様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科研費
KAKENHI

科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ４ 月 １５ 日現在

機関番号：１４４０１
研究種目：基盤研究(C)
研究期間：2011～2013
課題番号：２３５００３４７
研究課題名（和文）因子分析の新たな母数モデルに基づく新展開

研究課題名（英文）New Developments in Factor Analysis Underlain by Fixed Models

研究代表者
足立　浩平（Adachi, Kohei）

大阪大学・その他の研究科・教授

研究者番号：６０２９９０５５

交付決定額（研究期間全体）：（直接経費）　3,900,000円、（間接経費）　1,170,000円

研究成果の概要（和文）：　新たな母数モデル「データ＝共通因子×負荷＋独自因子×独自分散の平方根＋誤差」に基づく同時因子分析を考案して，最小二乗法および重みつき最小二乗法のアルゴリズムを開発して，シミュレーションによる挙動の確認・実データ解析による有用性の例証と，高階数近似とみなせる数学的性質の考究を行った．開発したアルゴリズムの特徴は，データ・フィッテングの形をとりながらも，負荷行列と独自分散の推定には，データ行列がなくても標本共分散さえ与えられれば十分である点にある．以上に加えて，共通・独自因子得点の不確定性のあり方の研究と，それの推定値の算定法の提案も行った．

研究成果の概要（英文）：I developed a factor analysis (FA) procedure underlain by a new fixed factor model in which common and unique factor scores are treated as fixed parameters. This procedure can be called joint FA (JFA), as all parameters are estimated jointly. I presented an algorithm for minimizing the least squares function of JFA. A feature of the algorithm is that the optimal loadings and unique variances can be estimated only if inter-variable covariances are available. Simulation studies showed that the true loadings and unique variances are recovered well, and the usefulness of JFA was illustrated in real data examples. A property of JFA interesting theoretically is that JFA can be regarded as higher rank approximation of a data matrix. I further used the property to give the explanation of the common and unique factor indeterminacy which is a generalized version of Mulaik's (1976) cone.

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：行動計量分析　多変量解析　因子分析　母数モデル　同時因子分析　高階数近似

１．研究開始当初の背景

変数の平均が 0 の n 個体×p 変数のデータ行列 $\mathbf{X}$ に対する m 因子の探索的因子分析モデルは，$\mathbf{X} = \mathbf{FA}' + \mathbf{E}$ のように行列表記できる．ここで，$\mathbf{F}(n \times m)$ は平均 0 の共通因子，$\mathbf{A}(p \times m)$ は負荷行列，$\mathbf{E}(n \times p)$ は誤差行列であり，$\mathbf{E}$ の各行は平均 0 の確率変数 $\mathbf{e}'$ が占め，次の 2 条件を満たす．[1] 共通因子からの独立性: $\mathbf{F}$ の行と $\mathbf{e}'$ の無相関である; [2] 独自因子の変数間の独立性: $\mathbf{e}'$ の共分散行列 Ψ^2 は非負値対角行列である.

以上のモデルは，$\mathbf{F}$ の行を変数と見なす変量モデルと母数と見なす母数モデルに大別される．因子分析は一般性の高い多変量解析法であるが，$\mathbf{F}$, $\mathbf{A}$, Ψ^2 の同時推定ができないという問題がある．この事から，現在は変量モデルが主流を占め, $\mathbf{A}$ と Ψ^2 の推定後に $\mathbf{F}$ を求める二段階推定法がとられる．しかし，母数モデルの新展開には，次に記す十分な意義を認めうる.

1904 年のスピアマンの草案から 100 年以上を経た現在，幅広く利用されながらも，未解決問題が行動計量学者・数理統計学者を惹きつける因子分析において，新たな母数モデルに基づく因子分析は，既存法が出来なかった同時母数推定を可能にして，負の独自性の不適解回避の長所も併せ持つ．さらに，「モデルがデータより多くを語る」と表せるかもしれない高階数近似を，今後の多変量解析発展の一方向とさせる可能性を持つ.

