Chapter 11
Factor Analysis

11.1. Introduction

We will utilize the same notations as in the previous chapters. Lower-case letters $x, y, \ldots$ will denote real scalar variables, whether mathematical or random. Capital letters $X, Y, \ldots$ will be used to denote real matrix-variate mathematical or random variables, whether square or rectangular matrices are involved. A tilde will be placed on top of letters such as $\tilde{x}, \tilde{y}, \tilde{X}, \tilde{Y}$ to denote variables in the complex domain. Constant matrices will for instance be denoted by A, B, C. A tilde will not be used on constant matrices unless the point is to be stressed that the matrix is in the complex domain. In the real and complex cases, the determinant of a square matrix A will be denoted by $|A|$ or $\det(A)$ and, in the complex case, the absolute value or modulus of the determinant of A will be denoted as $|\det(A)|$. When matrices are square, their order will be taken as $p \times p$, unless specified otherwise. When A is a full rank matrix in the complex domain, then AA^* is Hermitian positive definite where an asterisk designates the complex conjugate transpose of a matrix. Additionally, $\mathrm{d}X$ will indicate the wedge product of all the distinct differentials of the elements of the matrix X. Thus, letting the $p \times q$ matrix $X = (x_{ij})$ where the x_{ij}'s are distinct real scalar variables, $\mathrm{d}X = \wedge_{i=1}^{p} \wedge_{j=1}^{q} \mathrm{d}x_{ij}$. For the complex matrix $\tilde{X} = X_1 + iX_2$, $i = \sqrt{(-1)}$, where X_1 and X_2 are real, $\mathrm{d}\tilde{X} = \mathrm{d}X_1 \wedge \mathrm{d}X_2$.

Factor analysis is a statistical method aiming to identify a relatively small number of underlying (unobserved) factors that could explain certain interdependencies among a larger set of observed variables. Factor analysis also proves useful for analyzing causal mechanisms. As a statistical technique, Factor Analysis was originally developed in connection with psychometrics. It has since been utilized in operations research, finance and biology, among other disciplines. For instance, a score available on an intelligence test will often assess several intellectual faculties and cognitive abilities. It is assumed that a

certain linear function of the contributions from these various mental factors is producing the final score. Hence, there is a parallel to be made with analysis of variance as well as design of experiments and linear regression models.

11.2. Linear Models from Different Disciplines

In order to introduce the current topic, we will first examine a linear regression model and an experimental design model.

11.2.1. A linear regression model

Let x be a real scalar random variable and let $t_1, \ldots, t_r$ be either r real fixed numbers or given values of r real random variables. Let the conditional expectation of x, given $t_1, \ldots, t_r$, be of the form

$$E[x|t_1, \ldots, t_r] = a_o + a_1 t_1 + \cdots + a_r t_r$$

or the corresponding model be

$$x = a_o + a_1 t_1 + \cdots + a_r t_r + e$$

where $a_o, a_1, \ldots, a_r$ are unknown constants, $t_1, \ldots, t_r$ are given values and e is the error component or the sum total of contributions coming from unknown or uncontrolled factors plus the experimental error. For example, x might be an inflation index with respect to a particular base year, say 2010. In this instance, t_1 may be the change or deviation in the average price per kilogram of certain staple vegetables from the base year 2010, t_2 may be the change or deviation in the average price of a kilogram of rice compared to the base year, t_3 may be the change or deviation in the average price of flour per kilogram with respect to the base year, and so on, and t_r may be the change or deviation in the average price of milk per liter compared to the base year 2010. The notation t_j, $j = 1, \ldots, r$, is utilized to designate the given values as well as the corresponding random variables. Since we are taking deviations from the base values, we may assume without any loss of generality that the expected value of t_j is zero, that is, $E[t_j] = 0$, $j = 1, \ldots, r$. We may also take the expected value of the error term e to be zero, that is, $E[e] = 0$. Now, let x_1 be the inflation index, x_2 be the caloric intake index per person, x_3 be the general health index and so on. In all these cases, the same $t_1, \ldots, t_r$ can act as the independent variables in a regression set up. Thus, in such a situation, a multivariate linear regression model will have the following format:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1r} \\ a_{21} & a_{22} & \ldots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \ldots & a_{pr} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_r \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}, \tag{11.2.1}$$

and we may write this model as

$$X = M + \Lambda F + \epsilon$$

where $\Lambda = (\lambda_{ij})$ is a $p \times r$, $r \le p$, matrix of full rank r, ϵ is $p \times 1$ and F is $r \times 1$. In (11.2.1), $\lambda_{ij} = a_{ij}$ and $f_j = t_j$. Then, $E[X] = M + \Lambda E[F] + E[\epsilon] = M$ since we have assumed that $E[F] = O$ (a null matrix) and $E[\epsilon] = O$. When F and ϵ are uncorrelated, the covariance matrix associated with X, denoted by $\text{Cov}(X) = \Sigma$, is the following:

$$\begin{aligned}
\Sigma = \text{Cov}(X) &= E\{(X - M)(X - M)'\} = E\{(\Lambda F + \epsilon)(\Lambda F + \epsilon)'\} \\
&= \Lambda \, \text{Cov}(F) \, \Lambda' + \text{Cov}(\epsilon) + O \\
&= \Lambda \, \Phi \, \Lambda' + \Psi
\end{aligned} \tag{11.2.2}$$

where the covariance matrices of F and ϵ are respectively denoted by $\Phi > O$ (positive definite) and $\Psi > O$. In the above formulation, F is taken to be a real vector random variable. In a simple linear model, the covariance matrix of ϵ, namely Ψ, is usually taken as $\sigma^2 I$ where $\sigma^2 > 0$ is a real scalar quantity and I is the identity matrix. In a more general setting, Ψ can be taken to be a diagonal matrix whose diagonal elements are positive; in such a model, the e_j's are uncorrelated and their variances need not be equal. It will be assumed that the covariance matrix Ψ in (11.2.2) is a diagonal matrix having positive diagonal elements.

