} \sum_{j=0}^{s-k} \gamma_{k,j} h_{3k+j-1}(x) \right\}$$

が提案される. ここに, h_j はエルミート多項式であって, $s = 1, 2$ のときの Q は

$$Q_0(x) = \Phi(x), \quad Q_1(x) = \Phi(x) - \frac{1}{6} \phi(x) \tilde{\kappa}^{(3)} (x^2 - 1)$$

である。このような近似の誤差評価および誤差限界が求められるようになってきているが、それには次の評価式を利用する。

$$\begin{aligned} \sup_x |P(T \leq x) - Q_s(x)| &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{|t|} |C_Z(t) - C_{Z,s}(t)| dt \\ &= \frac{1}{2\pi} (I_1[v] + I_2[v] + I_3[v]), \end{aligned}$$

が成り立つ。ここに、 $0 < v < 1$, $m = \frac{1}{2}(n - p - \frac{1}{2})\kappa^{(2)}$

$$\begin{aligned} I_1[v] &= \int_{-mv}^{mv} \frac{1}{|t|} |C_Z(t) - C_{Z,s}(t)| dt, \\ I_2[v] &= \int_{|t| > mv} \frac{1}{|t|} |C_{Z,s}(t)| dt \quad \text{and} \quad I_3[v] = \int_{|t| > mv} \frac{1}{|t|} |C_Z(t)| dt \end{aligned}$$

これらの評価式を用いて、オーダーが $O(m^{-s})$ であるような誤差限界 B_s が求められている。

2組の変数 $\mathbf{x}; p \times 1$, $\mathbf{y}; q \times 1$ の正準相関分析における $\mathbf{x}_1; p_1 \times 1$ と $\mathbf{y}_1; q_1 \times 1$ の十分性、あるいは、残りの変数の冗長性の尤度比基準に対して、キュムラントは

$$\begin{aligned} \kappa^{(s)} &= (-1)^s \left\{ \psi_{p-p_1}^{(s-1)} \left(\frac{n-p_1-q}{2} \right) - \psi_{p-p_1}^{(s-1)} \left(\frac{n-p_1}{2} \right) \right. \\ &\quad \left. + \psi_{q-q_1}^{(s-1)} \left(\frac{n-p_1-q_1}{2} \right) - \psi_{q-q_1}^{(s-1)} \left(\frac{n-q_1}{2} \right) \right\} \end{aligned}$$

となる。また、漸近展開の妥当性に関する十分条件は

$$n - p - q \rightarrow \infty, \quad \frac{(n-p)(n-q)(n-p_1-q_1)}{(n-p_1)(n-q_1)(n-p-q)} > 1$$

として与えられる。さらに、誤差限界 B_s が求められるが、そのオーダーは $O(m^{-s})$ である。ここに

$$m = m_0(\kappa^{(2)})^{1/2}, \quad m_0 = \frac{1}{2}(n - p - q - \frac{1}{2}).$$

誤差限界は数値的に求めることができ、かなりよい誤差限界が求められる。

表：誤差限界

				$n = 50, s = 1$				$n = 50, s = 2$			
p	q	p_1	q_1	bound	v	m	$\kappa_{\lambda}^{(2)}$	bound	v	m	$\kappa_{\lambda}^{(2)}$
5	5	1	1	.0763	.6	3.1	.16	.0706	.65	3.1	.16
5	5	3	3	.1116	.65	2.6	.13	.1138	.7	2.6	.13
10	10	2	2	.0186	.5	5.3	.36	.0109	.5	5.3	.36
10	10	6	6	.0285	.5	4.5	.31	.0195	.55	4.5	.31
15	15	3	3	.0078	.45	6.2	.64	.0034	.5	6.2	.64
15	15	9	9	.0121	.5	5.5	.56	.0063	.55	5.5	.56
20	20	4	4	.0057	.6	5.2	1.1	.0022	.65	5.2	1.1
20	20	12	12	.0085	.6	4.8	1.0	.0039	.65	4.8	1.0

8. 工学分野での多変量解析の応用

Wooding (1956) は、信号のノイズ解析の目的のために多変量の複素正規分布を導入し、Goodman (1963) は、複素正規分布と複素（中心）ウィシャート分布の基本性質を、実数の正規、ウィシャート分布と対比して紹介した。近年、携帯電話や無線 LAN システムで導入されている MIMO (Multiple-Input Multiple-Output) システム技術では、複素正規分布と複素ウィシャート分布が伝送路容量の解析に使われている。MIMO システムとは、送信機と受信機の双方に複数のアンテナを設置したシステムである。