２．研究の目的

因子分析のモデルは $\mathbf{X} = \mathbf{FA}' + \mathbf{E}$ と書けるが，誤差行列 $\mathbf{E}$ に制約を課すことが特徴である．従って，$\mathbf{X}$ が標本であれば，$\mathbf{X}$ と $\mathbf{FA}' + \mathbf{E}$ は等号では結べず, 近似記号で結んで $\mathbf{X} \cong \mathbf{FA}' + \mathbf{E}$ と書かざるを得ない．この式の $\mathbf{E}$ を $\mathbf{E} = \mathbf{U}\Psi$ と再表現したモデル

$$\mathbf{X} \cong \mathbf{FA}' + \mathbf{U}\Psi \qquad (1)$$

を因子分析の基礎にすえる．ただし，$\mathbf{F}$ は n 個体×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 は独自分散の平方根を含む対角行列であり，

$$n^{-1}\mathbf{F}'\mathbf{F} = \mathbf{I}_m (m \text{ 次の単位行列}), \qquad (2)$$

$$n^{-1}\mathbf{U}'\mathbf{U} = \mathbf{I}_p, \qquad (3)$$

$$\mathbf{F}'\mathbf{U} = \mathbf{O}(\text{零行列}) \qquad (4)$$

を満たす．ここで，$\mathbf{F}$, $\mathbf{A}$, $\mathbf{U}$, Ψ の全てを母数と見なすのが，既存の母数モデルとは異なる点である．このモデルに基づく因子分析を，全てのモデル・パラメータの同時推定を目指す点で，同時因子分析(JFA)と名づけ，その**アルゴリズム**の開発・**数理的性質**の考究・**既存の因子分析との関係**の明確化を行う．JFA に期待できることに，既存の因子分析では出来なかった因子得点と他の母数の**同時推定**・**不適解の回避**・データの低階数近似である主成分分析に対して因子分析を**高階数近似**と位置づける**新たな多変量解析体系化**の提唱がある.

３．研究の方法

まず，(1)の近似の悪さを数量化する目的関数を定義して，目的関数を最小にするためのアルゴリズムを構成する．そして，アルゴリズムを実行するためのコンピュータ・プログラムを作成して，パラメータの真値からモデルに従って生成された人工データに適用して，その解がパラメータの真値を再現しているかどうかを検証する．この過程で，アルゴリズムをより効率的なものにするための改良も試みる．さらに，アルゴリズムを実データに適用して，その有用性を例証する.

目的関数の値を最小にする解を得られたとしても，解は一意でない可能性がある．特に因子分析の場合には，因子得点の解は一意でない，つまり，不確定性があることが知られる．そこで，複数ある因子得点の解の中から一組の解を選択する方法，あるいは, 一意に定まって解をよく近似する推定値の算定法を考える.

モデル，および，そのパラメータの解の数理的性質を考究する．特に，本研究では，線形代数の視点から，因子分析がデータ行列の高階数近似と見なせる可能性を追求する.

４．研究成果

モデル(1)から重みつき最小二乗基準

$$f(\mathbf{Z},\mathbf{B}) = \|\mathbf{X} - \mathbf{FA}' - \mathbf{U}\Psi\|_{\mathbf{W}}^2 = \|\mathbf{X} - \mathbf{ZB}'\|_{\mathbf{W}}^2 \qquad (5)$$

が導かれる．ここで，$\mathbf{Z} = [\mathbf{F}, \mathbf{U}]$, $\mathbf{B} = [\mathbf{A}, \Psi]$ であり，$\|\mathbf{M}\|_{\mathbf{W}}^2 = \mathrm{tr}\mathbf{MWM}'$ である．制約条件(2), (3), (4)は,

$$n^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{I}_{m+p} \qquad (6)$$

にまとめられる. 所与の $\mathbf{W}$ について, 条件(6)のもとに(5)を最小にする問題として同時因子分析(JFA)を定式化し，JFA のアルゴリズムを開発した．その長所は，データ行列 $\mathbf{X}$ がなくても，共分散行列 $\mathbf{S} = n^{-1}\mathbf{X}'\mathbf{X}$ が与えられれば，$\mathbf{A}$ と Ψ の最適解が得られる点にある．この $\mathbf{S}$ の十分性を，$\mathbf{W} = \mathbf{I}$ として次の段落に記す.

