RECENT DEVELOPMENT OF EIGENVALUES AND EIGENVECTORS FOR COVARIANCE MATRIX

Takakazu Sugiyama 杉山髙一(中央大学)
Shin-ichi Tsukada 塚田真一(明星大学)
Yuichi Takeda 竹田裕一(神奈川工科大学)
Hidetoshi Murakami 村上秀俊(防衛大学校)

序

$\Pi_g(\mu_g, \Sigma_g)$：平均ベクトル μ_g, 共分散行列 Σ_g の第 g 母集団

$\lambda_j^{(g)}$：共分散行列 Σ_g の第 j 番目の固有値
ただし $\lambda_1^{(g)} > \cdots > \lambda_p^{(g)}$.

$\eta_j^{(g)}$：$\lambda_j^{(g)}$ に対応する固有ベクトル

$\{\boldsymbol{X}_1^{(g)}, \ldots, \boldsymbol{X}_{N_g}^{(g)}\}$：$p$ 変量確率標本

序

標本平均ベクトル：$\bar{\boldsymbol{X}}_g$

標本共分散行列：$S_g = \dfrac{1}{N_g - 1} \displaystyle\sum_{i=1}^{N_g} \left(\boldsymbol{X}_i^{(g)} - \bar{\boldsymbol{X}}_g \right) \left(\boldsymbol{X}_i^{(g)} - \bar{\boldsymbol{X}}_g \right)'$

$l_j^{(g)}$：S_g の j 番目に大きい標本固有値
ただし，$l_1^{(g)} > \cdots > l_p^{(g)}$

$\boldsymbol{h}_j^{(g)}$：$l_j^{(g)}$ に対応する標本固有ベクトル

序

オーストラリアに住む白人とアボリジニの手の大きさや形に違いがあるか

中学２年生と中学３年生の成績に違いがあるか

以下のような仮説において検定を行なう.

$$H_0 : \lambda_j^{(1)} = \cdots = \lambda_j^{(k)},$$
$$H_1 : \text{not } H_0.$$

序

主成分スコアを

$$y_{j\alpha}^{(g)} = \boldsymbol{h}'^{(g)}_{j} \left(\boldsymbol{X}_{\alpha}^{(g)} - \bar{\boldsymbol{X}}^{(g)} \right), \qquad \alpha = 1, \ldots, N_g.$$

と定義し, k 母集団における主成分スコア

$$\begin{aligned} Y_1 &= \left\{ y_{j1}^{(1)}, y_{j2}^{(1)}, \ldots, y_{jN_1}^{(1)} \right\}, \\ Y_2 &= \left\{ y_{j1}^{(2)}, y_{j2}^{(2)}, \ldots, y_{jN_2}^{(2)} \right\}, \\ &\vdots \\ Y_k &= \left\{ y_{j1}^{(k)}, y_{j2}^{(k)}, \ldots, y_{jN_k}^{(k)} \right\} \end{aligned}$$

を得る.

$R_{j\alpha}^{(i)}$ を $y_{j\alpha}^{(g)}$ を昇順に並べた順位とする.

序

$$\mathrm{E}\left[l_j^{(g)}\right] = \lambda_j^{(g)} + \frac{\lambda_j^{(g)}}{N_g}\sum_{i\neq j}\frac{\lambda_i^{(g)}}{\lambda_j^{(g)}-\lambda_i^{(g)}} + O(N_g^{-2}),$$

$$\mathrm{Var}\left[y_{j\alpha}^{(g)}\right] = \lambda_j^{(g)} - \frac{2}{N_g-1}\sum_{i\neq j}^{p}\frac{\mathrm{E}\left[x_{i\alpha}^{(g)2}x_{j\alpha}^{(g)2}\right]}{\lambda_i^{(g)}-\lambda_j^{(g)}} + O(N_g^{-2})$$

ノンパラメトリック検定を行なうにあたり, 主成分スコア間は無相関になる必要があるが, $y_{\alpha i}$ and $y_{\alpha k}$ は無相関ではない.

一般性を失うことなく $E(\boldsymbol{x}_i) = \mathbf{0}$ となる.