実数の多変量解析でも固有値、固有ベクトルの果たす役割は重要であるが、MIMO システムの伝送路容量の解析において、複素中心ウィシャート行列の固有値分布が利用される（唐沢(2003)）。

実ウィシャート行列の固有値の分布では、James (1960) の定義したゾーナル多項式や行列変数の超幾何関数が分布関数の記述に現れるが、複素ウィシャート行列の固有値分布についても、実数の場合と同様に、James (1964) の定義した複素ゾーナル多項式（シュアー多項式）や複素行列変数の超幾何関数が分布関数の記述に現れる。実数のゾーナル多項式は、1990 年まで明示的な表現が知られていなかったが、複素ゾーナル多項式は、対称群の表現論ではシュアー多項式とも呼ばれ、行列式の有理式による明示的表現が知られていた。Ratanarajha et al. (2005) は、Muirhead (1982) にまとめられている方法を複素ウィシャート行列に適用し、最大固有値と最小固有値の分布関数を求めた。

複素数の確率ベクトル Z が次の密度をもつとき、

$$f(Z) = \frac{1}{\pi^m |\Sigma|} \exp(-Z^* \Sigma^{-1} Z)$$

Z は複素正規分布 $CN(0_m, \Sigma)$ に従うという。ただし、 $|\Sigma|$ は Σ の行列式であって、正定値性 $|\Sigma| > 0$ を仮定する。

複素正規分布 $CN_m(0_m, \Sigma)$ からの n 個の無作為標本を Z_1, Z_2, \dots, Z_n , $n \geq m$ とし、これらを並べた $m \times n$ の確率行列を $A = (Z_1, Z_2, \dots, Z_n)^T$ とする。このとき、 $W = A^* A$ を複素ウィシャート行列といい、この従う分布を、自由度 n 、共分散行列 Σ の複素ウィシャート分布という。複素ウィシャート分布を記号 $\widetilde{W}_m(n, \Sigma)$ で表す。複素ウィシャート行列 W の密度関数は、

$$f(W) = \frac{|W|^{n-m}}{|\Sigma|^n \widetilde{\Gamma}_m(n)} \exp[-\text{tr}(\Sigma^{-1} W)] \quad (8.4)$$

で表される。ただし、 $\widetilde{\Gamma}_m(a)$ を複素多変量ガンマ関数といい、

$$\widetilde{\Gamma}_m(a) = \pi^{\frac{1}{2}m(m-1)} \prod_{i=1}^m \Gamma(a - i + 1) \quad (8.5)$$

で定義する。さらに、 $\Gamma(a)$ は通常のガンマ関数である。

$m \times m$ の行列 Y をエルミート行列とし、その固有値を y_1, y_2, \dots, y_m とする。複素ゾーナル多項式は、対称群の表現論にでてくるシュアー多項式、表現の次元

$$S_\kappa(y_1, \dots, y_m) = \frac{\det(y_j^{\kappa_i + k - j})}{\det(y_j^{k - j})}, \quad \chi_{[\kappa]}(1) = k! \frac{\prod_{i < j}^m (\kappa_i - \kappa_j - i + j)}{\prod_{i=1}^m (\kappa_i + m - i)!}$$

を用いて,

$$\mathcal{C}_\kappa(\mathbf{Y}) = \mathcal{C}_\kappa(y_1, \dots, y_m) = \chi_{[\kappa]}(1) \mathcal{S}_\kappa(y_1, \dots, y_m)$$

で定義される。また, 次式が成立する。

$$\sum_{\kappa \in P_m^k} \mathcal{C}_\kappa(\mathbf{Y}) = (\text{tr } \mathbf{Y})^k$$

複素行列 \mathbf{Y} の超幾何関数 ${}_pF_q$ を複素ゾーナル多項式の無限級数として,

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; X) = \sum_{k=0}^{\infty} \sum_{\kappa \in P_m^k} \frac{[a_1]_\kappa \cdots [a_p]_\kappa \mathcal{C}_\kappa(\mathbf{Y})}{[b_1]_\kappa \cdots [b_q]_\kappa k!} \quad (8.6)$$

