Aplikace matematiky

Petr Kratochvíl

The determination of factors in linear models of factor analysis

Aplikace matematiky, Vol. 35 (1990), No. 5, 350–355

Persistent URL: `http://dml.cz/dmlcz/104416`

Terms of use:

© Institute of Mathematics AS CR, 1990

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

35 (1990) APLIKACE MATEMATIKY No. 5, 350—355

THE DETERMINATION OF FACTORS IN LINEAR MODELS OF FACTOR ANALYSIS

PETR KRATOCHVÍL

(Received December 21, 1988)

Summary. The author shows that a decomposition of a covariance matrix $\Sigma = \mathbf{AA}'$ implies the corresponding model, i.e. the existence of factors f_j such that $x_i = \Sigma a_{ij} f_j$ is true. The result is applied to the general linear model of factor analysis. A procedure for computing the factor score is proposed.

Keywords: Factor score, linear model, existence of factors, singular value decomposition.

AMS Classification: 62H25.

1. INTRODUCTION

The usual models for factor analysis are based on the assumption of existence of a set of unobserved random variables. These variables are called factors and are used for description of a given set of observed variables. Sometimes the factors are considered to be nonrandom quantities that vary from one individual to another. Analysing the procedure of the application we see that the existence of factors is never verified, which causes certain problems if we have to teach factor analysis or to explain it to specialists. Really, there is no simple method that would enable us to prove apriori the existence of factors. It is peculiar to assume something which is not verified. A question arises whether the above assumption is necessary. In this paper we give a negative answer to the question. We show that the existence of factors in a linear factor model is a consequence of a decomposition of a covariance matrix. Now we describe the idea more precisely. In the sequel we suppose that *all variables have finite variances* and that they are expressed as *deviations from their means. The covariance matrix is supposed to be nontrivial*, i.e. $\Sigma \neq 0$.

Let $\mathbf{g}$ be a column vector of m components and $\mathbf{u}$ a column vector of p components such that for a suitable $p \times m$ matrix Λ of real coefficients the observable column vector $\mathbf{x}$ of p components can be written as

$$\mathbf{x} = \Lambda \mathbf{g} + \mathbf{u} . \tag{1}$$

The components of $\mathbf{g}$ are called common factors and the components of $\mathbf{u}$ are called specific factors. The specific factors are supposed to be distributed independently of $\mathbf{g}$ and the covariance matrix $\mathscr{E}\mathbf{uu}' = \mathbf{\Psi}$ is supposed to be diagonal and non-singular. Let $\mathscr{E}\mathbf{u} = \mathbf{0}$, $\mathscr{E}\mathbf{g} = \mathbf{0}$ and let $\mathscr{E}\mathbf{gg}' = \mathbf{I}$ = the identity matrix.

A well-known consequence of the linear model (1) is the following decomposition of the covariance matrix $\mathbf{\Sigma}$ of the observed $\mathbf{x}$:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}' + \mathbf{\Psi}\,. \tag{2}$$

The factor analysis has two steps. First the covariance matrix $\mathbf{\Sigma}$ is computed and then the decomposition (2) is performed. In the paper [4] the authors have shown that the decomposition (2) of the covariance matrix $\mathbf{\Sigma}$ implies the existence of factors such that the above mentioned properties are true, the components of $\mathbf{g}$ being uncorrelated and with unit variances. However, the matrix formalism was not used and the proof was awkward.

In this paper we present a simple proof of existence of factors and derive consequences of our method for the problem of factor scores. Although the factors are determined ambiguously, the score could possess better properties. The problem of factor scores has been discussed in the famous paper by Anderson and Rubin [2]. A survey of modern methods for factor score is given in the monograph [3]. See also the monograph [1].

Let a matrix $\mathbf{A}$ be partitioned into two submatrices $\mathbf{A} = (\mathbf{\Lambda}\ \mathbf{\Phi})$, where $\mathbf{\Phi}$ is positive definite, diagonal and $\mathbf{\Phi}^2 = \mathbf{\Psi}$. Then $\mathbf{AA}' = \mathbf{\Lambda\Lambda}' + \mathbf{\Psi}$, i.e.

$$\mathbf{\Sigma} = \mathbf{AA}'\,. \tag{3}$$

We see that (2) is a special case of the simple decomposition (3) which is therefore quite general. Similarly we can show that the linear model (1) may be considered as a special case of a model

$$\mathbf{x} = \mathbf{Af}\,, \tag{4}$$

where $\mathbf{f}$ is a column vector of $m + p$ components and can be partitioned into two subvectors $\mathbf{f}' = (\mathbf{g}'\quad \mathbf{u}'\mathbf{\Phi}^{-1})$.