$\Sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ かつ λ_k が単根のとき, $y_{\alpha i}$ と $y_{\alpha k}$ の共分散は次のようになる :

序

$$E[y_{\alpha i}y_{\alpha k}] = E\left[\sum_{u=1}^{p}\sum_{v=1}^{p} h_{u\alpha}h_{v\alpha}x_{ui}x_{vk}\right]$$

$$\begin{aligned}
= \quad & -\frac{2}{n^2}\sum_{l\neq\alpha}^{p}\lambda_{l\alpha}^2 m_{\alpha l}^{21}m_{\alpha l}^{21} - \frac{1}{n^2}\sum_{u\neq\alpha}^{p}\lambda_{u\alpha}^2(m_{\alpha u}^{21}m_{\alpha u}^{21} + m_{\alpha}^{3}m_{\alpha u}^{12}) \\
& + \frac{1}{n^2}\sum_{\substack{l,u\neq\alpha \\ u\neq l}}^{p}\lambda_{u\alpha}\lambda_{l\alpha}(m_{ul}^{21}m_{\alpha l}^{21} + m_{ul\alpha}^{111}m_{ul\alpha}^{111}) - \frac{1}{n^2}\sum_{v\neq\alpha}^{p}\lambda_{v\alpha}^2(m_{\alpha}^{3}m_{\alpha v}^{12} + m_{\alpha v}^{21}m_{\alpha v}^{21}) \\
& + \frac{1}{n^2}\sum_{\substack{v,l\neq\alpha \\ v\neq l}}^{p}\lambda_{v\alpha}\lambda_{l\alpha}(m_{vl\alpha}^{111}m_{vl\alpha}^{111} + m_{vl}^{21}m_{\alpha l}^{21}) \\
& + \frac{1}{n^2}\sum_{u,v\neq\alpha}^{p}\lambda_{u\alpha}\lambda_{v\alpha}(m_{\alpha u}^{12}m_{\alpha v}^{12} + m_{\alpha uv}^{111}m_{\alpha uv}^{111}) + O(n^{-3})
\end{aligned}$$

序

ただし, $\lambda_{\alpha\beta} = (\lambda_\alpha - \lambda_\beta)^{-1}$, 3 次モーメントを $m_i^3 = E(x_i x_i x_i)$, $m_{ik}^{21} = E(x_i x_i x_k)$, $m_{ik}^{12} = E(x_i x_k x_k)$, $m_{ikt}^{111} = E(x_i x_k x_t)$ と表す.

対称な母集団に対して

$$E[y_{\alpha i} y_{\alpha k}] = 0$$

となる.

非対称な母集団の場合, n の大きさによって 3 次モーメントへ影響ある.

2母集団の場合(Ansari-Bradley検定)

Sugiyama and Ushizawa (1998) :

$$Z = \frac{W - E(W)}{\sqrt{V(W)}},$$

ただし

$$W = \frac{N_1(N_1 + N_2 + 1)}{2} - \sum_{\alpha=1}^{N_1} \left| R_{j\alpha}^{(1)} - \frac{N+1}{2} \right|.$$

2母集団の場合 (Takeda, 2001)

Takeda (2001) :

$$T_1 = l_j^{(2)} / l_j^{(1)}$$

$j = 1$ に対して, T_1 の精密分布は

$$\begin{aligned}
h(t_1) \;=\; & C(\Sigma_1, n_1) C(\Sigma_2, n_2) n_1^{\tilde{n}_1} n_2^{\tilde{n}_2} \\
& \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} \sum_{\kappa} \sum_{\kappa'} \frac{\left(\frac{p+1}{2}\right)_\kappa}{\left(\frac{n_1+p+1}{2}\right)_\kappa} \frac{\left(\frac{p+1}{2}\right)_{\kappa'}}{\left(\frac{n_2+p+1}{2}\right)_{\kappa'}} \frac{C_\kappa\left(n_1 \tilde{\Sigma}_1^{-1}\right)}{k!} \frac{C_{\kappa'}\left(n_2 \tilde{\Sigma}_2^{-1}\right)}{k'!} \\
& \Big\{ (\tilde{n}_1 + k)(\tilde{n}_2 + k') \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k'\right) \Delta_0^{-1} \\
& - \left(\mathrm{tr}\; n_1 \tilde{\Sigma}_1\right) (\tilde{n}_2 + k') \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k' + 1\right) \Delta_1^{-1} \\
& - (\tilde{n}_1 + k) \left(\mathrm{tr}\; n_2 \tilde{\Sigma}_2\right) \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k' + 1\right) \Delta_1^{0} \\
& + \left(\mathrm{tr}\; n_1 \tilde{\Sigma}_1\right) \left(\mathrm{tr}\; n_2 \tilde{\Sigma}_2\right) \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k' + 2\right) \Delta_2^{0} \Big\},
\end{aligned}$$

で与えられている.

