ブロック行列分解による因子分析の諸性質

足立浩平

(大阪大学人間科学研究科)

キーワード: 探索的因子分析・高階数近似・因子得点の円錐

Some Properties of Factor Analysis by Block Matrix Factorization

Kohei ADACHI

(Graduate School of Human Sciences, Osaka University)

Key words: Exploratory factor analysis, Higher rank approximation, Factor score cone

1. ブロック行列分解による因子分析

列中心化された n 個体 $\times p$ 変数のデータ行列を $\mathbf{X}$ と表すと、因子分析(FA)モデルは、次のように書ける。

$$\mathbf{X} \cong \mathbf{F}\mathbf{A}' + \mathbf{U}\Psi' \quad (1)$$

ここで、 $\mathbf{F}$ は n 個体 $\times m$ 因子の共通因子得点行列、 $\mathbf{A}$ は $p \times m$ の負荷行列、 $\mathbf{U}$ は $n \times p$ の独自因子得点行列、 Ψ は対角行列で、その対角要素の二乗は独自分散と呼ばれる。ただし、

$n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m$ (m 次の単位行列)、 $n^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_p$ 、 $\mathbf{F}'\mathbf{U} = \mathbf{O}$ (零行列) (2) という制約条件がおかれる。

(1) が母集団を記述するモデルであれば、その左辺と右辺は等号で結べるが、 $\mathbf{X}$ が標本であれば、(1) のように左辺と右辺は近似記号で結ばなければならない。それは、誤差に相当する $\mathbf{U}\Psi$ が (2) のように制約されるからである。

モデル(1)から次の最小二乗基準が導かれる。

$$f(\mathbf{Z}, \mathbf{B}) = \|\mathbf{X} - \mathbf{F}\mathbf{A}' - \mathbf{U}\Psi'\|^2 = \|\mathbf{X} - \mathbf{Z}\mathbf{B}'\|^2 \quad (3)$$

ここで、 $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ 、 $\mathbf{B} = [\mathbf{A}, \Psi]$ である。(2) の 3 つの条件は

$$n^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{I}_{m+p} \quad (4)$$

にまとめられ、(4) のもとで (3) を最小にする解法が de Leeuw (2004) によって示唆され、Unkel & Trendafilov (2010) と Adachi (2012) によって精緻化されている。(3) の最小化は、データ行列を 2 つのブロック行列に近似的に分解する試みなので、ブロック行列分解による因子分析(BMF-FA)と呼ばれ得る。この BMF-FA の幾つか性質を示すことが、本研究の目的である。

2. 十分統計量としての標本共分散

BMF-FA において $\mathbf{B}$ を推定するためには、データ行列 $\mathbf{X}$ が観測されなくても、変数間の共分散行列 $\mathbf{S} = n^{-1}\mathbf{X}'\mathbf{X}$ が与えられれば十分であることを本節で示す。

まず、所与の $\mathbf{B}$ のもとで (3) を最小化するために、(3) を

$$f(\mathbf{Z} | \mathbf{B}) = (\mathbf{Z} \text{ に無関係な項}) - 2n\text{tr}(n^{-1}\mathbf{X}'\mathbf{Z})\mathbf{B}' \quad (5)$$

と書き換えよう。 $\mathbf{X}\mathbf{B}$ の特異値分解を $n^{-1/2}\mathbf{X}\mathbf{B} = \mathbf{K}\mathbf{D}\mathbf{L}'$ 、ただし、 $\mathbf{D}$ は p 次対角行列であるとする、(5) を最小にする $\mathbf{Z}$ は、

$$\mathbf{Z} = n^{1/2}\mathbf{K}\mathbf{L}' + n^{1/2}\mathbf{K}_{\perp}\mathbf{L}_{\perp}' \quad (6)$$

と表せる。ただし、 $[\mathbf{K}, \mathbf{L}][\mathbf{K}_{\perp}, \mathbf{L}_{\perp}] = \mathbf{I}_{p+m}$ である。ここで、(5) と (6) を見比べると、(5) の最小化は、 $\mathbf{Z}$ を求めなくても、

$$n^{-1}\mathbf{X}'\mathbf{Z} = n^{-1}\mathbf{B}'\mathbf{B}'\mathbf{X}'\mathbf{Z} = \mathbf{B}'^{-1}\mathbf{D}\mathbf{L}' \quad (7)$$

のときに達成されることがわかる。

次に、所与の $\mathbf{Z}$ のもとで (3) の最小化するため、(3) を

$$f(\mathbf{B} | \mathbf{Z}) = n\|\mathbf{X}'\mathbf{Z} - \mathbf{B}\|^2 + (\mathbf{B} \text{ に無関係な項}) \quad (8)$$