In the following propositions *we shall consider the general expressions* (3) *and* (4) *only*. Thus the general model of component analysis and also the model of factor analysis will be treated.

2. THE EXISTENCE OF FACTORS

Lemma 1. *Let $0 < r \leqq s \leqq t$ be integer numbers. Let a $t \times r$ matrix $\mathbf{Y}$ and an $s \times r$ matrix $\mathbf{Z}$ be such that their transposes are semiorthogonal, i.e. $\mathbf{Y}'\mathbf{Y} = \mathbf{Z}'\mathbf{Z} = \mathbf{I}_r$ = the identity matrix. Then there exists* a $t \times s$ *matrix $\mathbf{W}$ such that*

$\mathbf{W}'\mathbf{W} = \mathbf{I}_s$ *and the decomposition*

$$\mathbf{Y} = \mathbf{W}\mathbf{Z} \tag{5}$$

is true.

Proof. The columns of $\mathbf{Z}$ are orthogonal s vectors. Hence we can choose $s - r$ vectors $\mathbf{z}_{r+1}, \mathbf{z}_{r+2}, \ldots, \mathbf{z}_s$ such that we get a base of the space of all s vectors. Let the chosen vectors be columns of an $s \times (s - r)$ matrix $\mathbf{Z}_1$. Define a matrix $\mathbf{G}$ that is partitioned into the submatrices $\mathbf{Z}$ and $\mathbf{Z}_1$, $\mathbf{G} := (\mathbf{Z}\ \mathbf{Z}_1)$. It is square and semiorthogonal, hence orthogonal, $\mathbf{G}'\mathbf{G} = \mathbf{G}\mathbf{G}' = \mathbf{I}_s$. Similarly we can choose $s - r$ orthogonal t vectors and joint them to the columns of $\mathbf{Y}$ in such a way that a partitioned $t \times s$ matrix $\mathbf{F} := (\mathbf{Y}\ \mathbf{Y}_1)$ has a semiorthogonal transpose. Define a $t \times s$ matrix $\mathbf{W} := \mathbf{F}\mathbf{G}'$. Then $\mathbf{W}\mathbf{G} = \mathbf{F}$. Now we have $\mathbf{W}'\mathbf{W} = \mathbf{G}\mathbf{F}'\mathbf{F}\mathbf{G}' = \mathbf{G}\mathbf{G}' = \mathbf{I}_s$. The equality between the first blocks of

$$(\mathbf{Y}\ \mathbf{Y}_1) = \mathbf{F} = \mathbf{W}\mathbf{G} = \mathbf{W}(\mathbf{Z}\ \mathbf{Z}_1) = (\mathbf{W}\mathbf{Z}\ \ \mathbf{W}\mathbf{Z}_1) \quad \text{yields (5)}.$$

Remark. Recall that the dimension of a space of random variables is defined as the maximal number of mutually uncorrelated random variables on the basic space. A finite dimension is an exceptional case and in applications we may ignore an assumption about the dimension since it is fulfilled.

In the following proof we use the notion of an orthogonal base. An orthogonal base of a subspace is a family of mutually uncorrelated random variables with unit variances and such that each random variable of the given subspace can be expressed as a linear combination. This expression is given for $\mathbf{x}$ by the formula (8) below. The subspace containing $\mathbf{x}$ and also the orthogonal base are not uniquely determined, which implies that the factors are not uniquely determined. This fact has caused the conviction that the factor analysis is a questionable theory. We hope that our results throw light on the matter.

Theorem 1. *Let* $\mathbf{x}$ *be a column vector of* p *components which are variables with mean zero and finite variances* $\mathbf{x}' = (x_1, x_2, \ldots, x_p)$, *and let*

$$\Sigma = \mathbf{A}\mathbf{A}' \tag{6}$$

be a decomposition of the covariance matrix, where $\mathbf{A}$ *is a* $p \times s$ *nonzero matrix. Let the dimension of the space of all random variables be not less than* s.

