

令和 4 年 5 月 5 日現在

機関番号：14401

研究種目：基盤研究(C) (一般)

研究期間：2018～2021

課題番号：18K11191

研究課題名(和文) データ近似定式化による因子分析の解の性質の解明

研究課題名(英文) Studies on Properties of the Data-Fitting Solution in Factor Analysis

研究代表者

足立 浩平 (Adachi, Kohei)

大阪大学・人間科学研究科・教授

研究者番号：60299055

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：因子分析とは、多数の変数の原因となる少数の共通因子を抽出する統計解析法である。ここで、変数と共通因子の例として、テスト問題に対する回答の得点と知的特性、および、行動パターンと性格特性が掲げられる。この因子分析の行列代数に基づく計算法に焦点をあて、次に列挙する解の性質を証明にした。[1] 共通因子とそれとは別の独自因子に係数を乗じた部分の解は、一通りしか存在しないことを示した。[2] 解への条件を付加すると、共通因子・独自因子・誤差が完全に分利した解が得られることを示した。[3] 因子分析に類似する統計解析法に主成分分析があるが、それと因子分析の解の違いを明示する不等式を導出した。

研究成果の学術的意義や社会的意義

多変量解析と総称される統計解析法の中でも、因子分析は、創案から100年以上の歴史を持ち、かつ、現在普及する統計ソフトウェアに常備されるポピュラーな手法であるが、その解の性質の細部が明確でなかったが、本研究で明らかになった。

因子分析は、心理学をはじめとした人文・社会科学から、自然科学に渡って広く利用される統計解析法であるため、本研究成果は、因子分析を利用する人文社会・自然科学の研究者、さらには、企業のデータサイエンティストに、因子分析の結果を解釈するための指針を与えることができる。

研究成果の概要(英文)：Factor analysis is a statistical procedure for extracting a few common factors that cause a number of variables. For example, the variables and common factors are the scores for test items and intelligence properties; they are also exemplified by behavioral patterns and personality properties. I focused on the matrix-algebraic solution of the factor analysis to elucidate the following properties of the solution. [1] The solution for the multiplication of the common factors and the other (unique) factors multiplied by coefficients can be uniquely determined. [2] Imposing an additional condition into the solution allows us to obtain the solution, in which the common factors, unique factors, and errors are completely decomposed. [3] The inequalities exist that elucidate the differences between factor analysis and principal component analysis, with the latter similar to the former.

研究分野：統計科学

キーワード：統計学 多変量解析 因子分析 行列代数

1. 研究開始当初の背景

列平均が 0 である n 個体 $\times p$ 変数のデータ行列を $X = (x_{ij})$ と表そう。因子分析(FA)を創始した Spearman(1904)の着想に基づいて、FA のモデルを行列で表すと、次のように書ける。

$$\begin{matrix} X & F & \Lambda' & U & \Psi & E \\ \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} & = & \begin{bmatrix} f_{11} & \cdots & f_{1m} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nm} \end{bmatrix} & + & \begin{bmatrix} \lambda_{11} & \lambda_{12} & \lambda_{1p} \\ \vdots & \vdots & \vdots \\ \lambda_{m1} & \lambda_{jm} & \lambda_{pm} \end{bmatrix} & + & \begin{bmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & & \vdots \\ u_{n1} & \cdots & u_{np} \end{bmatrix} & + & \begin{bmatrix} \psi_1 & & 0 \\ & \ddots & \\ 0 & & \psi_p \end{bmatrix} & + & \begin{bmatrix} e_{11} & \cdots & e_{1p} \\ \vdots & & \vdots \\ e_{n1} & \cdots & e_{np} \end{bmatrix} . \quad (1)
 \end{matrix}$$

ここで、 $F(n \times m)$ は共通因子得点、 $U(n \times p)$ は独自因子得点の行列であり、列平均は 0、

$$F'F = nI_m, \quad U'U = nI_p, \quad F'U = O \quad (\text{ただし、} I_m \text{ は } m \text{ 次の単位行列、} O \text{ は零行列}) \quad (2)$$

