

# FACTOR ANALYSIS IN XLISP-STAT

JAN DE LEEUW

ABSTRACT. This note explains algorithms and corresponding Xlisp-Stat computer programs for exploratory factor analysis. It gives brief introductions to the various factor analysis techniques.

## CONTENTS

1. Introduction	1
2. Common Factor Analysis	2
2.1. Upper Bounds for uniqueness	3
2.2. Rank p Approximation	3
2.3. Least Squares Factor Analysis	3
2.4. Maximum Likelihood Factor Analysis	4
2.5. Swain-type Loss Functions	4
3. Minimum Trace Factor Analysis	4
4. Image Analysis	4
5. Brief Notes on Nonlinear Factor Analysis	4

## 1. INTRODUCTION

Suppose  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_m$  is a sequence of  $m$  centered random variables, with  $\mathbf{V}(\underline{x}_j) < \infty$  for all  $j$ , and suppose

$$(1.1) \quad C = \{c_{j\ell}\} \triangleq \{\mathbf{E}(\underline{x}_j \underline{x}_\ell)\}$$

is their *covariance* or *dispersion matrix*.

**Definition 1.1.** *Linear Factor Analysis* is a class of techniques that studies decompositions of the form

$$(1.2) \quad \underline{x}_j = \hat{\underline{x}}_j + \tilde{\underline{x}}_j,$$

where, for each  $j$ ,

$$(1.3) \quad \mathbf{C}(\hat{\underline{x}}_j, \tilde{\underline{x}}_j) = 0.$$

*Remark 1.2.* This definition is, of course, rather vague. It means that we think of factor analysis as any technique that decomposes observed variables into two uncorrelated parts. If we remember the basic Tukey-ism, that

$$\text{Data} = \text{Structure} + \text{Error},$$

we see that this notion is very general indeed.

To give Definition 1.1 some substantial content, we have to add some additional qualifications. Usually they are of the form

$$(1.4) \quad \mathcal{P}_1\{\hat{\underline{x}}_1, \dots, \hat{\underline{x}}_m\},$$

which simply means that the structural parts  $\hat{\underline{x}}_j$  are supposed to have some (usually desirable) property  $\mathcal{P}_1$ , and

$$(1.5) \quad \mathcal{P}_2\{\tilde{\underline{x}}_1, \dots, \tilde{\underline{x}}_m\},$$

which means that the *errors* or *residuals* have some other property  $\mathcal{P}_2$ , which usually emphasizes their error-like behavior.

*Remark 1.3.* Of course property 1.3 implies that

$$(1.6) \quad \mathbf{C}(\underline{x}_j, \underline{x}_\ell) = \mathbf{C}(\hat{\underline{x}}_j, \hat{\underline{x}}_\ell) + \mathbf{C}(\tilde{\underline{x}}_j, \tilde{\underline{x}}_\ell),$$

which means that the additive decomposition 1.2 of the random variables implies an additive decomposition of their dispersions. Of course the properties 1.4 and 1.5 in general also imply certain properties of the dispersion matrices, which we will write as  $\hat{C}$  and  $\tilde{C}$  below.

## 2. COMMON FACTOR ANALYSIS

**Definition 2.1.** Common factor analysis is defined by the two properties

$$(2.1) \quad \text{rank}(\hat{\underline{x}}_1, \dots, \hat{\underline{x}}_m) = p < m,$$

$$(2.2) \quad \tilde{\underline{x}}_j \perp \tilde{\underline{x}}_\ell \text{ for all } j, \ell.$$

It is clear that Definition 2.1 implies that

$$(2.3) \quad \text{rank}(\hat{C}) = p < m,$$

$$(2.4) \quad \text{diag}(\tilde{C}) = \tilde{C}.$$

It may not be immediately obvious (and, indeed, it took Spearman 25 years to prove) that the converse is also true. If equations 2.3 are true, then equations 2.1 also have a solution (in fact, they have more than one, which is the fundamental indeterminacy of common factor analysis). We give a simple proof of these facts below.

**Definition 2.2.** Random variable  $\hat{x}_j$  is called the *common part* of  $\underline{x}_j$ , while  $\tilde{x}_j$  is called the *unique part*. The matrix  $U^2 \triangleq \text{diag}(\tilde{C})$  contains *uniquenesses*, while  $H^2 \triangleq \text{diag}(C) - U^2$  contains *communalities*. Any basis for the space spanned by the  $\hat{x}_j$  is a set of *common factors*. Coordinates of the  $\hat{x}_j$  in a space of common factors are called *factor loadings*.