11.2.2. A design of experiment model

Consider a completely randomized experiment where one set of treatments are undertaken. In this instance, the experimental plots are assumed to be fully homogeneous with respect to all the known factors of variation that may affect the response. For example, the observed value may be the yield of a particular variety of corn grown in an experimental plot. Let the set of treatments be r different fertilizers $F_1, \ldots, F_r$, the effects of these fertilizers being denoted by $\alpha_1, \ldots, \alpha_r$. If no fertilizer is applied, the yield from a test plot need not be zero. Let μ_1 be a general effect when F_1 is applied so that we may regard α_1 as a deviation from this effect μ_1 due to F_1. Let e_1 be the sum total of the contributions coming from all unknown or uncontrolled factors plus the experimental error, if any, when F_1 is applied. Then a simple linear one-way classification model for F_1 is

$$x_1 = \mu_1 + \alpha_1 + e_1,$$

with x_1 representing the yield from the test plot where F_1 was applied. Then, corresponding to $F_1, \ldots, F_r$, $r = p$ we have the following system:

$$
\begin{aligned}
x_1 &= \mu_1 + \alpha_1 + e_1 \\
&\vdots \\
x_p &= \mu_p + \alpha_p + e_p
\end{aligned}
$$

or, in matrix notation,

$$X = M + \Lambda F + \epsilon \tag{11.2.3}$$

where

$$
X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad
\epsilon = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix}, \quad
F = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix}, \quad \text{and } \Lambda =
\begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}.
$$

In this case, the elements of Λ are dictated by the design itself. If the vector F is fixed, we call the model specified by (11.2.3), the fixed effect model, whereas if F is assumed to be random, then it is referred to as the random effect model. With a single observation per cell, as stated in (11.2.3), we will not be able to estimate the parameters or test hypotheses. Thus, the experiment will have to be replicated. So, let the j-th replicated observation vector be

$$
X_j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{pj} \end{bmatrix}, \quad j = 1, \ldots, n,
$$

Σ, Φ, and Ψ remaining the same for each replicate within the random effect model. Similarly, for the regression model given in (11.2.1), the j-th replication or repetition vector will be $X_j' = (x_{1j}, \ldots, x_{pj})$ with Σ, Φ and Ψ therein remaining the same for each sample.

We will consider a general linear model encompassing those specified in (11.2.1) and (11.2.3) and carry out a complete analysis that will involve verifying the existence and uniqueness of such a model, estimating its parameters and testing various types of hypotheses. The resulting technique is referred to as Factor Analysis.

11.3. A General Linear Model for Factor Analysis

Consider the following general linear model:

$$X = M + \Lambda F + \epsilon \tag{11.3.1}$$

where

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \; M = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix}, \; \epsilon = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1r} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pr} \end{bmatrix} \text{ and } F = \begin{bmatrix} f_1 \\ \vdots \\ f_r \end{bmatrix}, \; r \le p,$$

with the μ_j's, λ_{ij}'s, f_j's being real scalar parameters, the x_j's, $j = 1, \dots, p$, being real scalar quantities, and Λ being of dimension $p \times r$, $r \le p$, and of full rank r. When considering expected values, variances, covariances, etc., X, F, and ϵ will be assumed to be random quantities; however, when dealing with estimates, X will represent a vector of observations. This convention will be employed throughout this chapter so as to avoid a multiplicity of symbols for the variables and the corresponding observations.

From a geometrical perspective, the r columns of Λ, which are linarly independent, span an r-dimensional subspace in the p-dimensional Euclidean space. In this case, the $r \times 1$ vector F is a point in this r-subspace and this subspace is usually called the *factor space*. Then, right-multiplying the $p \times r$ matrix Λ by a matrix will correspond to employing a new set of coordinate axes for the factor space.

Factor Analysis is a subject dealing with the identification or unique determination of a model of the type specified in (11.3.1), as well as the estimation of its parameters and the testing of various related hypotheses. The subject matter was originally developed in connection with intelligence testing. Suppose that a test is administered to an individual to evaluate his/her mathematical skills, spatial perception, language abilities, etc., and that the score obtained is recorded. There will be a component in the model representing the expected score. If the test is administered to 10th graders belonging to a particular school, the grand average of such test scores among all 10th graders across the nation could be taken as the expected score. Then, inputs associated to various intellectual faculties or combinations thereof will come about. All such factors may be contributing towards the observed test score. If $f_1, \dots, f_r$, are the contributions coming from r factors corresponding to specific intellectual abilities, then, when a linear model is assumed, a certain linear combination of these inputs will constitute the final quantity accounting for the observed test score. A test score, x_1, may then result from a linear model of the following form:

$$x_1 = \mu_1 + \lambda_{11} f_1 + \lambda_{12} f_2 + \dots + \lambda_{1r} f_r + e_1$$

