主成分分析を用いて因子負荷行列を計算する際に主要因子が推定できる条件

伊藤雅明（広島大学大学院 工学研究科）
佐藤学（広島県立保健福祉大学 総合教育センター）*

1 はじめに

主成分分析は，本来は因子分析とは異なる目的のための手法であるが，因子分析的に用いられることが多い．そのとき，相関行列 (または分散共分散行列) の固有根の大きい方から対応する固有ベクトルから計算した「因子負荷」を“ 順に ”解釈しようとすることが多い．ここでは，その手順を機械的に適用すると不適切となる例を提示する．

このため，p 次元データ $\boldsymbol{x}$ が因子分析の模型 $\boldsymbol{x} = \boldsymbol{\mu} + \Lambda \boldsymbol{f} + \boldsymbol{u}$ に従うと仮定する．ただし Λ は p 行 k 列の因子負荷行列である．$\boldsymbol{x}$ の母分散共分散行列 Σ は，$E\{\boldsymbol{f}\} = \boldsymbol{0}$, $E\{\boldsymbol{u}\} = \boldsymbol{0}$, $E\{\boldsymbol{f}\boldsymbol{u}'\} = O$, $E\{\boldsymbol{u}\boldsymbol{u}'\} = \Psi$(対角行列) という仮定のもとで，$\Sigma = \Lambda\Lambda' + \Psi$ となる．$D = \mathrm{diag}\,\Sigma$ とおき，$D^{-1/2}\Lambda$, $D^{-1/2}\Psi D^{-1/2}$ を改めて Λ, Ψ とそれぞれ書くことにすると，母相関行列 P は $P = \Lambda\Lambda' + \Psi$ と表される．

Θ を対角成分に P の固有根を大きさの順に並べた対角行列，Q をその固有値に対応する長さが 1 の固有ベクトルを並べた直交行列とすると，$Q'PQ = \Theta$ なので，

$$P = Q\Theta Q' = (Q\Theta^{1/2})(Q\Theta^{1/2})' = \tilde{\Lambda}\tilde{\Lambda}' + E$$

となる．ただし，$\tilde{\Lambda}$ は $Q\Theta^{1/2}$ のはじめの m 列で構成された行列であり，主成分分析で計算した m 因子の“ 因子負荷行列 ”とよばれる．

以下の議論では，Λ は一意に定まるための Anderson と Rubin(1956; Theorem 5.1) が与えた十分条件を満たし $\mathrm{diag}(I - \Lambda\Lambda')$ が正定値になるように与えて，$\Psi = \mathrm{diag}(I - \Lambda\Lambda')$ として P を定める．ここで I は単位行列である．そして $\tilde{\Lambda}$ が Λ を近似できているかを調べる．

*連絡先　〒 723-0053 広島県三原市学園町 1-1 広島県立保健福祉大学 総合教育センター　sato@hpc.ac.jp

2　数値例

固有根の大きい方から対応する固有ベクトルから計算した「因子負荷」を“ 順に ”解釈すると不適切である例を掲げる．これは Sato(1992) の Table 6.1 で与えられた例の拡張である．

$$\Lambda = \begin{pmatrix} x & 0 \\ x & 0 \\ .75 & .15 \\ .75 & .15 \\ .75 & -.15 \\ .75 & -.15 \end{pmatrix} = (\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}) \text{ とおく．}$$

(1) $\tilde{\Lambda}$ の第 2 列が $\boldsymbol{\lambda_2}$ を反映している場合

$x = .72$ のとき．

$$\tilde{\Lambda} = \begin{pmatrix} .777 & \mathbf{.0} & .491 & .394 \\ .777 & \mathbf{.0} & -.491 & .394 \\ .793 & \mathbf{.355} & .0 & -.193 \\ .793 & \mathbf{.355} & .0 & -.193 \\ .793 & \mathbf{-.355} & .0 & -.193 \\ .793 & \mathbf{-.355} & .0 & -.193 \end{pmatrix}.$$

固有根: 3.723　.505　.482　.460

(2) $\tilde{\Lambda}$ の第 3 列が $\boldsymbol{\lambda_2}$ を反映している場合

$x = .69$ のとき．

$$\tilde{\Lambda} = \begin{pmatrix} .754 & .512 & \mathbf{.0} & .411 \\ .754 & -.512 & \mathbf{.0} & .411 \\ .792 & .0 & \mathbf{.355} & -.196 \\ .792 & .0 & \mathbf{.355} & -.196 \\ .792 & .0 & \mathbf{-.355} & -.196 \\ .792 & .0 & \mathbf{-.355} & -.196 \end{pmatrix}.$$

