MAXIMUM LIKELIHOOD IN GENERALIZED FIXED SCORE FACTOR ANALYSIS

JAN DE LEEUW

ABSTRACT. We study the weighted least squares fixed rank approximation problem in which the weight matrices depend on unknown parameters. The classical example is fixed score factor analysis (FSFA), where the weights depend on the unknown uniquenesses.

1. INTRODUCTION

Factor Analysis (FA) is a class of techniques to approximate a data matrix by a matrix of the same dimension but of lower rank. In order words, we have an $n \times m$ matrix X and an integer $1 \leq p \leq \min(n,m)$, and we want to find an $n \times p$ matrix A of *factor scores* and an $m \times p$ matrix B of *factor loadings* such that $X \approx AB'$.

The obvious translation into an optimization problem with a real-valued loss function is to minimize $\|X - AB'\|$ over A and B, with $\| \bullet \|$ some norm or semi-norm. The norm is usually unitarily invariant, and of course more often than not of the weighted least squares type. A fairly general treatment is in De Leeuw [1984], also see Zha [1991].

In *Common Factor Analysis (CFA)* we want, in addition, that the residuals $X - Y$ are approximately orthogonal by columns. Thus we want to find A and B such that $X \approx AB'$ as well as $\mathbf{offdiag}\{(X - AB')'(X - AB')\} \approx 0$. Because we now want two things, instead of just one, the choice of the loss function becomes much less straightforward.

But before we discuss some of the more common loss function alternatives below, we generalize CFA by replacing the requirement of orthogonality of the residuals by the more general requirement that $(X - AB')'(X - AB')$ is approximately in $\mathscr{S}_m$, where $\mathscr{S}_m$ is a subset of the convex cone of positive semi-definite matrices of order m. The diagonal matrices defining CFA are just a special case of this.

Date: Wednesday 14th January, 2009 — 11h 4min — Typeset in TIMES ROMAN.

2. Multinormal Maximum Likelihood

Consider the multinormal maximum likelihood problem

$$\inf_{\Sigma\in\mathscr{S}_m}\inf_{\mathbf{rank}(Y)\leq p}\Delta_F(Y,\Sigma), \tag{1a}$$

where

$$\Delta_F(Y,\Sigma)=\log\mathbf{det}(\Sigma)+\mathbf{tr}\,\Sigma^{-1}(X-Y)'(X-Y), \tag{1b}$$

and where $\mathscr{S}_m$ is a subset of the convex cone of positive definite matrices of order m.

This is a straightforward generalization of maximum likelihood estimation in fixed score factor analysis as described by Young [1940]. In factor analysis the set $\mathscr{S}$ are the diagonal matrices. For that reasons we call the technique in which we can handle more general classes of weight matrices *Generalized Fixed Score Factor Analysis* or GFSFA.

In FSFA, as Lawley [1942] found, implementation of a straightforward block relaxation algorithm, in which we alternate minimization over Y of rank p for fixed Σ and minimization over diagonal Σ for fixed Y, leads to an unbounded decreasing sequence of loss function values. Thus the minimum is not attained, and the minimizer, which would give the maximum likelihood estimate in the FSFA model, does not exist. An actual proof was given by Anderson and Rubin [1956, section 7.7].

In this short note we generalize the result to GFSFA, with more general sets $\mathscr{S}_m$.

Theorem 2.1. *Suppose $\Sigma_n\in\mathscr{S}$ is a sequence of positive definite matrices of order m, converging to Σ_0, where Σ_0 has rank $m-1$. Suppose u is the normalized eigenvector satisfying $\Sigma_0u=0$ and assume $\Sigma_nu=\lambda_nu$ for all n. Then if $p>0$ we have*

$$\lim_{n\rightarrow\infty}\min_{\mathbf{rank}(Y)\leq p}\Delta(\Sigma_n,Y)=-\infty.$$

Proof. Suppose u is the unique (up to sign) unit length vector in the null space of Σ_0. Because Σ_0 commutes with Σ_n the vector u is also an eigenvector of Σ_n. Set $Y=Xuu'$ so that $X-Y=X(I-uu')$ and $(X-Y)'(X-Y)=(I-uu')X'X(I-uu')$. Moreover $\mathbf{rank}(Y)\leq 1\leq p$. Then, by continuity the smallest eigenvalue and the corresponding eigenvector,

$$\lim_{n\rightarrow\infty} \mathbf{tr}\ \Sigma_n^{-1}(X-Y)'(X-Y) = \lim_{n\rightarrow\infty} \mathbf{tr}\ (I-uu')\Sigma_n^{-1}(I-uu')X'X = \mathbf{tr}\ \Sigma_0^{+}X'X,$$

which is a finite non-negative number. On the other hand

$$\lim_{n\rightarrow\infty} \log \mathbf{det}(\Sigma_n) = -\infty.$$

□

Example 2.1. For Anderson and Rubin's result we take a sequence of positive definite diagonal matrices for which exactly one diagonal element converges to zero. All other diagonal elements can be fixed at one. Clearly Theorem 2.1 also applies if $\mathscr{S}$ contains the set of all diagonal matrices, for example if $\mathscr{S}$ is the set of all positive definite tri-diagonal matrices.