with $\lambda_{11}, \ldots, \lambda_{1r}$ being the coefficients of $f_1, \ldots, f_r$, where $f_1, \ldots, f_r$, are contributions from r factors toward x_1; these factors may be called the *main intellectual factors* in this case, and the coefficients $\lambda_{11}, \ldots, \lambda_{1r}$ may be referred to as the *factor loadings* for these main factors. In this context, μ_1 is the general expected value and e_1 is the error component or the sum total of contributions originating from all unknown factors plus the experimental error, if any. Note that the contributions $f_1, \ldots, f_r$ due to the main intellectual factors can vary from individual to individual, and hence it is appropriate to treat $f_1, \ldots, f_r$ as random variables rather than fixed unknown quantities. These $f_1, \ldots, f_r$ are not observable as in the case of the design model in (11.2.3), whereas in the regression type model specified by (11.2.1), they may take on the recorded values of the observable variables which are called the independent variables. Thus, the model displayed in (11.3.1) may be analyzed either by treating $f_1, \ldots, f_r$ as fixed quantities or as random variables. If they are treated as random variables, we can assume that $f_1, \ldots, f_r$ follow some joint distribution. Usually, joint normality is presupposed for $f_1, \ldots, f_r$. Since $f_1, \ldots, f_r$ are deviations from the general effect μ_1 due to certain main intellectual faculties under consideration, it may be assumed that the expected value is a null vector, that is, $E[F] = O$. We will denote the covariance matrix associated with F as Φ: $\text{Cov}(F) = \Phi > O$ (real positive definite). Note that the model's error term e_j is always a random variable. Letting $x_1, \ldots, x_p$ be the test scores on p individuals, we have the error vector $\epsilon' = (e_1, \ldots, e_p)$. Without any loss of generality, we may take the expected value of ϵ as being a null vector, that is, $E[\epsilon] = O$. For a very simple situation, we may assume the covariance matrix associated with ϵ to be $\text{Cov}(\epsilon) = \sigma^2 I$ where $\sigma^2 > 0$ is a real positive scalar quantity and I is the identity matrix. For a somewhat more general situation, we may take $\text{Cov}(\epsilon) = \Psi$ where Ψ is a real positive definite diagonal matrix, or a diagonal matrix with positive diagonal elements. In the most general case, we may take Ψ to be a real positive definite matrix. It will be assumed that Ψ is diagonal with positive diagonal elements in the model (11.3.1), and that F and ϵ are uncorrelated. Thus, letting Σ be the covariance matrix of X, we have

$$\begin{aligned}
\Sigma = E[(X - M)(X - M)'] &= E[(\Lambda F + \epsilon)(\Lambda F + \epsilon)'] \\
&= \Lambda E(FF')\Lambda' + E(\epsilon\epsilon') + O \\
&= \Lambda \Phi \Lambda' + \Psi
\end{aligned} \tag{11.3.2}$$

where Σ, Φ and Ψ, with $\Sigma = \Lambda \Phi \Lambda' + \Psi$, are all assumed to be real positive definite matrices.

11.3.1. The identification problem

Is the model specified by (11.3.1) unique or could it represent different situations? In other words, does it make sense as a model as stated? Given any $r \times r$ nonsingular matrix A, let $AF = F^*$ and $\Lambda A^{-1} = \Lambda^*$. Then, $\Lambda^* F^* = \Lambda A^{-1} AF = \Lambda F$. In other words,

$$X = M + \Lambda F + \epsilon = M + \Lambda^* F^* + \epsilon. \tag{11.3.3}$$

Consequently, the model in (11.3.1) is not identified, that is, it is not uniquely determined.

The identification problem can also be stated as follows: Does there exist a real positive definite $p \times p$ matrix $\Sigma > O$ containing $p(p+1)/2$ distinct elements, which can be uniquely represented as $\Lambda \Phi \Lambda' + \Psi$ where Λ has pr distinct elements, $\Phi > O$ has $r(r+1)/2$ distinct elements and Ψ is a diagonal matrix having p distinct elements? There is clearly no such matrices as can be inferred from (11.3.3). Note that an $r \times r$ arbitrary matrix A represents r^2 distinct elements. It can be observed from (11.3.3) that we can impose r^2 conditions on the parameters in Λ, Φ and Ψ. The question could also be posed as follows: Can the $p(p+1)/2$ distinct elements in Σ plus the r^2 elements in A (r^2 conditions) uniquely determine all the elements of Λ, Ψ and Φ? Let us determine how many elements there are in total. Λ, Ψ and Φ have a total of $pr + p + r(r+1)/2$ elements while A and Σ have a total of $r^2 + p(p+1)/2$ elements. Hence, the difference, denoted by δ, is

$$\delta = \frac{p(p+1)}{2} + r^2 - \left[pr + \frac{r(r+1)}{2} + p \right] = \frac{1}{2}[(p-r)^2 - (p+r)]. \tag{11.3.4}$$

Observe that the right-hand side of (11.3.2) is not a linear function of Λ, Φ and Ψ. Thus, if $\delta > 0$, we can anticipate that existence and uniqueness will hold although these properties cannot be guaranteed, whereas if $\delta < 0$, then existence can be expected but uniqueness may be in question. Given (11.3.2), note that

$$\Sigma = \Psi + \Lambda \Phi \Lambda' \Rightarrow \Sigma - \Psi = \Lambda \Phi \Lambda'$$

where $\Lambda \Phi \Lambda'$ is positive semi-definite of rank r, since the $p \times r$, $r \leq p$, matrix Λ has full rank r and $\Phi > O$ (positive definite). Then, the existence question can also be stated as follows: Given a $p \times p$ real positive definite matrix $\Sigma > O$, can we find a diagonal matrix Ψ with positive diagonal elements such that $\Sigma - \Psi$ is a real positive semi-definite matrix of a specified rank r, which is expressible in the form BB' for some $p \times r$ matrix B of rank r where $r \leq p$? For the most part, the available results on this question of existence can be found in Anderson (2003) and Anderson and Rubin (1956). *If a set of parameters exist and if the model is uniquely determined, then we say that the model is identified, or*

alternatively, identifiable. The concept of identification or identifiability within the context of Factor Analysis has been studied by Ihara and Kano (1995), Wegge (1996), Allman et al. (2009) and Chen et al. (2020), among others.