固有根: 3.650　.524　.505　.491

(3) $\tilde{\Lambda}$ の第 4 列が $\boldsymbol{\lambda_2}$ を反映している場合

$x = .66$ のとき．

$$\tilde{\Lambda} = \begin{pmatrix} .732 & .531 & .427 & \mathbf{.0} \\ .732 & -.531 & .427 & \mathbf{.0} \\ .792 & .0 & -.197 & \mathbf{.355} \\ .792 & .0 & -.197 & \mathbf{.355} \\ .792 & .0 & -.197 & \mathbf{-.355} \\ .792 & .0 & -.197 & \mathbf{-.355} \end{pmatrix}.$$

固有根: 3.579　.564　.521　.505

x がわずかに変化しただけで，(1), (2), (3) の場合に分かれる．また，P の第 2 固有根 θ_2 はいずれの場合も 1 より小さいが，(1) の場合よりも (2) や (3) の場合の方が θ_2 が大きいことに注意したい．

3　解析的な結果

ここでは因子負荷行列として，

$$\Lambda = \begin{pmatrix} x & 0 \\ x & 0 \\ b & a \\ b & a \\ b & -a \\ b & -a \end{pmatrix} = (\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2})$$ を扱う．ただし，$0 < x < 1,\ -1 < b < 1,\ b \neq 0,$

$a > 0,\ a^2 + b^2 < 1$ とする．

$P = \Lambda\Lambda' + \mathrm{diag}(I - \Lambda\Lambda')$ の固有根 $\varphi_1, \ldots, \varphi_6$ は

$$\begin{aligned}
\varphi_1 &= 1 + (-a^2 + 3b^2 + x^2 + \sqrt{(-a^2+3b^2+x^2)^2 + 4x^2(a^2+5b^2)}\,)/2 > 1, \\
\varphi_2 &= 1 + 3a^2 - b^2, \\
\varphi_3 &= 1 - x^2 < 1, \\
\varphi_4 &= 1 + (-a^2 + 3b^2 + x^2 - \sqrt{(-a^2+3b^2+x^2)^2 + 4x^2(a^2+5b^2)}\,)/2 < 1, \\
\varphi_5 &= \varphi_6 = 1 - a^2 - b^2 < 1
\end{aligned}$$

である．

$\varphi_1, \ldots, \varphi_4$が重根でないときには，長さを１としていない対応する固有ベクトル$\boldsymbol{\gamma_1}, \ldots, \boldsymbol{\gamma_4}$は，

$$(\boldsymbol{\gamma_1}\ \boldsymbol{\gamma_2}\ \boldsymbol{\gamma_3}\ \boldsymbol{\gamma_4}) = \begin{pmatrix} s & 0 & 1 & t \\ s & 0 & -1 & t \\ 1 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 \\ 1 & -1 & 0 & -1 \\ 1 & -1 & 0 & -1 \end{pmatrix},$$

ただし

$$\begin{aligned}
s &= \frac{x(a^2 + 5b^2 + x^2 + \sqrt{a^4 + 9b^4 + 26b^2x^2 + x^4 + a^2(-6b^2 + 2x^2)}\,)}{b(-a^2 + 3b^2 + 3x^2 + \sqrt{a^4 + 9b^4 + 26b^2x^2 + x^4 + a^2(-6b^2 + 2x^2)}\,)}, \\
t &= \frac{x(a^2 + 5b^2 + x^2 - \sqrt{a^4 + 9b^4 + 26b^2x^2 + x^4 + a^2(-6b^2 + 2x^2)}\,)}{b(a^2 - 3b^2 - 3x^2 + \sqrt{a^4 + 9b^4 + 26b^2x^2 + x^4 + a^2(-6b^2 + 2x^2)}\,)}
\end{aligned}$$

と表される．

φ_2 に対応する固有ベクトル $\boldsymbol{\gamma_2}$ が $\boldsymbol{\lambda_2}$ を反映しているので φ_2 の順位を調べると，次の結果が得られる．

(1) $|b| \leq \sqrt{3}a$ のとき，φ_2 は第 1 または第 2 固有根である．

(2) $|b| > \sqrt{3}a$ のとき，

(2-1) $b^2 - 3a^2 < x^2$ であれば，φ_2 は第 2 固有根である．

(2-2) $\dfrac{(b^2-a^2)(b^2-3a^2)}{a^2+b^2} < x^2 < b^2 - 3a^2$ であれば，φ_2 は第 3 固有根である．

(2-3) $x^2 < \dfrac{(b^2-a^2)(b^2-3a^2)}{a^2+b^2}$ であれば，φ_2 は第 4 固有根である．

(2-4) φ_2 は第 5 および第 6 固有根にはならない．

参考文献

Anderson, T. W. and Rubin H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **5**, 111–150.

Sato, M. (1992). A study of an identification problem and substitute use of principal component analysis in factor analysis. *Hiroshima Mathematical Journal*, **22**, 479–524.