Example 2.2. Suppose $\mathscr{S}$ are the equi-correlation matrices, i.e. correlation matrices with all correlations equal to ρ. We can choose Σ_0 as the matrix with $\rho = -\frac{1}{n-1}$. This also proves, for example, that Theorem 2.1 applies to the symmetric positive definite Toeplitz correlation matrices, which has the equi-correlation matrices as a subset.

Note that the condition in the Theorem is sufficient for the log likelihood to be unbounded, but it is not necessary. We have a similar result if $\mathbf{rank}(\Sigma_0) = m-2$ and $p > 1$.

3. Remedies

3.1. **Smaller Covariance Structures.** One obvious way around the problem that maximum likelihood estimates do not exist is to work with a smaller $\mathscr{S}_m$, which does not have a singular matrix of rank $m-1$ in its closure.

In FSFA Young [1940] and Whittle [1952], for example, suggests to work with the scalar matrices $\sigma^2 I$. More generally, we could use $\sigma^2 D$, where D is any positive definite matrix. Jöreskog [1962], for example, suggests to choose D equal to the diagonal matrix with elements $s_j^2(1-R_j^2)$, where s_j^2 is the sample variance of variable j and R_j^2 is the multiple correlation of variable j with the remaining $m-1$ variables. In any case, we can find the maximum likelihood estimate of Y by computing the first p singular values and singular vectors of $XD^{\frac{1}{2}}$ and then compute the maximum likelihood estimate of σ^2 from the sum of squares of the $m-p$ smallest singular values. Computations are trivial, compared to the complicated iterative procedures in the general case.

In GFSFA we can replace the set of all Toeplitz matrices, for example, by the set of all matrices with elements $\sigma_{ij} = \sigma^2\rho^{|i-j|}$, where $-1 < \rho < +1$. These are special Toeplitz matrices, known as Kac-Murdoch-Szego matrices, after Kac et al. [1953], or AR(1) covariance matrices. They are nonsingular if $\sigma^2 > 0$, and the only non-zero singular matrices in the closure are the matrices with $\rho = \pm 1$, which are of rank 1.

Alternatively we can also use constraints on the parameters that rule out degeneracies. In the equi-correlation case, for example, we may require $\rho \geq 0$, which rules out the singular matrix that causes the problems.

3.2. **Smaller Mean Structures.** The constraint $\mathbf{rank}(Y) \leq p$ is equivalent to $Y = AB'$, where A and B are unknown matrices of dimension $n \times p$ and $m \times p$. We can require, instead, that $Y = ATB'$, with A and B known, or partially known, and T unknown. This is a Pothoff-Roy model, described in the context of growth curve analysis by Pothoff and Roy [1964]. Clearly there is a range of models here, going from A and B completely known (the regression situation) to A and B completely unknown (the principal component situation). The more structure we impose on Y, the less structure we have to impose on Σ, presumably.

3.3. **Augmented Least Squares.**

3.4. **Random Scores.** The log-likelihood so far has been based on the model of n i.i.d. random variables $\underline{x}_i \sim \mathcal{N}(Ba_i, \Sigma)$. We use the convention of underlining random variables [Hemelrijk, 1966].

Alternatively we can assume that $\mathcal{N}(Ba_i, \Sigma)$ is the conditional distribution of $\underline{x}_i$ given $\underline{a}_i = a_i$, where the $\underline{a}_i$ are i.i.d $\mathcal{N}(0, \Omega)$. It follows that $\underline{x}_i \sim \mathcal{N}(0, \Sigma + B\Omega B')$.

In this case the negative log-likelihood is

$$\Delta_R(B, \Sigma, \Omega) = \log \mathbf{det}(\Sigma + B\Omega B') + \mathbf{tr}\, X(\Sigma + B\Omega B')^{-1} X'.$$

This loss function can be written in different forms, using familiar matrix identities (see, for example, De Hoog et al. [1990] or De Leeuw and Meijer [2008, p. 65]).

$$\begin{aligned}\mathbf{tr}\, X(\Sigma+B\Omega B')^{-1}X' &= \min_A \left\{\mathbf{tr}\,\Sigma^{-1}(X-AB')'(X-AB')+A'\Omega^{-1}A\right\} = \\ &= \mathbf{tr}\, X(\Sigma^{-1}-\Sigma^{-1}B(\Omega^{-1}+B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1})^{-1}X'.\end{aligned}$$

$$\begin{aligned}\log\mathbf{det}(\Sigma+B\Omega B') &= \log\mathbf{det}(\Sigma)+\log\mathbf{det}(\Omega)+\log\mathbf{det}(\Omega^{-1}+B'\Sigma^{-1}B) = \\ &= \log\mathbf{det}(\Sigma)+\log\mathbf{det}(B'\Sigma^{-1}B)+\log\mathbf{det}(\Omega+(B'\Sigma^{-1}B)^{-1}).\end{aligned}$$