2母集団の場合 (Takeda, 2001)

ただし

$$
\begin{aligned}
C(\mathbf{\Sigma}, n) &= |\mathbf{\Sigma}|^{-\frac{1}{2}n} \Gamma_p\left(\frac{p+1}{2}\right) \Big/ 2^{\frac{1}{2}np} \Gamma_p\left(\frac{n+p+1}{2}\right), \\
\Gamma_p(n) &= \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(n - \frac{1}{2}(i-1)\right), \\
\Delta_j^i &= \frac{t_1^{\tilde{n}_2+k'+i}}{\left(\mathrm{tr}\ n_1\tilde{\Sigma}_1^{-1} + t_1\,\mathrm{tr}\ n_2\tilde{\Sigma}_2^{-1}\right)^{\tilde{n}_1+\tilde{n}_2+k+k'+j}},
\end{aligned}
$$

$\tilde{n}_g = pn_g/2$ and $\tilde{\Sigma}_g^{-1} = \Sigma_g^{-1}/2$ である.

2母集団の場合 (Hino *et al*., 2009)

Hino, Murakami and Sugiyama (2009):

$$T_2 = \sqrt{N} \log(l_j^{(2)}/l_j^{(1)}),$$

ただし $N = N_1 + N_2$ である.

統計的本質は T_1 と同じであるが, 極限分布が標準正規分布となるので, 検定を行なう際に有用である.

極限分布は

$$P\left(\frac{T_2}{\sigma} \leq x\right) = \Phi(x) - \frac{1}{\sqrt{N}}\phi(x)\{a_1\sigma^{-1} + a_3\sigma^{-3}h_2(x)\} + o(N^{-\frac{1}{2}})$$

である.

多母集団の場合 (k標本Ansari-Bradly)

Murakami, Hino and Tsukada (2007) :

N が偶数の場合,

$$AB_{ke} = \frac{48(N-1)}{N(N^2-4)} \sum_{i=1}^{k} N_i \left(\bar{A}_j{}^{(i)} - \frac{N+2}{4} \right)^2 .$$

N が奇数の場合,

$$AB_{ko} = \frac{48N^2}{N(N+1)(N^2+3)} \sum_{i=1}^{k} N_i \left(\bar{A}_j{}^{(i)} - \frac{(N+1)^2}{4N} \right)^2 .$$

ただし, $\bar{A}_j{}^{(i)}$

$$\bar{A}_j{}^{(i)} = \frac{1}{N_i} \sum_{m=1}^{N_i} \left(\frac{N+1}{2} - \left| R_{jm}^{(i)} - \frac{N+1}{2} \right| \right) .$$

である.

多母集団の場合 (k標本Mood検定)

Murakami, Hino and Tsukada (2007) :

$$M_k = \frac{180}{N(N+1)(N^2-4)} \sum_{i=1}^{k} N_i \left(\bar{M}_j^{(i)} - \frac{N^2-1}{12} \right)^2 .$$

ただし,

$$\bar{M}_j^{(i)} = \frac{1}{N_i} \sum_{m=1}^{N_i} \left(R_{jm}^{(i)} - \frac{N+1}{2} \right)^2 .$$

である.

多母集団の場合 (Murakami *et al*., 2008)

Murakami, Tsukada and Takeda (2008) :

$$T_M = \frac{N}{2k} \sum\sum_{1 \le g_1 < g_2 \le k} \left(\log \frac{l_j^{(g_2)}}{l_j^{(g_1)}} \right)^2 .$$

ただし,

$$N = \sum_{g=1}^{k} N_g, \quad l_j^{(g)} = \frac{1}{N_g} \sum_{\alpha=1}^{N_g} \left(y_{j\alpha}^{(g)} - \bar{y}_j^{(g)} \right)^2 \quad \text{and} \quad \bar{y}_j^{(g)} = \frac{1}{N_g} \sum_{\alpha=1}^{N_g} y_{ji}^{(g)}$$

である. また, T_M の極限分布は自由度 $k-1$ の χ^2 分布となる.
(正規母集団の下)

多母集団の場合 (Tsukada and Murakam, 2008)

Tsukada and Murakami(2008) :

$$\sum_{g_1<g_2}^{k} \frac{N r_{g_1} r_{g_2}}{2r} \left(\log l_\alpha^{(g_1)} - \log l_\alpha^{(g_2)}\right)^2,$$

ただし $r_g = N_g/N$, $r = \sum_{g=1}^{k} r_g$ and $N = \sum_{g=1}^{k} N_g$.