Assuming that $\Phi = I$ will impose $r(r + 1)/2$ conditions. However, $r^2 = \frac{r(r+1)}{2} + \frac{r(r-1)}{2}$. Thus, we may impose $r(r - 1)/2$ additional conditions after requiring that $\Phi = I$. Observe that when $\Phi = I$, $\Lambda^* \Phi \Lambda'^* = \Lambda^* \Lambda'^* = \Lambda A^{-1} A'^{-1} \Lambda'$, and if this is equal to $\Lambda \Lambda'$ assuming that $\Phi = I$, this means that $(A'A)^{-1} = I$ or $A'A = I$, that is, A is an orthonormal matrix. Thus, under the condition $\Phi = I$, the arbitrary $r \times r$ matrix A becomes an orthonormal matrix. In this case, the transformation $Y = \Lambda A$ is an orthonormal transformation or a rotation of the coordinate axes. The following $r \times r$ symmetric matrix of $r(r + 1)/2$ distinct elements

$$\Delta = \Lambda' \Psi^{-1} \Lambda \tag{11.3.5}$$

is needed for solving estimation and hypothesis testing problems; accordingly, we can impose $r(r - 1)/2$ conditions by requiring Δ to be diagonal with distinct diagonal elements, that is, $\Delta = \text{diag}(\eta_1, \ldots, \eta_r)$, $\eta_j > 0$, $j = 1, \ldots, r$. This imposes $\frac{r(r+1)}{2} - r = \frac{r(r-1)}{2}$ conditions. Thus, for the model to be identifiable or for all the parameters in Λ, Φ, Ψ to be uniquely determined, we can impose the condition $\Phi = I$ and require that $\Delta = \Lambda' \Psi^{-1} \Lambda$ be diagonal with positive diagonal elements. These two conditions will provide $\frac{r(r+1)}{2} + \frac{r(r-1)}{2} = r^2$ restrictions on the model which will then be identified.

When $\Phi = I$, the main factors are orthogonal. If Φ is a diagonal matrix (including the identity matrix), the covariances are zeros and it is an orthogonal situation, in which case we say that the *main factors are orthogonal*. If Φ is not diagonal, then we say that the *main factors are oblique*.

One can also impose $r(r - 1)/2$ conditions on the $p \times r$ matrix Λ. Consider the first $r \times r$ block, that is, the leading $r \times r$ submatrix or the upper $r \times r$ block in the $p \times r$ matrix, which we will denote by B. Imposing the condition that this $r \times r$ block B is lower triangular will result in $r^2 - \frac{r(r+1)}{2} = \frac{r(r-1)}{2}$ conditions. Hence, $\Phi = I$ and the condition that this leading $r \times r$ block B is lower triangular will guarantee r^2 restrictions, and the model will then be identified. One can also take a preselected $r \times r$ matrix B_1 and then impose the condition that $B_1 B$ be lower triangular. This will, as well, produce $\frac{r(r-1)}{2}$ conditions. Thus, $\Phi = I$ and $B_1 B$ being lower triangular will ensure the identification of the model.

When we impose conditions on Φ and Ψ, the unknown covariance matrices must assume certain formats. Such conditions can be justified. However, could conditions be put on Λ, the factor loadings? Letting the first $r \times r$ block B in the $p \times r$ matrix Λ be lower triangular is tantamount to assuming that $\lambda_{12} = 0 = \lambda_{13} = \cdots = \lambda_{1r}$ or, equivalently, that

$f_2, \ldots, f_r$ do not contribute to the model for determining x_1 in $X' = (x_1, x_2, \ldots, x_p)$. Such restrictions are justified if we can design the experiment in such a way that x_1 depends on f_1 alone and not on $f_2, \ldots, f_r$. In psychological tests, it is possible to design tests in such a way that only certain main factors affect the scores. Thus, in such instances, we are justified to utilize a triangular format such that, in general, there are no contributions from $f_{i+1}, \ldots, f_r$, toward x_i or, equivalently, the factor loadings $\lambda_{i\,i+1}, \ldots, \lambda_{ir}$ equal zero for $i = 1, \ldots, r - 1$. For example, suppose that the first r tests are designed in such a way that only $f_1, \ldots, f_i$ and no other factors contribute to x_i or, equivalently, $x_i = \mu_i + \lambda_{i1} f_1 + \cdots + \lambda_{ii} f_i + e_i$, $i = 1, \ldots, r$. We can also measure the contribution from f_i in λ_{ii} units or we can take $\lambda_{ii} = 1$. By taking $B = I_r$, we can impose r^2 conditions without requiring that $\Phi = I$. This means that the first r tests are specifically designed so that x_1 only has a one unit contribution from f_1, x_2 only has a one unit contribution from f_2, and so on, x_r receiving a one unit contribution from f_r. When B is taken to be diagonal, the factor loadings are $\lambda_{11}, \lambda_{22}, \ldots, \lambda_{rr}$, respectively, so that only f_i contributes to x_i for $i = 1, \ldots, r$. Accordingly, the following are certain model identification conditions:

(1): $\Phi = I$ and $\Lambda' \Psi^{-1} \Lambda$ is diagonal with distinct diagonal elements;

(2): $\Phi = I$ and the leading $r \times r$ submatrix B in the $p \times r$ matrix Λ is triangular;

(3): $\Phi = I$ and $B_1 B$ is lower triangular where B_1 is a preselected matrix;

(4): The leading $r \times r$ submatrix B in the $p \times r$ matrix Λ is an identity matrix.

Observe that when $r = p$, condition (4) corresponds to the design model considered in (11.2.3).