と制約される。また、 E は非系統的な誤差の行列、 $\Lambda(p \times m)$ は因子負荷量の行列、 Ψ は $p \times p$ の対角行列である。(1)で、 X より F の横幅が狭いのは、 p 種の変数がより少数の m 個の共通因子によって説明され、これとは対照的に、 U の横幅が X と同じで、 Ψ が対角であることは、 U の各列の独自因子が X の対応する列(変数)に独自に作用することを表し、 Ψ の対角要素の 2 乗を独自分散と呼ぶ。条件(2)のもとに、モデル(1)の誤差行列 E の Frobenius ノルムの平方

$$\|E\|^2 = \text{tr}E'E = \|X - (FA' + U\Psi)\|^2 \quad (3)$$

を最小化する F, Λ, U, Ψ を求める解法が開発されたのは、近年である(Adachi, 2012; Sočan, 2003; Unkel & Trendafilov, 2010)。

普及法は、 F, U の要素を一定の分布に従う確率変数と見なすと、(1)のモデル部 $FA' + U\Psi$ の列間共分散行列が $\Lambda\Lambda' + \Psi^2$ と書けることから、これと標本共分散行列 $S_{XX} = n^{-1}X'X$ の乖離度

$$\|S_{XX} - (\Lambda\Lambda' + \Psi^2)\|^2, \quad \text{または、その負の対数尤度バージョン} \quad (4)$$

を最小にする Λ と Ψ^2 を求めるものである。なお、(3)の最小化と(4)の最小化がほぼ同じ Λ と Ψ の解を与えることは、数値的に確認されている(足立, 2015; Stegeman, 2016)。

多くの多変量解析法が、(1)を含め「データ=モデル部+誤差」のようにデータ近似として定式化されるが、FA の普及法は、因子得点 F, U が消去された(4)の最小化、すなわち、「変数 \times 変数の共分散近似(広義に言えばモーメント近似)」であり、 F, U の最適解に基づく議論ができない。

2. 研究の目的

(1), (3)に基づくデータ近似 FA の解の性質を研究することが応募課題となる。データ近似 FA では、 F と U の最適解を考慮でき、普及法の枠組内では見出せない FA の正体を明確化させ得ると考えられる。さらに、(3)の最小化は、 F, Λ, U, Ψ が全て固定された(未知)行列である点で、行列分解としての FA とも呼べ、解の性質が行列代数に基づいて見出される。数学の中でも行列代数が分析手法の位置づけの明確化に極めて役立つ事は(Seber, 2008)、本研究が有益な諸定理を提供することを確約する。

3. 研究の方法

モデル部・残差の一意性・高階数近似としての因子得点の不定性・変数 \times 共通・独自因子の共分散と因子分析の最適性・より制約の強いデータ近似 FA との比較・主成分分

析(PCA)の解との比較を，定理の証明・シミュレーション研究・実データへの適用などによって行う．

4．研究成果

2018年度の研究成果は，次の[1], [2]のように要約される．

[1] データ近似としてFAのモデルは，(1)つまり「データ = 共通因子得点 x 因子負荷量 + 独自因子得点 x 独自標準偏差 + 誤差」のように表されるが，「モデル部 = 共通因子得点 x 因子負荷量 + 独自因子得点 x 独自標準偏差」が各個体の各変数について同定できる，つまり，算出可能であることを証明し，定理として提示した．この事実は，データからモデル部を減じて得られる誤差の値，つまり，残差が，個体 x 変数ごとに具体的数値として求められることを示し，各観測値について，FAモデルからの逸脱を評価できることを意味する．この効用を実データを用いて例証した．

[2] 共通因子得点と独自因子得点の最適解は，一意には定まらないが，最適解が満たすべき数式を明示的に示した．この数式は，幾何学的には，共通因子得点と独自因子得点を要素とするベクトルが，データ行列の線形結合をベクトル化したものを中心とした円錐を構成することを示す．