で定義する。ただし, $\{a_i\}_{i=1}^p$ と $\{b_i\}_{i=1}^q$ は任意の複素数であり, 分割 κ に対するポツホハンマー記号 $[a]_\kappa$ を, 通常のパツホハンマー記号 $(a)_k = a(a+1)\cdots(a+k-1)$, $(a)_0 = 1$ を用いて

$$[a]_\kappa = \prod_{i=1}^m (a - i - 1)_{\kappa_i},$$

で定義する。さらに, 2つの行列 \mathbf{X}, \mathbf{Y} に対する超幾何関数を

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; X, Y) = \sum_{k=0}^{\infty} \sum_{\kappa \in P_m^k} \frac{[a_1]_\kappa \cdots [a_p]_\kappa \mathcal{C}_\kappa(X) \mathcal{C}_\kappa(Y)}{[b_1]_\kappa \cdots [b_q]_\kappa k! \mathcal{C}_\kappa(I_m)} \quad (8.7)$$

で定義する。

いま, 共分散行列 Σ を正定値と仮定し, したがって, その固有値 $\lambda_1, \dots, \lambda_m$ は全て正としてよい, また, \mathbf{W} の固有値を $\ell_1 > \dots > \ell_m > 0$ とする。このとき, ℓ_1, \dots, ℓ_m の同時密度関数は,

$$f(\ell_1, \dots, \ell_m) = \frac{\pi^{m(m-1)} (\det \Sigma)^{-n}}{\tilde{\Gamma}_m(m) \tilde{\Gamma}_m(n)} \prod_{j=1}^m \ell_j^{n-m} \prod_{j < k} (\ell_j - \ell_k)^2 {}_0F_0(-D, \Sigma^{-1}),$$

で与えられる。

Sugiyama(1972) は, 行列 \mathbf{W} の最大固有値 ℓ_1 の cdf. を次のように与えている。

$$P(\mathbf{S} < z\mathbf{I}) = P(\mathbf{A} < nz\mathbf{I}) = (\tilde{\Gamma}_p(n))^{-1} \int_{0 < \mathbf{B} < x} \Sigma^{-1} \text{etr}(-\mathbf{B}) |\mathbf{B}|^{n-p} (d\mathbf{B}).$$

この結果は, Ratnarajah et al. (2005) によって与えられている次の結果式になる。

$$\Pr(\ell_1 < x) = \frac{\tilde{\Gamma}_m(m)}{\tilde{\Gamma}_m(n+m)} \frac{x^{mn}}{(\det \Sigma)^{n-1}} {}_1F_1(n; n+m; -x \Sigma^{-1}),$$

Sugiyama (1972), Takemura and Sheena (2005) では, 実数の場合に, 最大固有値の分布が自由度 n の χ^2 に従うことが示されているが, 複素の場合も同様のことが示されることを橋口博樹氏が下記のガンマ分布近似の結果を示している。

1. ℓ_1/λ_1 の分布は, 形状パラメータ n , 尺度パラメータ 1 のガンマ分布で近似できる

2. ℓ_m/λ_m の分布は, 形状パラメータ $n - m + 1$, 尺度パラメータ 1 のガンマ分布で近似できる.

また, Sugiyama(1972) は複素正規分布の場合の最大根の χ^2 分布の積による近似式

$$P(\ell_1 < x) = \prod_{i=1}^p P(\chi_{2(n-i+1)}^2 \leq 2nx\Lambda_i)$$

を与えている。さらに、行列Wが Non-central Wishart 分布に従う場合の最大根の近似分布を同様の型で与えている。

1970 年前半に、複素正規分布に基づく論文を書いたりしていたときは、実数の場合の単なる拡張に思えて、いろいろな計算を行いながら、論文として発表しないで終わっていた。2000 年に入り、MIMO システムの伝送路容量の解析が必要であることから、電子工学分野の方々が、複素正規分布の場合の研究を進め、論文を発表している。このように、統計的方法を現実に応用できる分野が生まれてきて、統計学の研究が進むというのが、純粋数学とは異なる点であろう。工学の分野で、複素正規分布の研究がなされたことを例にとって述べたが、同様な統計学の研究が、経済学、経営学、心理学、社会学、医学、薬学、法律学、人類学、農学、環境科学等で行われている。統計数学は、それらの基礎を構成する重要な役割を果たしている。

多変量解析で優れた研究をされている現職の教授方、小西貞則氏、早川 毅氏、竹村 彰通氏、国友 直人氏、青嶋 誠氏、久保川 達也氏等の研究成果に、ページ数等の関係でほとんど触れられなかったことを最後に記しておきます。