- Mood test
- Ansari-Bradley test
- χ^2 criterion (Fligner and Killeen, 1976)
- bootstrap (Boos and Brownie, 1989)
- randomization test (Wludyka and Sa, 2004)
- Levene test (Levene, 1960)
- Brown-Forsythe test (Brown and Forsythe,1974)
- O'Brien's Test(1979)

と検出力の比較を行なった.

シミュレーション

- Multivariate Normal random number

$$N(\mathbf{0}, \boldsymbol{\Sigma}^{(i)})$$

- Multivariate Contaminated Normal random number

$$0.95 \times N(\mathbf{0}, \boldsymbol{\Sigma}^{(i)}) + 0.05 \times N(\mathbf{0}, 9\boldsymbol{\Sigma}^{(i)})$$

- Multivariate Skew Normal random number

$$MSN(\alpha^{(i)}, \mathbf{0}, \boldsymbol{\Sigma}^{(i)})$$

シミュレーション

$$N_1 = N_2 = N_3 = 50,$$

$$N_1 = N_2 = N_3 = 100,$$

$$N_1 = 150, N_2 = 100, N_3 = 50,$$

$$N_1 = 200, N_2 = 150, N_3 = 100$$

p=3の場合の結果を紹介する

k標本Ansari-Bradley 検定と k 標本 Mood 検定に対して100,000回のシミュレーション

Murakami, Tsukada and Takeda (2008) で提案された統計量に対して, 100,000回の繰り返しと 10,000回のパ並べ替えをおこなった.

シミュレーション

Case 1:

$\lambda_1^{(1)} = 6, \lambda_2^{(1)} = 3, \lambda_3^{(1)} = 1; \quad \lambda_1^{(2)} = 6, \lambda_2^{(2)} = 3, \lambda_3^{(2)} = 1; \quad \lambda_1^{(3)} = 6, \lambda_2^{(3)} = 3, \lambda_3^{(3)} = 1,$

Case 2:

$\lambda_1^{(1)} = 10, \lambda_2^{(1)} = 3, \lambda_3^{(1)} = 1; \quad \lambda_1^{(2)} = 8, \lambda_2^{(2)} = 3, \lambda_3^{(2)} = 1; \quad \lambda_1^{(3)} = 6, \lambda_2^{(3)} = 3, \lambda_3^{(3)} = 1,$

Case 3:

$\lambda_1^{(1)} = 9, \lambda_2^{(1)} = 5, \lambda_3^{(1)} = 2; \quad \lambda_1^{(2)} = 6, \lambda_2^{(2)} = 3, \lambda_3^{(2)} = 1; \quad \lambda_1^{(3)} = 6, \lambda_2^{(3)} = 3, \lambda_3^{(3)} = 1,$

Case 4:

$\lambda_1^{(1)} = 9, \lambda_2^{(1)} = 5, \lambda_3^{(1)} = 2; \quad \lambda_1^{(2)} = 7.5, \lambda_2^{(2)} = 4, \lambda_3^{(2)} = 1.5; \quad \lambda_1^{(3)} = 6, \lambda_2^{(3)} = 3, \lambda_3^{(3)} = 1.$

シミュレーション (正規母集団)