11.3.2. Scaling or units of measurement

A shortcoming of any analysis being based on a covariance matrix Σ is that the covariances depend on the units of measurement of the individual variables. Thus, modifying the units will affect the covariances. If we let y_i and y_j be two real scalar random variables with variances σ_{ii} and σ_{jj} and associated covariance σ_{ij}, the effect of scaling or changes in the measurement units may be eliminated by considering the variables $z_i = y_i / \sqrt{\sigma_{ii}}$ and $z_j = y_j / \sqrt{\sigma_{jj}}$ whose covariance $\text{Cov}(z_i, z_j) \equiv r_{ij}$ is actually the correlation between y_i and y_j, which is free of the units of measurement. Letting $Y' = (y_1, \ldots, y_p)$ and $D = \text{diag}(\frac{1}{\sqrt{\sigma_{11}}}, \ldots, \frac{1}{\sqrt{\sigma_{pp}}})$, consider $Z = DY$. We note that $\text{Cov}(Y) = \Sigma \Rightarrow \text{Cov}(Z) = D\Sigma D = R$ which is the correlation matrix associated with Y.

In psychological testing situations or referring to the model (11.3.1), when a test score x_j is multiplied by a scalar quantity c_j, then the factor loadings $\lambda_{j1}, \ldots, \lambda_{jr}$, the error term e_j and the general effect μ_j are all multiplied by c_j, that is, $c_j x_j = c_j \mu_j + c_j(\lambda_{j1} f_1 + \cdots + \lambda_{jr} f_r) + c_j e_j$. Let $\text{Cov}(X) = \Sigma = (\sigma_{ij})$, that is, $\text{Cov}(x_i,\ x_j) = \sigma_{ij}$, $X' = (x_1, \ldots, x_p)$ and $D = \text{diag}(\frac{1}{\sqrt{\sigma_{11}}}, \ldots, \frac{1}{\sqrt{\sigma_{pp}}})$. Consider the model

$$DX = DM + D\Lambda F + D\epsilon \Rightarrow \text{Cov}(DX) = D\Sigma D = D\Lambda\Phi\Lambda'D + D\Psi D. \quad (11.3.6)$$

If $X^* = DX$, $M^* = DM$, $\Lambda^* = D\Lambda$, and $\epsilon^* = D\epsilon$, then we obtain the following model and the resulting covariance matrix:

$$X^* = M^* + \Lambda^* F + \epsilon^* \Rightarrow \Sigma^* = \text{Cov}(X^*) = \Lambda^*\text{Cov}(F)\Lambda^{*'} + \Psi^*$$
$$\Rightarrow D\Sigma D = D\Lambda\Phi\Lambda'D + D\Psi D$$
$$\Rightarrow R = \Lambda^*\Phi\Lambda^{*'} + \Psi^* \quad (11.3.7)$$

where $R = (r_{ij})$ is the correlation matrix in X. An interesting point to be noted is that the identification conditions $\Phi = I$ and $\Lambda^{*'}\Psi^{*-1}\Lambda^*$ being diagonal become the following: $\Phi = I$ and $\Lambda^{*'}\Psi^{*-1}\Lambda^* = \Lambda'DD^{-1}\Psi^{-1}D^{-1}D\Lambda = \Lambda'\Psi^{-1}\Lambda$ which is diagonal, that is, $\Lambda'\Psi^{-1}\Lambda$ is invariant under scaling transformations on the model or under $X^* = DX$ and $\Psi^* = D\Psi D$.

11.4. Maximum Likelihood Estimators of the Parameters

A simple random sample of size n from the model $X = M + \Lambda F + \epsilon$ specified in (11.3.1) is understood to be constituted of independently and identically distributed (iid) X_j's, $j = 1, \ldots, n$, where

$$X_j = M + \Lambda F + E_j,\ j = 1, \ldots, n,\ X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{pj} \end{bmatrix},\ E_j = \begin{bmatrix} e_{1j} \\ e_{2j} \\ \vdots \\ e_{pj} \end{bmatrix}, \quad (11.4.1)$$

and the X_j's are iid. Let f_j and E_j be independently normally distributed. Let $E_j \sim N_p(O,\ \Psi)$ and $X_j \sim N_p(\mu,\ \Sigma)$, $\Sigma = \Lambda\Phi\Lambda' + \Psi$ where $\Phi = \text{Cov}(F) > O$, $\Psi = \text{Cov}(\epsilon) > O$ and $\Sigma > O$, Ψ being a diagonal matrix with positive diagonal elements. Then, the likelihood function is the following:

$$L = \prod_{j=1}^{n} \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_j - M)'\Sigma^{-1}(X_j - M)}$$
$$= \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{j=1}^{n}(X_j - M)'\Sigma^{-1}(X_j - M)}. \quad (11.4.2)$$

The sample matrix will be denoted by a boldface $\mathbf{X} = (X_1, \ldots, X_n)$. Let J be the $n \times 1$ vector of unities, that is, $J = (1, 1, \ldots, 1)'$. Then,

$$\mathbf{X} = (X_1, \ldots, X_n) = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \ldots & x_{pn} \end{bmatrix}$$

$$\Rightarrow \frac{1}{n} \mathbf{X} J = \begin{bmatrix} \frac{1}{n}(\sum_{j=1}^{n} x_{1j}) \\ \vdots \\ \frac{1}{n}(\sum_{j=1}^{n} x_{pj}) \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \bar{X}$$

where $\bar{X}$ is the sample average vector or the sample mean vector. Let the boldface $\bar{\mathbf{X}}$ be the $p \times n$ matrix $\bar{\mathbf{X}} = (\bar{X}, \bar{X}, \ldots, \bar{X})$. Then,

$$(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})' = S = (s_{ij}), \; s_{ij} = \sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j), \tag{11.4.3}$$

where S is the sample sum of products matrix or the "corrected" sample sum of squares and cross products matrix. Note that

$$\bar{X} = \frac{1}{n} \mathbf{X} J$$

$$\Rightarrow \bar{\mathbf{X}} = (\bar{X}, \ldots, \bar{X}) = \mathbf{X}\left(\frac{1}{n} J J'\right)$$

$$\Rightarrow \mathbf{X} - \bar{\mathbf{X}} = \mathbf{X}\left(I - \frac{1}{n} J J'\right).$$