2019年度の研究成果は，データ近似定式化に基づいて，100年近くの期間，興味を惹きながら，明確な答えがでなかったFAと主成分分析(PCA)の解の相違を明示する4つの不等式を導いたことである．第1の不等式は，主成分負荷量の平方和が因子負荷量の平方和を下回らないことを示し，これは，個々の主成分負荷量の絶対値が因子負荷量の絶対値より大きくなる傾向を示唆する．第2の不等式は，FAの残差平方和がPCAの残差平方和を上回らないこと，つまり，前者の方がデータへよく(または同等に)適合することを示す．第3の不等式は，FAの独自分散の総和が，PCAの残差分散の総和からFAの残差分散の総和を引いた値を下回らない事を示し，これは，FAの残差分散が十分に小さいときに，各変数の独自分散がPCAの残差分散より大きくなる傾向を示唆する．第4の不等式は，主成分負荷量の平方和が，因子負荷量の平方和と独自分散の和を上回らないことを示し，第1の不等式とは対照的に，前者の下限を示す．以上のFAとPCAの解の関係は，FAのデータ近似定式化に基づくが，他の定式化に基づくFAとPCAの解にも，経験的に一般化できることを，シミュレーションによって確認した．

2020年度の研究成果は，次の[1]~[3]のように要約される．

[1] データ近似FAの解から得られる変数 x 独自因子得点の共分散行列の非対角要素が0でない事，つまり，各変数がそれに対応しない独自因子と相関する事実を確認した．これは，「独自因子は対応する変数に排他的に作用する」というFAの理想から逸脱する．

[2] 上記の逸脱を容認するRationalityとして，変数と独自因子得点の解の共分散行列の非対角要素変数の平方和が，データ近似FAで最小化される二乗誤差に比例することを証明した．

[3] データ近似FAに，上記のFAの理想からの逸脱を起こさないように解を制約する条件を組み込んだStegeman(2016)の方法と，この制約のない[1], [2]のデータ近似FAの解を数値的に比較した結果，後者の解が現在普及するFAの解と酷似するのに対して，これら2つの解Stegeman(2016)の制約つきFAの解がやや異なる事を確認した．

2021年度に行った研究の成果は，次の[1]~[3]のように要約される．

[1] データ近似FAの解から得られる共通因子得点と誤差が無相関であること，および，独自因子得点とそれに対応する変数の誤差は無相関であるが，独自因子得点とそれに対応しない変数の誤差は相関することを証明した．

- [2] データ近似FAに制約する条件を組み込んだStegeman(2016)の方法の解では，共通因子得点と誤差が無相関であり，かつ，独自因子得点と変数の誤差も無相関であることを証明した．
- [3] 上記の結果より，Stegeman(2016)の方法は，共通因子得点・独自因子得点・誤差の3パートが互いに無相関であることを制約条件としたFAと再定式化できるのに対して，これらの制約を緩めた方法として，データ近似FAを再定式化できることを示した．

<引用文献>

- Adachi, K. (2012). Some contributions to data-fitting factor analysis with empirical comparisons to covariance-fitting factor analysis. *Journal of the Japanese Society of Computational Statistics*, 25, 25-38.
- 足立浩平 (2015). 因子分析への行列集約アプローチ. *日本統計学会誌*, 44, 363-382.
- Seber, G. A. F. (2008). *A matrix handbook for statisticians*. Hoboken, New Jersey: Wiley.
- Sočan, G. (2003). *The incremental value of minimum rank factor analysis*. PhD Thesis, University of Groningen: Groningen.
- Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Stegeman, A. (2016). A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts. *Computational Statistics and Data Analysis*, 99, 189-203.
- Unkel, S., & Trendafilov, N. T. (2010). Simultaneous parameter estimation in exploratory factor analysis: An expository review. *International Statistical Review*, 78, 363-382.

5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 10件 / うち国際共著 5件 / うちオープンアクセス 2件）