と書き換えよう。(7)、(8)、 $\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$ を見比べると、(5) を最小にする $\mathbf{B} = [\mathbf{A}, \Psi]$ は、次式で与えられることがわかる。

$$\mathbf{A} = \mathbf{B}'^{-1}\mathbf{D}\mathbf{L}'\mathbf{H}_m, \quad \Psi = n^{-1}\text{diag}(\mathbf{B}'^{-1}\mathbf{D}\mathbf{L}'\mathbf{H}_p) \quad (9)$$

ここで、 $\mathbf{H}_m = [\mathbf{I}_m, \mathbf{O}_p]'$ 、 $\mathbf{H}_p = [\mathbf{O}_m, \mathbf{I}_p]'$ である。

以上より、(7) と (9) を交互に繰り返せば、 $\mathbf{B}$ の解に収束する

ことがわかる。ここで、(7) と (9) の更新に必要な $\mathbf{L}\mathbf{D}\mathbf{L}'$ は、固有値分解 $\mathbf{S} = \mathbf{L}\mathbf{D}^2\mathbf{L}'$ で求められるため、 $\mathbf{S}$ の十分性がわかる。

3. 既存解法との経験的比較

BMF-FA が既存の最尤 FA と同様の解を導くかどうかを、シミュレーションによって検討する。 $1 \leq m \leq 5$ 、 $4m \leq p \leq 9m$ 、 $8p \leq n \leq 12p$ を満たすように n, m, p をランダムに定めて、 $\mathbf{F}$ と $\mathbf{U}$ の正規性を仮定して生成された 1000 個の $\mathbf{S}$ に、両方法を適用した。そして、 $\mathbf{A}$ の要素、および、 Ψ の対角要素について、両方法間の差の絶対値の平均(絶対偏差)を求めた。その結果、1000 個の解を通した $\mathbf{A}$ の絶対偏差の 99 パーセンタイルは 0.024、平均は 0.004、 Ψ の絶対偏差のそれらは 0.025、0.007 であり、両方法の解はほぼ同じであることが示された。

4. 高階数近似としての因子分析

最小二乗基準(3)は、次のようにも書き換えられる。

$$f(\mathbf{Z} | \mathbf{B}) = \|\mathbf{Z} - \mathbf{X}\mathbf{B}\|^2 + (\mathbf{Z} \text{ に無関係な項}) \quad (10)$$

これは、階数 p の $\mathbf{X}\mathbf{B}$ を、(4) より階数 $m+p$ である $\mathbf{Z}$ によって近似する問題と見なせ、FA が高階数近似であることを表す。

5. 共通・独自因子得点の「円錐」

解(6)は、因子得点 $\mathbf{Z}$ の不確定性を示す。ここで、(6) 右辺のユニークに定まる項 $\mathbf{Z}_1 = n^{1/2}\mathbf{K}\mathbf{L}'$ は、 $\mathbf{Z}_1 = \mathbf{X}\mathbf{B}\mathbf{L}\mathbf{L}'^{-1}\mathbf{L}'$ のように $\mathbf{X}$ の関数として表せる。さらに、 $\mathbf{Z}$ と $\mathbf{Z}_1$ の隔たりは、

$$\|\mathbf{Z} - \mathbf{Z}_1\|^2 = \|n^{1/2}\mathbf{K}_{\perp}\mathbf{L}_{\perp}'\|^2 = n\text{tr}(\mathbf{L}_{\perp}'\mathbf{L}_{\perp}) = nm \quad (11)$$

のように一定である。従って、行列をベクトルで表すと、下図のように、 $\mathbf{Z}$ は、 $\mathbf{X}$ と同方向に伸びる $\mathbf{Z}_1$ を中心とした円錐をなす。Mulaik (1976) の円錐が共通因子を表すのに対して、下図は共通・独自因子の円錐を描いている点で異なる。

6. 結語

EMF-FA は、その目的関数(3)が、主成分分析(PCA)の最小二乗基準に $\mathbf{U}\Psi$ を付加した形をとる点で、FA が PCA の拡張であることを示すと同時に、4 節に記したように、低階数近似の PCA とは逆に、FA が高階数近似であることを示す。

文献

- Adachi, K. (2012) *J. Jpn. Soc. Comp. Statist.*, **25**, 25-38.
 de Leeuw, (2014) In *Recent developments of structural equation models*, pp. 121-134, Kluwer Academic Publishers
 Mulaik, S.A. (1976). *Psychometrika*, **41**, 249-262.
 Unkel, S., & Trendafilov, N.T. (2010) *Inter. Statist. Rev.*, **78**, 363-382.