Thus,

$$S = \mathbf{X}\left(I - \frac{1}{n} J J'\right)\left(I - \frac{1}{n} J J'\right)' \mathbf{X}' = \mathbf{X}\left(I - \frac{1}{n} J J'\right)\mathbf{X}'. \tag{11.4.4}$$

Since $(X_j - M)'\Sigma^{-1}(X_j - M)$ is a real scalar quantity, we have the following:

$$
\begin{aligned}
\sum_{j=1}^{n}(X_j - M)'\Sigma^{-1}(X_j - M) &= \sum_{j=1}^{n}\mathrm{tr}(X_j - M)'\Sigma^{-1}(X_j - M)\\
&= \sum_{j=1}^{n}\mathrm{tr}[\Sigma^{-1}(X_j - M)(X_j - M)']\\
&= \mathrm{tr}[\Sigma^{-1}\sum_{j=1}^{n}(X_j - \bar{X} + \bar{X} - M)(X_j - \bar{X} + \bar{X} - M)']\\
&= \mathrm{tr}[\Sigma^{-1}\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})']\\
&\quad + n\,\mathrm{tr}[\Sigma^{-1}(\bar{X} - M)(\bar{X} - M)']\\
&= \mathrm{tr}(\Sigma^{-1}S) + n(\bar{X} - M)'\Sigma^{-1}(\bar{X} - M). \qquad (11.4.5)
\end{aligned}
$$

Hence,

$$
L = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}}\mathrm{e}^{-\frac{1}{2}\{\mathrm{tr}(\Sigma^{-1}S)+n(\bar{X}-M)'\Sigma^{-1}(\bar{X}-M)\}}. \qquad (11.4.6)
$$

Differentiating (11.4.6) with respect to M, equating the result to a null vector and solving, we obtain an estimator for M, denoted by $\hat{M}$, as $\hat{M} = \bar{X}$. Then, $\ln L$ evaluated at $M = \bar{X}$ is

$$
\begin{aligned}
\ln L &= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}S)\\
&= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Lambda\Phi\Lambda' + \Psi| - \frac{1}{2}\mathrm{tr}[(\Lambda\Phi\Lambda' + \Psi)^{-1}S]. \qquad (11.4.7)
\end{aligned}
$$

11.4.1. Maximum likelihood estimators under identification conditions

The derivations in the following sections parallel to those found in Mathai (2021). One of the conditions for identification of the model is $\Phi = I$ and $\Lambda'\Psi^{-1}\Lambda$ being a diagonal matrix with positive diagonal elements. We will examine the maximum likelihood estimators (MLE)/maximum likelihood estimates (MLE) under this identification condition. In this case, it follows from (11.4.7) that

$$
\ln L = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Lambda\Lambda' + \Psi| - \frac{1}{2}\mathrm{tr}[(\Lambda\Lambda' + \Psi)^{-1}S]. \qquad (11.4.8)
$$

By expanding the following determinant in two different ways by applying properties of partitioned determinants that are stated in Sect. 1.3, we have the following identities:

$$\begin{vmatrix} \Psi & -\Lambda \\ \Lambda' & I_r \end{vmatrix} = |\Psi| \, |I + \Lambda' \Psi^{-1} \Lambda| = |\Psi + \Lambda \Lambda'|. \tag{11.4.9}$$

Hence, letting

$$\Delta = \Lambda' \Psi^{-1} \Lambda = \text{diag}(\delta_1' \delta_1, \delta_2' \delta_2, \ldots, \delta_r' \delta_r),$$

we have

$$\ln |\Lambda \Lambda' + \Psi| = \ln |\Psi| + \ln |I + \Lambda' \Psi^{-1} \Lambda|$$

$$= \sum_{j=1}^{p} \ln \psi_{jj} + \sum_{j=1}^{r} \ln(1 + \delta_j' \delta_j), \tag{11.4.10}$$

where $\delta_j' = \Lambda_j' \Psi^{-\frac{1}{2}}$, Λ_j is the j-th column of Λ and ψ_{jj}, $j = 1, \ldots, p$, is the j-th diagonal element of the diagonal matrix Ψ, the identification condition being that $\Phi = I$ and $\Lambda' \Psi^{-1} \Lambda = \Delta = \text{diag}(\delta_1' \delta_1, \ldots, \delta_r' \delta_r)$. Accordingly, if we can write $\text{tr}(\Sigma^{-1} S) = \text{tr}[(\Lambda \Lambda' + \Psi)^{-1} S]$ in terms of ψ_{jj}, $j = 1, \ldots, p$, and $\delta_j' \delta_j$, $j = 1, \ldots, r$, then the likelihood equation can be directly evaluated from (11.4.8) and (11.4.10), and the estimators can be determined. The following result will be helpful in this connection.

Theorem 11.4.1. *Whenever $\Lambda \Lambda' + \Psi$ is nonsingular, which in this case, means real positive definite, the inverse is given by*

$$(\Lambda \Lambda' + \Psi)^{-1} = \Psi^{-1} - \Psi^{-1} \Lambda (\Delta + I)^{-1} \Lambda' \Psi^{-1} \tag{11.4.11}$$

where the Δ is defined in (11.4.10).

It can be readily verified that pre and post multiplications of $\Psi^{-1} - \Psi^{-1} \Lambda (\Delta + I)^{-1} \Lambda' \Psi^{-1}$ by $\Lambda \Lambda' + \Psi$ yield the identity matrix I_p.