Then there exists a column vector $\mathbf{f}$ *of uncorrelated random variables with unit variances and mean zero such that the model*

$$\mathbf{x} = \mathbf{A}\mathbf{f} \tag{7}$$

is true.

Proof. Since the dimension is supposed to be large enough, we may assume that all x_i, $i = 1, 2, \ldots, p$, belong to a subspace with an orthogonal base $\mathbf{g}' =$

$= (g_1, g_2, \ldots, g_t)$, $t \geqq s$. Let $\mathbf{B}$ denote a $p \times t$ matrix of the corresponding coordinates. We have

(8) $$\mathbf{x} = \mathbf{Bg}\,.$$

Since the covariance matrix of $\mathbf{g}$ is the identity matrix, the relation (8) implies

(9) $$\Sigma = \mathbf{BB}'\,,$$

which can be easily computed. Now (6) and (9) give $\mathbf{AA}' = \mathbf{BB}'$. Therefore the first and second factors in Singular Value Decomposition of $\mathbf{A}$ and $\mathbf{B}$ coincide, i.e. $\mathbf{A} = \mathbf{VDZ}'$ and $\mathbf{B} = \mathbf{VDY}'$ where the diagonal $r \times r$ matrix $\mathbf{D}$ is positive definite and the matrices $\mathbf{V}$, $\mathbf{Y}$, $\mathbf{Z}$ possess semiorthogonal transposes. The assumptions of Lemma 1 are fulfilled, hence there exists a matrix $\mathbf{W}$ with a semiorthogonal transpose and such that (5) is true. Define a column vector $\mathbf{f} = \mathbf{W}'\mathbf{g}$. We get

$$\mathbf{x} = \mathbf{Bg} = \mathbf{VDY}'\mathbf{g} = \mathbf{VDZ}'\mathbf{W}'\mathbf{g} = \mathbf{Af}\,,$$

i.e. (7) is true.

An easy computation gives the covariance matrix $\mathbf{C_f}$ of the vector $\mathbf{f}$:

$$\mathbf{C_f} = \mathbf{W}'\mathbf{C}_g\mathbf{W} = \mathbf{W}'\mathbf{I}_t\mathbf{W} = \mathbf{W}'\mathbf{W} = \mathbf{I}_s\,,$$

hence the components of $\mathbf{f}$ are uncorrelated with unit variances and the proof is complete.

The decomposition (6) is a well-known and easy consequence of the relation (7), and as we have shown in Theorem 1, the relation (7) is a consequence of (6) if a certain general condition is fulfilled. Therefore the existence of a model (7) and the decomposition (6) can be considered as equivalent properties of $\mathbf{x}$.

Evidently, each semiorthogonal matrix $\mathbf{V}'$ is a pseudoinverse of $\mathbf{V}$. According to Theorem 1, $\mathbf{x} = \mathbf{Vh}$ possesses a solution $\mathbf{h} = \mathbf{DZ}'\mathbf{f}$, which implies the following

Corollary 1. *Let* $\mathbf{A} = \mathbf{VDZ}'$ *be a Singular Value Decomposition of* $\mathbf{A}$ *and let the suppositions of Theorem* 1 *be fulfilled. Then*

(10) $$\mathbf{VV}'\mathbf{x} = \mathbf{x}\,.$$

Corollary 2. *Denote* $\mathbf{h} = \mathbf{D}^{-1}\mathbf{V}'\mathbf{x}$ *and let the assumptions of Corollary* 1 *be true. Then*

$$\mathbf{Z}'\mathbf{f} = \mathbf{h} \quad \textit{is equivalent to} \quad \mathbf{Af} = \mathbf{x}\,.$$

Proof. The relation $\mathbf{Z}'\mathbf{f} = \mathbf{h}$ implies $\mathbf{Af} = \mathbf{VDZ}'\mathbf{f} = \mathbf{VDD}^{-1}\mathbf{V}'\mathbf{x} = \mathbf{x}$. The sufficiency of the condition is evident.

In the following propositions we analyse the determination of factors. We show that the factors $\mathbf{f}$ can be expressed as sums where the first part is uniquely determined and the second part is uncorrelated with the first and can be arbitrarily selected.