まず, 所与の $\mathbf{B}$ のもとで(5)を最小化するために，$\mathbf{W} = \mathbf{I}$ とした(5)を

$$f(\mathbf{Z}\,|\,\mathbf{B}) = (\mathbf{Z} \text{ に無関係な項}) - 2n\mathrm{tr}(n^{-1}\mathbf{X}'\mathbf{Z})\mathbf{B}' \qquad (7)$$

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

$$\mathbf{Z} = n^{1/2}\mathbf{KL}' + n^{1/2}\mathbf{K}_\perp\mathbf{L}_\perp' \qquad (8)$$

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, (7)と(8)を見比べると，(5)の最小化は, $\mathbf{Z}$ を求めなくて,

$$n^{-1}\mathbf{X}'\mathbf{Z} = n^{-1}\mathbf{B}'^{+}\mathbf{B}'\mathbf{X}'\mathbf{Z} = \mathbf{B}'^{+}\mathbf{LDL}' \qquad (9)$$

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(5)の最小化するため，(5)を

$$f(\mathbf{B}\,|\,\mathbf{Z}) = n\|n^{-1}\mathbf{X}'\mathbf{Z} - \mathbf{B}\|^2 + (\mathbf{B} \text{ に無関係な項}) \qquad (10)$$

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

$$\mathbf{A}=\mathbf{B}'^{+}\mathbf{LDL}'\mathbf{H}_m,\ \Psi = n^{-1}\mathrm{diag}(\mathbf{B}'^{+}\mathbf{LDL}'\mathbf{H}^p) \quad (11)$$

ここで, $\mathbf{H}_m$= [$\mathbf{I}_m$, ${}_m\mathbf{O}_p$]′, $\mathbf{H}^p$ = [${}_p\mathbf{O}_m$, $\mathbf{I}_p$]′である．

以上より，(9)と(11)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，(9)と(11)の更新に必要な $\mathbf{LDL}'$は，固有値分解 $\mathbf{S}$= $\mathbf{LD}^2\mathbf{L}'$で求められるため，$\mathbf{S}$ の十分性がわかる．なお，(8)は，因子得点の解は一意でないが，既存の因子分析ではできなかった負荷行列・独自分散とともに因子得点を同時推定できることを示している.

以上に記した JFA のアルゴリズムを開発して，そのプログラムを作成した上で，JFA によって十分な精度で，真のパラメータの値を推定できることを，シミュレーションによって確認した．さらに，既存の最尤 FA と同等の解を与えることを，シミュレーションと実データ解析によって確認した.

JFA の定式化から，因子分析を高階数近似と見なせることを，次のようにして確認した. 最小二乗基準(5)は,

$$f(\mathbf{Z}\,|\,\mathbf{B}) = \|\mathbf{Z} - \mathbf{XB}\|^2 + (\mathbf{Z}\text{ に無関係な項})\ . \quad (12)$$

と書き換えられ，これは，階数 p の $\mathbf{XB}$ を，(6)より階数 $m + p$ である因子得点行列 $\mathbf{Z}$ 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.

解(8)は，因子得点 $\mathbf{Z}$ の不確定性を示す．ここで，(8)右辺のユニークに定まる項 $\mathbf{Z}_1 = n^{1/2}\mathbf{KL}'$は，$\mathbf{Z}_1 = \mathbf{XBL}\Theta^{-1}\mathbf{L}'$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,

$$\|\mathbf{Z} - \mathbf{Z}_1\|^2 = \|n^{1/2}\mathbf{K}_\perp\mathbf{L}_\perp'\|^2 = n\mathrm{tr}\,\mathbf{L}_\perp'\mathbf{L}_\perp = nm \quad (13)$$

のように一定である．従って，行列をベクトルで表すと，$\mathbf{Z}$ 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．これは，共通因の不確定性を表す Mulaik(1976)の円錐を一般化したもの見なせる．(13)より，$\mathbf{Z}_1$ は，全ての解の中心をなし, $\mathbf{Z}_1$ を「中心推定値」として，因子得点の推定値とすることを提案した．また, 所与の行列に近い $\mathbf{Z}$ を求めるプロクラステス因子同定法も考案した.

(5)の最小化の特徴は，Ψを推定するため，Ψ^2 の対角要素である独自分散は非負で, 独自分散がマイナスとなる不適解が得られないことである．このことに関連する成果として，EM アルゴリズムによる因子分析では，独自分散が 0 をとることもなく，常に正でありえることを数学的に証明した．さらに，既存の因子分析の EM アルゴリズムの更新式を改良して，より安定した挙動を示す更新式を導いた.

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 7 件）

(1) Adachi, K. Three-Way Tucker2 Component Analysis Solutions of Stimuli x Responses x Individuals Data with Simple Structure and the Fewest Core Differences. *Psychometrika*, Vol. 76, No. 2, 2011, 285-305.
DOI: 10.1007/s11336-011-9208-6

(2) 足立浩平．　最小二乗置換によるサイズ固定クラスタリング．データ分析の理論と応用, 第 1 巻第 1 号, 11-22，2011 年.