11.4.2. Simplifications of $|\Sigma|$ and $\text{tr}(\Sigma^{-1} S)$

In light of (11.4.9) and (11.4.10), we have

$$|\Sigma| = |\Lambda \Lambda' + \Psi| = |\Psi| \, |\Lambda' \Psi^{-1} \Lambda + I|$$

$$= |\Psi| \, |I + \Delta| = \left\{ \prod_{j=1}^{p} \psi_{jj} \right\} \left\{ \prod_{j=1}^{r} (1 + \delta_j' \delta_j) \right\}.$$

Now, observe the following: In $\Lambda(\Delta+I)^{-1} = \Lambda \, \text{diag}(\frac{1}{1+\delta'_1\delta_1}, \ldots, \frac{1}{1+\delta'_r\delta_r})$, the j-th column of Λ is multiplied by $\frac{1}{1+\delta'_j\delta_j}$, $j = 1, \ldots, r$, and

$$\Lambda(\Delta + I)^{-1}\Lambda' = \sum_{j=1}^{r} \frac{1}{1 + \delta'_j\delta_j} \Lambda_j \Lambda'_j$$

where Λ_j is the j-th column of Λ and the δ_j's are specified in (11.4.10). Thus,

$$\ln |\Sigma| = \sum_{j=1}^{p} \ln \psi_{jj} + \sum_{j=1}^{r} \ln(1 + \delta'_j\delta_j) \tag{11.4.12}$$

and, on applying Theorem 11.4.1,

$$\text{tr}(\Sigma^{-1}S) = \text{tr}[(\Lambda\Lambda' + \Psi)^{-1}S] = \text{tr}[(\Psi^{-1}S)] - \text{tr}[\Psi^{-1}\Lambda(\Delta + I)^{-1}\Lambda'\Psi^{-1}S]$$

$$= \text{tr}(\Psi^{-1}S) - \sum_{j=1}^{r} \frac{1}{1 + \delta'_j\delta_j} \text{tr}(\Lambda_j\Lambda'_j(\Psi^{-1}S\Psi^{-1}))$$

$$= \text{tr}(\Psi^{-1}S) - \sum_{j=1}^{r} \frac{1}{1 + \delta'_j\delta_j} \text{tr}(\Lambda'_j(\Psi^{-1}S\Psi^{-1})\Lambda_j)$$

$$= \text{tr}(\Psi^{-1}S) - \sum_{j=1}^{r} \frac{1}{1 + \delta'_j\delta_j} \Lambda'_j(\Psi^{-1}S\Psi^{-1})\Lambda_j \tag{11.4.13}$$

where Λ_j is the j-th column of Λ, which follows by making use of the property $\text{tr}(AB) = \text{tr}(BA)$ and observing that $\Lambda'_j(\Psi^{-1}S\Psi^{-1})\Lambda_j$ is a quadratic form.

11.4.3. Special case $\Psi = \sigma^2 I_p$

Letting $\Psi = \sigma^2 I$ where σ^2 is a real scalar, $\Psi^{-1} = \sigma^{-2}I_p = \theta I_p$ where $\theta = \sigma^{-2}$, and the log-likelihood function can be simplified as

$$\ln L = -\frac{np}{2} \ln(2\pi) + \frac{np}{2} \ln \theta - \frac{n}{2} \sum_{j=1}^{r} \ln(1 + \delta'_j\delta_j)$$

$$- \frac{\theta}{2}\text{tr}(S) + \frac{\theta^2}{2} \sum_{j=1}^{r} \frac{1}{1 + \delta'_j\delta_j} \Lambda'_j S \Lambda_j$$

where $1 + \delta'_j \delta_j = 1 + \theta \Lambda'_j \Lambda_j$ with Λ_j being the j-th column of Λ. Consider the equation

$$\frac{\partial}{\partial \theta} \ln L = 0 \Rightarrow$$

$$\frac{np}{\theta} - n \sum_{j=1}^{r} \frac{\Lambda'_j \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \mathrm{tr}(S)$$

$$+ 2\theta \sum_{j=1}^{r} \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \theta^2 \sum_{j=1}^{r} \frac{\Lambda'_j \Lambda_j}{(1 + \theta \Lambda'_j \Lambda_j)^2} \Lambda'_j S \Lambda_j = 0. \qquad (11.4.14)$$

For a specific j, we have

$$\frac{\partial}{\partial \Lambda_j} \ln L = O \Rightarrow$$

$$-\frac{n}{2} \frac{2\theta \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} + \frac{\theta^2}{2} \frac{2 S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \frac{\theta^2}{2} \frac{\Lambda'_j S \Lambda_j}{[1 + \theta \Lambda'_j \Lambda_j]^2} (2\theta) \Lambda_j = O. \qquad (11.4.15)$$

Pre-multiplying (11.4.15) by Λ'_j yields

$$-\frac{n\theta \Lambda'_j \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} + \frac{\theta^2 \Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} - \frac{\theta^3 (\Lambda'_j S \Lambda_j)(\Lambda'_j \Lambda_j)}{(1 + \theta \Lambda'_j \Lambda_j)^2} = 0. \qquad (11.4.16)$$

Now, on comparing (11.4.16) with (11.4.14) after multiplying (11.4.14) by θ, we have

$$np - \theta \, \mathrm{tr}(S) + \theta^2 \sum_{j=1}^{r} \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} = 0. \qquad (11.4.17)$$

Multiplying the left-hand side of (11.4.15) by $[1 + \theta \Lambda'_j \Lambda_j]^2 / \theta$ gives

$$-n(1 + \theta \Lambda'_j \Lambda_j) \Lambda_j + \theta(1 + \theta \Lambda'_j \Lambda_j) S \Lambda_j - \theta^2 (\Lambda'_j S \Lambda_j) \Lambda_j = O. \qquad (11.4.18)$$