1. 著者名 Naoto Yamashita & Kohei Adachi	4. 巻 55
2. 論文標題 Permutimin: Factor Rotation to Simple Structure with Permutation of Variables	5. 発行年 2020年
3. 雑誌名 Multivariate Behavioral Research	6. 最初と最後の頁 17-29
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/00273171.2019.1598331	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 足立浩平, 伊藤真道, 宇野光平	4. 巻 32
2. 論文標題 行列分解に基づく因子分析の新展開	5. 発行年 2020年
3. 雑誌名 計算機統計学	6. 最初と最後の頁 61-77
掲載論文のDOI (デジタルオブジェクト識別子) 10.20551/jscswabun.32.1_61	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Naoto Yamashita & Kohei Adachi	4. 巻 37
2. 論文標題 A Modified k-Means Clustering Procedure for Obtaining a Cardinality-Constrained Centroid Matrix	5. 発行年 2020年
3. 雑誌名 Journal of Classification	6. 最初と最後の頁 509-525
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s00357-019-09324-6	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Kohei Adachi	4. 巻 10
2. 論文標題 Factor Analysis: Latent Variable, Matrix Decomposition, and Constrained Uniqueness Formulations	5. 発行年 2019年
3. 雑誌名 WIREs Computational Statistics	6. 最初と最後の頁 1-17
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/wics.1458	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kohei Adachi & Nickolay T. Trendafilov	4. 巻 31
2. 論文標題 Some inequalities contrasting principal component and factor analyses solutions	5. 発行年 2019年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 31-47
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-018-0024-4	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Kohei Uno, Kohei Adachi, & Nickolay T. Trendafilov	4. 巻 84
2. 論文標題 Clustered Common Factor Exploration in Factor Analysis	5. 発行年 2019年
3. 雑誌名 Psychometrika	6. 最初と最後の頁 1048-1067
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11336-019-09666-5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Kohei Adachi & Nickolay T. Trendafilov	4. 巻 83
2. 論文標題 Some Mathematical Properties of the Matrix Decomposition Solution in Factor Analysis	5. 発行年 2018年
3. 雑誌名 Psychometrika	6. 最初と最後の頁 407-424
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11336-017-9600-y	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kohei Adachi & Nickolay T. Trendafilov	4. 巻 12
2. 論文標題 Sparsest factor analysis for clustering variables: A matrix decomposition approach	5. 発行年 2018年
3. 雑誌名 Advances in Data Analysis and Classification	6. 最初と最後の頁 559-585
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11634-017-0284-z	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Jingyu Cai, Henk A. L. Kiers, & Kohei Adachi	4. 巻 11
2. 論文標題 Computational Identification of Confirmatory Factor Analysis Model with Simplimax Procedures	5. 発行年 2021年
3. 雑誌名 Open Journal of Statistics	6. 最初と最後の頁 1044-1061
掲載論文のDOI (デジタルオブジェクト識別子) 10.4236/ojs.2021.116062	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Kohei Adachi	4. 巻 online first
2. 論文標題 Factor Analysis Procedures Revisited from the Comprehensive Model with Unique Factors Decomposed into Specific Factors and Errors	5. 発行年 2022年
3. 雑誌名 Psychometrika	6. 最初と最後の頁 印刷中
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11336-021-09824-8	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計24件 (うち招待講演 3件 / うち国際学会 7件)

1. 発表者名 Kohei Adachi
2. 発表標題 Principal Component Versus Factor Analyses with Their Intermediate Procedure in Matrix Decomposition Formulation
3. 学会等名 Data Science, Statistics & Visualization 2020 (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 足立浩平
2. 発表標題 Factor Analysis as an Intermediate between Reduced- and Full-Rank PCA
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 足立浩平
2. 発表標題 因子分析の基本的定式化 潜在変数および行列分解によるアプローチ
3. 学会等名 第25回情報・統計科学シンポジウム (招待講演)
4. 発表年 2020年

1. 発表者名 伊藤真道・足立浩平
2. 発表標題 独自性を完全分離したスパース共通因子分
3. 学会等名 日本計算機統計学会第32回大会
4. 発表年 2019年

1. 発表者名 Kohei Adachi
2. 発表標題 Revisiting factor analysis: Three types of its formulation
3. 学会等名 The 3rd International Conference on Econometrics and Statistics (EcoSta 2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Masamichi Ito & Kohei Adachi
2. 発表標題 Uniqueness-dissociated Factor Analysis Modified for Providing Sparse Loadings and Extended Common Factors
3. 学会等名 Data Science, Statistics & Visualisation (DSSV) 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Kohei Adachi
2. 発表標題 A Comparison of Penalized and Penalty-Free Procedures for Sparse Factor Analysis
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 山本勇氣, 足立浩平
2. 発表標題 A Threshold-Constrained Sparse Regression Procedure and its Comparison with L0 approach
3. 学会等名 日本計算機統計学会第33回シンポジウム
4. 発表年 2019年

1. 発表者名 足立浩平
2. 発表標題 因子分析の 3 種の定式化 潜在変数・行列分解・独自性制約アプローチ
3. 学会等名 2019年度日本分類学会シンポジウム講演予稿集
4. 発表年 2019年