(3) Adachi, K. Some Contributions to Data-fitting Factor Analysis with Empirical Comprisons to Covariance-fitting Factor Analysis. *Journal of Japanese Society of Computational Statistics*, Vol. 25, No. 1, 25-38, 2012 年.
DOI: 10.5183/jjscs.1106001_197

(4) Adachi, K. Factor Analysis with EM Algorithm Never Gives Improper Solutions when Sample Covariance and Initial Parameter Matrices Are Proper. *Psychometrika*, Vol. 38, Issue 2, 380-394, 2013 年.
DOI: 10.1007/S11336-012-9299-8

(5) Adachi, K. Generalized joint Procrustes analysis. *Computational Statistics*, Volume 28, Issue 6, 2449-2464，2013 年.
DOI: 10.1007/s00180-013-0413-x

(6) Satomura, H., & Adachi, K. Oblique Rotation in Canonical Correlation Analysis Reformulated as Maximizing the Generalized Coefficient of Determination. *Psychometrika*, Vol. 78, No. 3, 526-537, 2013 年.
DOI: 10.1007/s11336-012-9310-4

(7) Adachi, K. A restrained condition number least squares technique with its applications to avoiding rank deficiency. *Journal of Japanese Society of Computational Statistics*, Vol. 26, 39-51, 2013 年.
DOI: 10.5183/jjscs.1208002_201

〔学会発表〕（計 26 件）

(1) Adachi, K. & Murakami. K　Three Kinds of Hierarchical Relations among PCA, Nonmetric PCA, and Multiple Correspondence Analysis. International Meeting of Psychometric Society, 2011, 2011 年 7 月 19 日, 香港, The Hong Kong Institute of Education.

(2) Adachi, K. Nonsingular Transformation of Tucker2 Solutions for Representing Stimulus-Response Relationships by Sparse Networks. International Meeting of Psychometric Society, 2011，2011 年 7 月 21 日，香港, The Hong Kong Institute of Education.

(3) Adachi, K. A Fixed Factor Analysis Procedure as an Extension of Principal Component Analysis. ISI (International Statistical Institute) 2011, 2011 年 8 月 22 日, ダブリン, Convention Centre Dublin.

(4) 足立浩平 & Trendafilov, N. T. Sparsimax: A Sparse omponent Analysis Procedure. 日本行動計量学会第 39 会大会，2011 年 9 月 13 日，岡山，岡山理科大学.

(5) Adachi, K. No Improper Solution Occurs in EM Factor Analysis. Joint Meeting of the Korea-Japan Conference of Computational Statistics and the 25th Symposium of Japanese Society of Computational Statistics, 2011 年 11 月 12 日，釜山, Haeunda Grand Hotel.

(6) Adachi, K. & Trendafilov, N. T. Sparse Component and Factor Analysis Procedures via Directly Constraining the Sparseness of Loadings. Symposium on Life Sciencces and Statistics，2011 年 11 月 3 日，豊中，大阪大学.

(7) Adachi, K., & Trendafilov, N. T. SPARSIMAX: Principal Component Analysis with Direct Sparsity Constraint on Loadings. Joint Meeting of the 2011 Taipei International Statistical Symposium and 7th Conference of the Asian Regional Section of the IASC, 2011 年 12 月 18 日，台北, Academia Sinica.

(8) Adachi, K. Least Squares Permutation and Its Applications to Factor Rotation and Fixed Size Clustering. The 4th Japanese- German Symposium on Classification, 2012 年 3 月 9 日，京都，同志社大学.

(9) 足立浩平. A Theorem Regarding the Covariances between Variables and Factors in Data-Fitting Factor Analysis. 日本計算機統計学会第 26 回大会，2012 年 5 月 13 日，香川，香川県社会福祉総合センター.

(10) Adachi, K. Murakami, T., Kiers, H. A. L., & ten Berge, J. M. F. Hierarchical Relationships among Multiple Corresp- ondence, Nonmetric Component, and Principal Component Analyses in Least Squares Formulations. COMPSTAT2012, 2012 年 8 月 30 日. キプロス・リマソール，Amathus Beach Hotel.