*Remark 2.3.* Observe that we can also write the common or multiple factor analysis model in the form

$$(2.5) \quad C = AA' + U^2,$$

where  $A$  is an  $m \times p$  matrix of rank  $p$  of factor loadings, and  $U^2$  is the non-negative diagonal matrix of uniquenesses. In the same way the random variable form of the model can be written as

$$(2.6) \quad \underline{x} = A\underline{z} + U\underline{y},$$

where  $\underline{z}$  is an orthonormal set of common factors, and  $\underline{y}$  is an orthonormal set of unique factors, which is moreover orthogonal to  $\underline{z}$ .

**Theorem 2.4.** A vector  $\underline{x}$  of random variables can be written in the form  $\underline{x} = A\underline{z} + U\underline{y}$  if and only if the dispersion matrix of  $\underline{x}$  can be written in the form  $C = AA' + U^2$ .

*Proof.* It is trivial that 2.6 implies 2.5. We construct the general solution to 2.6 to prove the converse.  $\square$

### 2.1. Upper Bounds for uniqueness.

**Theorem 2.5.**  $U \leq [\text{diag}(C^{-1})]^{-1}$ .

*Proof.* We know that  $U \leq C$ , which means  $C^{-1} \leq U^{-1}$ , which means  $\text{diag}(C^{-1}) \leq U^{-1}$ , which shows that  $U \leq [\text{diag}(C^{-1})]^{-1}$ .  $\square$

There is an alternative proof, which actually leads to a sharper bound. Define  $\beta_j$  as the vector of regression coefficients of variable  $j$  on the others. Also  $\sigma_j^2$  is the residual sum of squares of this regression. Define  $\bar{\beta}_j$  as the vector with element  $+1$  on position  $j$ , and the elements of  $-\beta_j$  on the other places. Observe that  $\bar{\beta}_j$  is equal to  $\sigma_j^2$  times column  $j$  of  $C^{-1}$ .

**Theorem 2.6.**  $u_j^2 \leq u_j^2 + \sum_{\ell \neq j} \beta_{j\ell}^2 u_\ell^2 \leq \sigma_j^2$

*Proof.* Because  $U \leq C$ , we also have  $\bar{\beta}_j' U_j^2 \bar{\beta}_j \leq \bar{\beta}_j' C \bar{\beta}_j$ . But the right hand side is equal to  $\sigma_j^2$ .  $\square$

### 2.2. Rank p Approximation.

### 2.3. Least Squares Factor Analysis.

## 2.4. Maximum Likelihood Factor Analysis.

## 2.5. Swain-type Loss Functions.

### 3. MINIMUM TRACE FACTOR ANALYSIS

Common Factor Analysis, in the covariance version, can be interpreted as finding a decomposition  $C = \hat{C} + \tilde{C}$  such that  $\tilde{C}$  is diagonal, and  $\text{rank}(\hat{C})$  is as small as possible. This means, using a more convenient notation, that we want to compute

$$(3.1) \quad \inf_U \{\text{rank}(C - U) \mid 0 \leq U \leq C\}$$

In Equation 3.1  $U$  varies over the diagonal matrices, and the interval over which  $U$  is allowed to vary should be interpreted in the Loewner sense. Thus  $U \leq C$  means that  $C - U$  is positive semi-definite, and  $0 \leq U$  means that  $U$  is positive semidefinite (which, in this case, simply means that the diagonal elements of  $U$  are non-negative). Observe that  $U^2$  varies over a convex region.

The rank of a matrix is not a continuous function of the matrix (in fact, it is a lower semi-continuous step-function), which means the minimization problem defined by 3.1 cannot really be solved by the usual numerical methods. In order to solve it, we have investigate solvability of the systems  $\text{rank}(C - U) = p$  for  $p = 1, 2, \dots$ . This can be done quite easily for  $p = 1$ , in which we find the famous tetrad conditions. See Bekker and De Leeuw [?] for a recent review. For  $p > 1$  things are not so simple, and basically unresolved. The equation  $\text{rank}(C - U) = p$  is equivalent to the system  $\det(A) = 0$ , where  $A$  varies over all submatrices of  $C - U$  of order  $p + 1$ . We can first check if this system is solvable for  $p = 1$ , then for  $p = 2$ , and so on.

A considerable more promising approach would be to look at the problem

$$(3.2) \quad \inf_U \{\text{trace}(C - U) \mid 0 \leq U \leq C\}$$

### 4. IMAGE ANALYSIS

### 5. BRIEF NOTES ON NONLINEAR FACTOR ANALYSIS

UCLA STATISTICS PROGRAM, 8118 MATHEMATICAL SCIENCES BUILDING,  
UNIVERSITY OF CALIFORNIA AT LOS ANGELES

*E-mail address:* deleeuw@stat.ucla.edu