1. 発表者名 Aya Nakashima & Kohei Adachi
2. 発表標題 Three-mode PCA for finding a solution intermediate between Tucker3 and Parafac
3. 学会等名 CMStatistics 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 Masamichi Ito & Kohei Adachi
2. 発表標題 An extended k-means clustering procedure with unique factors
3. 学会等名 CMStatistics 2019 (国際学会)
4. 発表年 2019年

1. 発表者名 山本勇氣・足立浩平
2. 発表標題 非ゼロ要素の下限を制約したスパース主成分分析
3. 学会等名 日本計算機統計学会第32回大会
4. 発表年 2018年

1. 発表者名 Jingyu Cai & Kohei Adachi
2. 発表標題 Sparse Factor Analysis Using a Majorization Algorithm for Providing Cardinality Prespecified Loadings
3. 学会等名 日本分類学会第37大会
4. 発表年 2018年

1. 発表者名 Naoto Yamashita, Kohei Adachi ,
2. 発表標題 A New Factor Analysis Procedure Constrained by Layered Simple Structure
3. 学会等名 International Meeting of Psychometric Society 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Yuki Yamamoto & Kohei Adachi
2. 発表標題 Sparse principal component analysis with the absolute values of nonzero loadings greater than a threshold
3. 学会等名 Data Science, Statistics & Visualization 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Kohei Adachi & Nickolay T. Trendafilov
2. 発表標題 Factor Analysis Unmasked by Matrix Decomposition Approach
3. 学会等名 Data Science, Statistics & Visualization 2018
4. 発表年 2018年

1. 発表者名 Jingyu Cai & Kohei Adachi
2. 発表標題 Maximum likelihood EM factor analysis of high-dimensional data with the sparsest constraint on loadings
3. 学会等名 COMPSTAT2018
4. 発表年 2018年

1. 発表者名 Kohei Adachi & Nickolay T. Trendafilov
2. 発表標題 Crucial differences between principal component and factor analyses solutions elucidated by some inequalities
3. 学会等名 COMPSTAT2018
4. 発表年 2018年

1. 発表者名 Kohei Adachi
2. 発表標題 A Generalization of FA and PCA Which Includes Them as Special “ Sparse ” Cases
3. 学会等名 Japanese Joint Statistical Meeting 2018
4. 発表年 2018年

1. 発表者名 山本勇氣・足立浩平
2. 発表標題 負荷量の下限を制約するスパース行列因子分析，
3. 学会等名 2018年度日本分類学会シンポジウム
4. 発表年 2018年

1. 発表者名 中島 文・足立浩平
2. 発表標題 Tucker3とParafacの中間モデルのスパース同定
3. 学会等名 2018年度日本分類学会シンポジウム
4. 発表年 2018年

1. 発表者名 Gaku Tsujii & Kohei Adachi
2. 発表標題 Sparse Asymmetric Multidimensional Scaling: Penalized Self-Distance Unfolding Approaches
3. 学会等名 日本計算機統計学会第32回シンポジウム
4. 発表年 2018年

1. 発表者名 足立浩平
2. 発表標題 因子負荷量とともに因子間相関の 単純構造を考慮した斜交回転法
3. 学会等名 日本分類学会第40回大会
4. 発表年 2021年

1. 発表者名 足立浩平
2. 発表標題 A Neglected Comprehensive Model of Factor Analysis Is Revived in Matrix Decomposition Approaches
3. 学会等名 日本計算機統計学会第35回シンポジウム
4. 発表年 2021年

〔図書〕 計2件

1. 著者名 Kohei Adachi	4. 発行年 2020年
2. 出版社 Springer	5. 総ページ数 457
3. 書名 Matrix-based introduction to multivariate data analysis: Second edition	

1. 著者名 足立浩平	4. 発行年 2019年
2. 出版社 勉誠出版	5. 総ページ数 15
3. 書名 文化情報学事典 (B3-1. 質的分析法; B3-10. 数量化 類)	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
英国	The Open University	Faculty of Sci, Tech, Eng & Math	School of Mathematics and Statistics
オランダ	University of Groningen	Faculty of Beh. & Soc Sciences	Psychometrics and Statistics