Proposition 1. *Let* $\Sigma = \mathbf{AA}'$ *be a decomposition of a covariance* $p \times p$ *matrix of a vector* $\mathbf{x}$ *and let* $\mathbf{A} = \mathbf{VDZ}'$ *be a Singular Value Decomposition of the* $p \times s$ *matrix* $\mathbf{A}$. *Denote by* $\mathbf{h}$ *the* r *vector* $\mathbf{h} = \mathbf{D}^{-1}\mathbf{V}'\mathbf{x}$, r *being rank of* Σ. *Let an arbitrary* $(s - r)$ *column vector* $\mathbf{g}_0$ *of mutually uncorrelated components with unit variances be chosen such that these components are uncorrelated with the components of* $\mathbf{h}$. *Choose an* $s \times (s - r)$ *matrix* $\mathbf{Z}_0$ *such that* $\mathbf{Z}'\mathbf{Z}_0 = 0$ *and* $\mathbf{Z}_0'\mathbf{Z}_0 = \mathbf{I}$.

Then $\mathbf{f} = \mathbf{Zh} + \mathbf{Z}_0\mathbf{g}_0$ *are factors, i.e.* $\mathbf{Af} = \mathbf{x}$ *is true.*

Proof.

(a) $\mathbf{AZ}_0\mathbf{g}_0 = \mathbf{VDZ}'\mathbf{Z}_0\mathbf{g}_0 = 0$ since $\mathbf{Z}'\mathbf{Z}_0 = 0$.

(b) $\mathbf{AZh} = (\mathbf{VDZ}')\,\mathbf{Z}(\mathbf{D}^{-1}\mathbf{V}'\mathbf{x}) = \mathbf{VV}'\mathbf{x} = \mathbf{x}$ by virtue of $\mathbf{DZ}'\mathbf{ZD}^{-1} = \mathbf{I}$ and Corollary 1. Note that the existence of uncorrelated components of $\mathbf{h}$ and $\mathbf{g}_0$ ensures the dimension condition in Theorem 1 and in Corollary 1.

Now we get $\mathbf{Af} = \mathbf{A}(\mathbf{Zh} + \mathbf{Z}_0\mathbf{g}_0) = \mathbf{AZh} + \mathbf{AZ}_0\mathbf{g}_0 = \mathbf{x}$.

Proposition 2. *Let* $\mathbf{A} = \mathbf{VDZ}'$ *be a Singular Value Decomposition of a* $p \times s$ *matrix* $\mathbf{A}$, *the diagonal* $r \times r$ *matrix* $\mathbf{D}$ *being positive definite. Let* $\mathbf{x}$ *and* $\mathbf{f}$ *be column vectors such that the model* $\mathbf{x} = \mathbf{Af}$ *is true and let the components of* $\mathbf{f}$ *be mutually uncorrelated with unit variances.*

Then there exists an $s \times (s - r)$ *matrix* $\mathbf{Z}_0$ *and an* $(s - r)$ *column vector* $\mathbf{g}_0$ *such that* $\mathbf{Z}'\mathbf{Z}_0 = 0$, $\mathbf{Z}_0'\mathbf{Z}_0 = \mathbf{I}$, *the components of* $\mathbf{g}_0$ *are mutually uncorrelated with unit variances and uncorrelated with components of the vector* $\mathbf{h} = \mathbf{D}^{-1}\mathbf{V}'\mathbf{x}$, *and* $\mathbf{f} = \mathbf{Zh} + \mathbf{Z}_0\mathbf{g}_0$.

Proof. Similarly as in the proof of Lemma 1, we can define a square orthogonal matrix $\mathbf{G}$ that is partitioned into submatrices $\mathbf{Z}$ and $\mathbf{Z}_0$, $\mathbf{G} = (\mathbf{ZZ}_0)$. The relation $\mathbf{G}'\mathbf{G} = \mathbf{I}$ implies $\mathbf{Z}'\mathbf{Z}_0 = 0$ and $\mathbf{Z}_0'\mathbf{Z}_0 = \mathbf{I}$. The relation $\mathbf{GG}' = \mathbf{I}$ implies