Table 1: Normal populations

		$N_1 = N_2 = N_3 = 50$				$N_1 = N_2 = N_3 = 100$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.041	0.242	0.224	0.169	0.046	0.478	0.428	0.327
	AB_{1k}	0.042	0.201	0.184	0.144	0.047	0.395	0.348	0.269
	T_M	**0.040**	**0.302**	**0.269**	**0.206**	**0.045**	**0.593**	**0.531**	**0.415**
$j = 2$	M_{2k}	0.048	0.049	0.335	0.253	0.049	0.050	0.615	0.488
	AB_{2k}	0.048	0.049	0.271	0.211	0.049	0.050	0.511	0.403
	T_M	**0.047**	**0.049**	**0.408**	**0.315**	**0.048**	**0.050**	**0.733**	**0.601**
$j = 3$	M_{3k}	0.056	0.056	0.564	0.445	0.053	0.053	0.867	0.763
	AB_{3k}	0.055	0.055	0.465	0.370	0.052	0.052	0.776	0.664
	T_M	**0.057**	**0.056**	**0.666**	**0.550**	**0.054**	**0.054**	**0.942**	**0.871**

シミュレーション (正規母集団)

Table 2: Normal populations

		$N_1 = 150, N_2 = 100, N_3 = 50$				$N_1 = 200, N_2 = 150, N_3 = 100$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.045	0.367	0.453	0.241	0.047	0.613	0.641	0.420
	AB_{1k}	0.046	0.309	0.375	0.207	0.048	0.518	0.538	0.349
	T_M	**0.043**	**0.458**	**0.408**	**0.307**	**0.046**	**0.737**	**0.687**	**0.529**
$j = 2$	M_{2k}	0.049	0.050	0.695	0.428	0.049	0.050	0.852	0.641
	AB_{2k}	0.049	0.050	0.594	0.361	0.049	0.050	0.761	0.545
	T_M	**0.049**	**0.050**	**0.680**	**0.534**	**0.049**	**0.050**	**0.898**	**0.768**
$j = 3$	M_{3k}	0.055	0.055	0.928	0.703	0.052	0.052	0.985	0.898
	AB_{3k}	0.055	0.055	0.861	0.613	0.051	0.052	0.956	0.823
	T_M	**0.060**	**0.060**	**0.933**	**0.815**	**0.055**	**0.055**	**0.994**	**0.962**

シミュレーション (Skew normal)

Table 6: Skew normal populations ($N_1 = N_2 = N_3 = 100$)

		$r = 0.5$				$r_1 = 0.5, r_2 = 0.3, r_3 = 0.2$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.050	0.336	0.500	0.387	0.050	0.356	0.479	0.371
	AB_{1k}	0.050	0.277	0.409	0.319	0.050	0.293	0.391	0.305
	T_M	**0.050**	**0.423**	**0.609**	**0.486**	**0.049**	**0.448**	**0.586**	**0.466**
$j = 2$	M_{2k}	0.050	0.063	0.578	0.456	0.048	0.058	0.585	0.459
	AB_{2k}	0.050	0.061	0.478	0.376	0.049	0.056	0.482	0.379
	T_M	**0.050**	**0.066**	**0.695**	**0.566**	**0.048**	**0.059**	**0.703**	**0.570**
$j = 3$	M_{3k}	0.053	0.053	0.851	0.743	0.052	0.053	0.859	0.753
	AB_{3k}	0.052	0.053	0.757	0.643	0.052	0.052	0.766	0.653
	T_M	**0.054**	**0.054**	**0.930**	**0.851**	**0.053**	**0.053**	**0.936**	**0.861**

シミュレーション (Skew normal)

Table 7: Skew normal populations ($N_1 = 150, N_2 = 100, N_3 = 50$)

		$r = 0.5$				$r_1 = 0.5, r_2 = 0.3, r_3 = 0.2$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.049	0.268	0.545	0.306	0.049	0.282	0.522	0.291
	AB_{1k}	0.049	0.229	0.455	0.260	0.049	0.241	0.435	0.248
	T_M	**0.049**	**0.336**	**0.511**	**0.385**	**0.048**	**0.354**	**0.486**	**0.367**
$j = 2$	M_{2k}	0.051	0.062	0.647	0.385	0.047	0.054	0.645	0.377
	AB_{2k}	0.051	0.061	0.548	0.326	0.048	0.055	0.545	0.318
	T_M	**0.052**	**0.073**	**0.625**	**0.484**	**0.047**	**0.062**	**0.621**	**0.476**
$j = 3$	M_{3k}	0.055	0.056	0.915	0.679	0.055	0.055	0.925	0.700
	AB_{3k}	0.055	0.055	0.843	0.589	0.054	0.055	0.856	0.609
	T_M	**0.060**	**0.062**	**0.919**	**0.793**	**0.060**	**0.062**	**0.930**	**0.814**