参考文献

1. 固有ベクトルの分布

- [1] Anderson, T. W. (1963). *Ann. Math. Statist.*, **34**, 122-148.
- [2] Flury, B.D. (1988). NewYork: Wiley.
- [3] Hayakawa, T. (1978). *Ann. Inst. Statist. Math.*, **30**, 51-55.
- [4] Kollo, T. and von Rosen, D. (2005). Springer.
- [5] Kzanowski, W.J. (1979). *J. Amer. Statist. Assoc.*, **74**, 703-707.
- [6] Schott, J.R. (1988). *Biometrika*, **75**, 229-236.
- [7] Sugiyama, T. (1965). *Ann. Math. Statist.*, **36**, 1875-1876.
- [8] Sugiyama, T. (1966). *Ann. Math. Statist.*, **37**, 995-1001.
- [9] Sugiyama, T. (1971). *Journal Japan Statistical Society*, **1**, 63-68.
- [10] Tsukada, S. (1998). *J. Japan Statist. Soc.*, **28**, 163-174.
- [11] Tsukada, S. (1999). Chuo University, Ph. D. thesis.
- [12] 塚田真一, 杉山高一 (1997). *計算機統計学*, **10**, 19-35.
- [13] 牛澤賢二 (1998). *計算機統計学*, **11**, 77-87.

2. 固有値の分布

- [1] Hino, E., Murakami, H. and Sugiyama, T. (2009). *Commun. Statist. - Simul. Comput.*, **38**, 1675 - 1689.
- [2] James, A. T. (1960). *Ann. Math. Statist.* **31**, 151-158.
- [3] James, A. T. (1964). *Ann. Math. Statist.* **35**, 475-501.

- [4] James, A. T. (1968). *Ann. Math. Statist.*, **39**, No.5 1711-1718.
- [5] Sugiyama, T. (1966). *Ann. Math. Statist.* **37**, 995-1001.
- [6] Sugiyama, T. (1967b). *Ann. Math. Statist.*. **38**, 1148-1151
- [7] Sugiyama, T. (1972). *J. Jpn. Statist. Soc.*, **3**, No.1 1-8.
- [8] Sugiyama, T. (1979). *Computer Science Monographs*, 12. Inst. Stat. Math.
- [9] Sugiyama, T., Takeda, Y. and Fukuda M. (1998). *J. Jpn. Soc. Comp. Statist.*, **11**.
- [10] Sugiyama, T., Fukuda M. and Takeda, Y. (1999). *Commun. Statist*, **28**, 825-837.
- [11] Takeda, Y. (2001). *J. Jpn. Soc. Comp. Statist.*, **14**, 1-10.
- [12] Fujikoshi, Y. (1977). *Ann. Inst. Statist. Math.*, **29**, Part A, 379-387.
- [13] Hashiguchi, H., Nakagawa, S. and Niki, N. (2000). *Math. Comp. Sim.* **51**, 489-496.
- [14] Hashiguchi, H. and Niki, N. (2006). *J. Japan. Soc. Comput. Statist.*, **19**, 45-56.
- [15] 早川 毅 (1971). *数学* **23**, 1-16.
- [16] 早川 毅 (1983). *一橋大学研究年報, 自然科学研究* **22**. 117-202.
- [17] Khatri, C. G. (1972). *J. Multivariate. Anal.*, **2**, 201-207.
- [18] Pillai, K. C. S. and Sugiyama, T. (1969). *Ann. Inst. Stat. Math.*, **21**, 321-327.
- [19] Sugiura, N. (1973). *Commun. in Statist.*, **1**. 393-417.
- [20] Sugiura, N. (1990). *J. Jpn. Statist. Soc.* **20**, 117-136.
- [21] Sakaori, F. (2002). *Commun. Statist. - Simul. Comput.*, **31**, No.4, 641- 651.
- [22] Sugiyama, T. (1965). *Ann. Math. Statist.*, **36**(6).
- [23] Sugiyama, T. (1967). *Ann. Math. Statist.*, **38**(4), 1148-1151.
- [24] Sugiyama, T. (1970). *Ann. Math. Statist.*, **41**(2), 665-657.
- [25] Sugiyama, T. (1972). *Ann. Math. Statist.*, **14**(1), 17-24.
- [26] Sugiyama, T. (1973). *J. Jpn. Statist. Soc.*, **3**(1), 1-8.