$$\mathbf{ZZ}' + \mathbf{Z}_0\mathbf{Z}_0' = \mathbf{I}\,. \tag{11}$$

Denote by $\mathbf{g}_0$ the $(s - r)$ vector $\mathbf{g}_0 = \mathbf{Z}_0'\mathbf{f}$. According to Corollary 2, $\mathbf{h} = \mathbf{Z}'\mathbf{f}$. Now the relation (11) implies $\mathbf{f} = \mathbf{If} = \mathbf{ZZ}'\mathbf{f} + \mathbf{Z}_0\mathbf{Z}_0'\mathbf{f} = \mathbf{Zh} + \mathbf{Z}_0\mathbf{g}_0$. An easy computation gives the covariance matrix of $\mathbf{g}_0$ which is equal to $\mathbf{Z}_0'\mathbf{Z}_0 = \mathbf{I}$. Similarly, the entries of $\mathbf{Z}'\mathbf{Z}_0 = 0$ are the correlation coefficients between the components of $\mathbf{h}$ and $\mathbf{g}_0$. The proof is complete.

3. CONCLUSIONS

As mentioned in Introduction, even the general linear model of factor analysis can be considered to be of the form (3) and (4). The existence of factors $\mathbf{f}$ and the model (4) is a consequence of the decomposition (3) as has been shown in Theorem 1. Let $\mathbf{A} = \mathbf{VDZ}'$ be a Singular Value Decomposition of $\mathbf{A}$. As we have seen in Proposition 2, each vector of factors can be written as $\mathbf{f} = \mathbf{Z}\mathbf{h} + \mathbf{Z}_0\mathbf{g}_0$ where $\mathbf{g}_0$ is uncorrelated with $\mathbf{h} = \mathbf{D}^{-1}\mathbf{V}'\mathbf{x}$, i.e. under general conditions it is also uncorrelated with $\mathbf{VDh} = \mathbf{VDD}^{-1}\mathbf{V}'\mathbf{x} = \mathbf{VV}'\mathbf{x} = \mathbf{x}$ according to Lemma 1 and Corollary 1. The natural value is $\hat{\mathbf{g}}_0 = 0$. Then the factor scores are computed from

$$\hat{\mathbf{f}} = \mathbf{Z}\mathbf{h} = \mathbf{Z}\mathbf{D}^{-1}\mathbf{V}'\mathbf{x}\,.$$

References

[1] *T. W. Anderson:* An Introduction to Multivariate Statistical Analysis, Second Edition, John Wiley & sons, 1984.

[2] *T. W Anderson, Herman Rubin:* Statistical inference in factor analysis Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability (Jerzy Neyman, ed.), Vol. V, University of California, Berkeley and Los Angeles, 1956.

[3] *P. Blahuš:* Factor analysis and its generalization. Matematický seminář SNTL, sv. 21, Praha 1985 (in Czech, russian translation to appear).

[4] *P. Kratochvíl, J. Nekola:* Modelling of the research activity with the use of factor analysis (Czech). Ekonomicko-matematický obzor 15 (1979), 295—309.

Souhrn

URČENÍ FAKTORŮ V LINEÁRNÍCH MODELECH FAKTOROVÉ ANALÝZY

PETR KRATOCHVÍL

V článku je ukázáno, že rozklad kovariační matice $\Sigma = \mathbf{AA}'$ implikuje platnost odpovídajícího modelu, tj. existenci faktorů f_j takových, že $x_i = \Sigma a_{ij} f_j$. Výsledek je užit na obecný lineární model faktorové analýzy. Je navržen postup výpočtu faktorových skóre.

Резюме

ОПРЕДЕЛЕНИЕ ФАКТОРОВ В ЛИНЕЙНЫХ МОДЕЛЯХ ФАКТОРНОГО АНАЛИЗА

PETR KRATOCHVÍL

В работе доказывается, что разложение ковариационной матрицы $\Sigma = \mathbf{AA}'$ влечет за собой соответствующую модель, это значит существование факторов f_j так, что $x_i = \Sigma\, a_{ij} f_j$ Резултат применяется к общей линейной модели факторного анализа. Предложен метод измерения факторов.

Author's address: RNDr. *Petr Kratochvíl*, CSc., Matematický ústav ČSAV, Žitná 25, 115 67 Praha 1.