(11) 足立浩平. 主成分分析の行列的基礎と非計量・三相配列・因子分析への発展. 2012 年度統計関連学会連合大会チュートリアルセッション，2012 年 9 月 9 日，札幌，かでる 2・7

(12) 足立浩平. 因子分析の EM アルゴリズムは不適解を与えない ― その証明と修正アルゴリズム ―. 2012 年度統計関連学会連合大会，2012 年 9 月 10 日，札幌，北海道大学.

(13) 足立浩平. 主成分分析 versus 因子分析: 低階数近似と高階数近似. 日本行動計量学会第 40 回大会，2012 年 9 月 16 日，新潟，新潟県立大学.

(14) Adachi, K. & Trendafilov, N. T. Factor analysis subject to direct sparseness constraint on loadings. ERCIM 2012，2012 年 12 月 1 日，スペイン・オビエド，Hotel Vetusta.

(15) Makino, M., & Adachi, K. A matrix decomposition approach to nonmetric factor analysis of multidimensionally quantified categorical data. ERCIM 2012，2012 年 12 月 3 日，スペイン・オビエド，Hotel Vetusta.

(16) 足立浩平. Generalized canonical correlation analysis reformulated as reduced rank approximation. 日本計算機統計学会第 27 回大会，2013 年 5 月

(17) Adachi, K. & Trendafilov, N. T. Sparse Orthogonal Factor Analysis. SIS2013 Statistical Conference，2013 年 6 月 19 日，イタリア・ブレシア，University of Brescia.

(18) Trendafilov, N. T. & Adachi, K. Sparse vs. Simple Structure Loadings. SIS2013 Statistical Conference，2013 年 6 月 19 日，イタリア・ブレシア，University of Brescia.

(19) Adachi, K. & Trendafilov, N. T. Computational Identification of Optimal Confirmatory Factor Analysis Model Using Sparseness Constraint. International Meeting Psychometric Society 2013, 2013 年 7 月 25 日，オランダ・アルンヘム，Musis Sacrum.

(20) Adachi, K. & Trendafilov, N. T. Sparse factor analysis for identifying optimal perfect simple structure. Joint Meeting of the IASC Satellite Conference and the 8th Asian Regional Section of the IASC，2013 年 8 月 23 日，ソウル，Yonsei University.

(21) Adachi, K. & Trendafilov, N. T. Perfect Simple Structure Factor Analysis with Alternating Least Squares Algorithm. 日本行動計量学会第 41 会大会，22013 年 9 月 4 日，船橋，東邦大学.

(22) Adachi, K. & Trendafilov, N. T. Decomposed Principal Component Analysis for Obtaining Interpretable Solutions. 2013 年度統

計関連学会連合大会，2013 年 9 月 11 日，豊中，大阪大学.

(23) Adachi, K. & Trendafilov, N. T. A Direct Cardinality Constraint Approach to Sparse Principal Components. 日本計算機統計学会第 27 回７シンポジウム，2013 年 11 月 16 日，熊本，熊本市民会館.

(24) Adachi, K. & Trendafilov, N. T. A New Sparse PCA procedure via Decomposing PCA Loss Function. CFE-ERCIM 2013, 2013 年 12 月 16 日，ロンドン，University of London.

(25) 足立浩平. 心理データ解析のための行列分解. 数学協働プログラム「人間行動への数理の応用による課題解決」研究集会，2014 年 2 月 15 日，東京，帝京大学・霞ヶ関キャンパス.

(26) 足立浩平. 行列分解型の因子分析と主成分分析: 高階数近似と低階数近似. 第８回日本統計学会春季集会, 2014 年 3 月 8 日, 京都, 同志社大学.

〔図書〕（計 3 件）

(1) 足立浩平・村上 隆. 非計量多変量解析法ー主成分分析から多重対応分析へー，朝倉書店，2011 年, 171 ページ.

(2) 足立浩平. 統計応用の百科事典 (松原望・美添泰人・岩崎 学・金 明哲・竹村和久・林 文・山岡和枝(編)),「非計量多次元尺度法・計量多次元尺度法」の項目. 丸善出版, 2011 年, 4 ページ.

(3) 足立浩平. 最新 心理学事典 (内田伸子・繁桝算男・杉山憲司((編)) 「多変量解析・回帰分析・主成分分析・クラスター分析・多次元尺度法」の項目，平凡社，2013 年, 9 ページ

〔産業財産権〕
○出願状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
足立 浩平（ADACHI KOHEI）

研究者番号：60299055

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：