3. 正準相関分析

- [1] Akaike, H. (1973). *Eds. B.N. Petrov and F. Csáki*, 267-281, Budapest: Akadémia.
- [2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, (3rd ed.). John Wiley & Sons, New York.
- [3] Fujikoshi, Y. (1982). *Ann. Inst. Statist. Math., Part A*, **34**, 523-530.
- [4] Fujikoshi, Y. (1985). *Multivariate Analysis*, -VI, 219-236, Elsevier.
- [5] Fujikoshi, Y., Sakurai, T., Kanda, S. and Sugiyama, T. (2009). *Journal of the Institute of Science and Engineering. Chuo University*, **14**, 31-49.
- [6] Ichikawa, M. and Konishi, S. (1999). *British J. of Math. and Statist. Psych.*, **52**, 285-302.
- [7] Konishi, S. and Kitagawa, G. (2008). *Springer*, New York.
- [8] Ogura, T. (2010). *Computational Statistics and Data Analysis*, **54**, 1117 - 1123.
- [9] Ogura, T., Fujikoshi, Y. and Sugiyama, T. (2011). to appear in *Commun. in Statist.*
- [10] Rao, C. R. (1973). *John Wiley*, New York.
- [11] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). American Sciences Press, Ohio.
- [12] Sugiyama, T. and Takeda, Y. (1999). *Bull. ISI 52nd Session, Finland*.
- [13] Sugiyama, T. and Ushizawa, K. (1992). *Commun. Statist.*, **21**, 947-960.

- [14] Sugiyama, T. and Yamada, T. (2006). *Comp. Statist. and Data Anal.*, **50**, 2111-2123.
- [15] Ushizawa, K. and Sugiyama, T. (1994). *J. Jpn Statist. Soc.*, **24**, 151-155.
- [16] Yamamoto, M., Sugiyama, T., Murakami, H. and Sakaori, F. (2007). *Comp. Statist. and Data Anal.*, **51**, 9, 4707 - 4716.

4. 平均ベクトル間の多変量多重比較法

- [1] Nishiyama, T. (2011). *submitted for publication*.
- [2] Nishiyama, T. and Seo, T. (2008). *Am. J. of Math. and M. Sci.*, **28**, 115-130.
- [3] Seo, T. (1995). *Hiroshima Mathematical Journal*, **23**, 387-422.
- [4] Seo, T., Mano, S. and Fujikoshi, Y. (1994). *J. of the Am. Statistical Assoc.*, **89**, 676-679.
- [5] Seo, T. and Siotani, M. (1992). *J. Jpn Statist. Soc.*, **22**, 123-137.
- [6] Siotani, M. (1959). *Ann. Inst. Statist. Math.*, **10**, 183-208.

5. 多元分割表における対称性のモデルと分解

- [1] Bhapkar, V. P. and Darroch, J. N. (1990). *Journal of Multivariate Analysis*, **34**, 173-184.
- [2] Tahata, K. and Tomizawa, S. (2008). *AStA Advances in Statistical Analysis; Journal of the German Statistical Society*, **92**, 255-269.
- [3] Tomizawa, S. and Tahata, K. (2007). *Journal de la Société Française de Statistique*, **148**, 3-36.

6. 判別分析および正準相関分析における固有値の高次元漸近分布

- [1] Fujikoshi, Y., Himeno, T. and Wakaki, H. (2008). *J. Statist. Plann. Inf.*, **138**, 3457-3466.
- [2] Fujikoshi, Y. and Sakurai, T. (2009). *J. Multivariate Anal.*, **100**, 231-242.
- [3] 藤越康祝 (2011). 2011 年度統計関連連合大会; 講演要旨.
- [4] Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985), American Sciences Press.

7. 高次元漸近展開近似の誤差限界

- [1] Akita, T. and Wakaki, H. (2010). *J. Multivariate Anal.*, **101**, 1086-1813.
- [2] Fujikoshi, Y. (2000). *J. Multivariate Anal.*, **73**, 1-17.
- [3] 藤越 康祝 (2003). *J. Jpn. Statist. Soc.*, **33**, 273-306.
- [4] Kato, N., Yamada, T. and Fujikoshi, Y. (2010). *J. Mult. Anal.*, **101**, 101-112.
- [5] Wakaki, H. (2009). In *Proceedings of Int. Conference on Multivariate Stat. Modeling and High Dimensional Data Mining*, Kayeseri, June 19-23, 2008.

8. 工学分野での多変量解析の応用

- [1] 唐沢好男. (2003), 電子情報通信学会論文誌, Vol.J86-B, 1706-1720
- [2] T. Ratnarajah, R. Vaillancourt and M. Alvo, SIAM J. MATRIX ANAL. APPL., 2005, Vol. 26, No. 2, pp. 441-456, (2005)
- [3] Sugiyama, T. (1972). *Ann. Inst. Statist. Math.*
- [4] A. Takemura and Y. Sheena, *Journal of Multivariate Analysis* 94, 271-299, (2005).