3.5. **Maximum Likelihood Ratio.** Suppose $\mathscr{S}_m \subset \mathscr{C}_m$ are two sets of positive definite matrices. The likelihood ratio test for testing $\mathscr{S}_m$ within $\mathscr{C}_m$ computes

$$\mathscr{L}_p(X) = \inf_{\mathbf{rank}(Y)\leq p}\ \inf_{\Sigma\in\mathscr{S}_m}\Delta(\Sigma,Y) - \inf_{\mathbf{rank}(Y)\leq p}\ \inf_{\Sigma\in\mathscr{C}_m}\Delta(\Sigma,Y).$$

As we have seen in many cases both terms are equal to $-\infty$, the maximum likelihood estimates do not exist, and the test cannot be used.

It was suggested by McDonald [1979], also see Etezadi-Amoli and McDonald [1983], to compute instead

$$\mathscr{M}_p(X) = \inf_{\mathbf{rank}(Y)\leq p}\left[\inf_{\Sigma\in\mathscr{S}_m}\Delta(\Sigma,Y) - \inf_{\Sigma\in\mathscr{C}_m}\Delta(\Sigma,Y)\right].$$

This amounts to computing "maximum likelihood ratio estimates" or MLR estimates. In the case of factor analysis we can choose $\mathscr{S}_m$ as the diagonal matrices and $\mathscr{C}_m$ as the set of all positive definite matrices. Then, with $S(X-Y) = (X-Y)'(X-Y)$,

$$\begin{aligned}\inf_{\Sigma\in\mathscr{S}_m}\Delta(\Sigma,Y) &= \log\mathbf{det}(\mathbf{diag}(S(X-Y)))+m,\\ \inf_{\Sigma\in\mathscr{C}_m}\Delta(\Sigma,Y) &= \log\mathbf{det}(S(X-Y))+m,\end{aligned}$$

and thus the MLR loss is

$$\mathscr{M}_p(X) = \inf_{\mathbf{rank}(Y)\leq p} -\log\mathbf{det}(R(X-Y)) = -\log\left\{\sup_{\mathbf{rank}(Y)\leq p}\mathbf{det}(R(X-Y))\right\},$$

where $R(X-Y)$ is the correlation matrix of the residuals $X-Y$. Thus we are maximizing the determinant of the correlation matrix of the residuals, which means making it as close to the identity as possible (in the determinant metric). This seems to be a perfectly respectable loss function for factor analysis, especially because $-\log\mathbf{det}(R(X-Y)) \geq 0$, which means the MLR loss function is bounded from below.

3.6. **Noncentral Wishart.** Anderson and Rubin [1956, section 11]

References

T.W. Anderson and H. Rubin. Statistical Inference in Factor Analysis. In J. Neyman, editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5*, pages 111–150. University of California Press, 1956.

F.R. De Hoog, T.P. Speed, and E.R. Williams. On a Matrix Indentity Associated with Generalized Least Squares. *Linear Algebra and Its Appplications*, 127: 449–456, 1990.

J. De Leeuw. Fixed-rank Approximation with Singular Weight Matrices. *Computational Statistics Quarterly*, 1:3–12, 1984.

J. De Leeuw and E. Meijer, editors. *Handbook of Multilevel Analysis*. Springer, 2008.

J. Etezadi-Amoli and R.P. McDonald. A Second Generation Nonlinear Factor Analysis. *Psychometrika*, 48:315–342, 1983.

J. Hemelrijk. Underlining random variables. *Statistica Neerlandica*, 20:1–7, 1966.

K.G. Jöreskog. On the Statistical Treatment of Residuals in Factor Analysis. *Psychometrika*, 27(4):335–354, 1962.

M. Kac, W.L. Murdock, and G. Szegö. On the Eigenvalues of Certain Hermitian Forms. *Journal of Rational Mechanics and Analysis*, 2:767–800, 1953.

D.N. Lawley. Further Investigations in Factor Estimation. *Proceedings of the Royal Society of Edinburgh*, 61:176–185, 1942.

R.P. McDonald. The Simultaneous Estimation of Factor Loadings and Scores. *British Journal of Mathematical and Statistical Psychology*, 32:212–228, 1979.

R. F. Pothoff and S. N. Roy. A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems. *Biometrika*, 51:313–326, 1964.

P. Whittle. On Principal Components and Least Squares Methods in Factor Analysis. *Skandinavisk Aktuarietidskrift*, 35:223–239, 1952.

G. Young. Maximum Likelihood Estimation and Factor Analysis of Data Matrices. *Psychometrika*, 6(1):49–53, 1940.

H. Zha. The Product-product Singular Value Decomposition of Matrix Triplets. *BIT*, 31:711–726, 1991.

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: `deleeuw@stat.ucla.edu`

URL, Jan de Leeuw: `http://gifi.stat.ucla.edu`