Then, by pre-multiplying (11.4.18) by Λ'_j, we obtain

$$-n(1 + \theta \Lambda'_j \Lambda_j) \Lambda'_j \Lambda_j + \theta[(1 + \theta \Lambda'_j \Lambda_j) \Lambda'_j S \Lambda_j - \theta^2 (\Lambda'_j S \Lambda_j) \Lambda'_j \Lambda_j = 0 \Rightarrow$$
$$\theta(\Lambda'_j S \Lambda_j) = n(1 + \theta \Lambda'_j \Lambda_j)(\Lambda'_j \Lambda_j), \qquad (11.4.19)$$

which provides the following representation of θ:

$$\theta = \frac{n \Lambda'_j \Lambda_j}{\Lambda'_j S \Lambda_j - n(\Lambda'_j \Lambda_j)^2} \quad \text{for } \Lambda'_j S \Lambda_j - n(\Lambda'_j \Lambda_j)^2 > 0 \qquad (i)$$

since θ must be positive. Further, on substituting $\theta \Lambda'_j S \Lambda_j \overset{(11.4.19)}{=} n(1 + \theta \Lambda'_j \Lambda_j)(\Lambda'_j \Lambda_j)$ in (11.4.18), we have

$$-n \Lambda_j + \theta [S - n(\Lambda'_j \Lambda_j)I] \Lambda_j = O, \qquad (ii)$$

which, on replacing θ by the right-hand side of *(i)* yields

$$-(\Lambda'_j S \Lambda_j) \Lambda_j + (\Lambda'_j \Lambda_j) S \Lambda_j = O$$

or, equivalently,

$$\left[S - \frac{\Lambda'_j S \Lambda_j}{\Lambda'_j \Lambda_j} I \right] \Lambda_j = O \Rightarrow \qquad (11.4.20)$$

$$[S - \lambda_j I] \Lambda_j = O \qquad (11.4.21)$$

where $\lambda_j = \frac{\Lambda'_j S \Lambda_j}{\Lambda'_j \Lambda_j}$. Observe that Λ_j in (11.4.20) is an eigenvector of S for $j = 1, \ldots, p$. Substituting the value of $\Lambda'_j S \Lambda_j$ from (11.4.19) into (11.4.17) gives the following estimate of θ:

$$\hat{\theta} = \frac{np}{\text{tr}(S) - n \sum_{j=1}^p \hat{\Lambda}'_j \hat{\Lambda}_j} \qquad (11.4.22)$$

whenever the denominator is positive as θ is by definition positive. Now, in light of (11.4.18) and (11.4.19), we can also obtain the following result for each j:

$$\hat{\theta} = \frac{n \hat{\Lambda}'_j \hat{\Lambda}_j}{\hat{\Lambda}'_j S \hat{\Lambda}_j - n(\hat{\Lambda}'_j \hat{\Lambda}_j)^2}, \ j = 1, \ldots, p, \qquad (11.4.23)$$

requiring again that the denominator be positive. In this case, $\hat{\Lambda}_j$ is an eigenvector of S for $j = 1, \ldots, p$. Let us conveniently normalize the $\hat{\Lambda}_j$'s so that the denominator in (11.4.22) and (11.4.23) remain positive.

Thus, $\hat{\Lambda}_j$ is an eigenvector of S with the corresponding eigenvalue λ_j for each j, $j = 1, \ldots, p$. Out of these, the first r of them, corresponding to the r largest eigenvalues, will also be estimates for the factor loadings Λ_j, $j = 1, \ldots, r$. Observe that we can multiply $\hat{\Lambda}_j$ by any constant c_1 without affecting equations (11.4.20) or (11.4.21). This constant c_1 may become necessary to keep the denominators in (11.4.22) and (11.4.23) positive. Hence we have the following result:

Theorem 11.4.2. *The sum of all the eigenvalues of S from Eq. (11.4.20), including the estimates of the r factor loadings* $\hat{\Lambda}_1, \ldots, \hat{\Lambda}_p$, *is given by*

$$\sum_{j=1}^{p} \frac{\hat{\Lambda}'_j S \hat{\Lambda}_j}{\hat{\Lambda}'_j \hat{\Lambda}_j} = \text{tr}(S). \tag{11.4.24}$$

It can be established that the representations of $\hat{\theta}$ given by (11.4.22) and (11.4.23) are one and the same. The equation giving rise to (11.4.23) is

$$\theta[\Lambda'_j S \Lambda_j - n(\Lambda'_j \Lambda_j)^2] = n \Lambda'_j \Lambda_j \text{ for each } j. \tag{iii}$$

Let us divide both sides of *(iii)* by $\Lambda'_j \Lambda_j$. Observe that $\frac{\Lambda'_j S \Lambda_j}{\Lambda'_j \Lambda_j} = \lambda_j$ is an eigenvalue of S for $j = 1, \ldots, p$, treating Λ_j as an eigenvector of S. Now, taking the sum over $j = 1, \ldots, p$, on both sides of *(iii)* after dividing by $\Lambda'_j \Lambda_j$, we have

$$\theta \left[\sum_{j=1}^{p} \lambda_j - n \sum_{j=1}^{p} (\Lambda'_j \Lambda_j) \right] = np \Rightarrow$$

$$\theta \left[\text{tr}(S) - n \sum_{j=1}^{p} \Lambda'_j \Lambda_j \right] = np, \tag{iv}$$

which is Eq. (11.4.22). This proves the claim.

Hence the procedure is the following: Compute the eigenvalues and the corresponding eigenvectors of the sample sum of products matrix S. The estimates for the factor loadings, denoted by $\hat{\Lambda}_j$, are available from the eigenvectors $\hat{\Lambda}_j$ of S after appropriate normalization to make the denominators in (11.4.22) and (1.4.23) positive. Take the first r largest eigenvalues of S and then compute the corresponding eigenvectors to obtain estimates for all the factor loadings. This methodology is clearly related to that utilized in Principal Component Analysis, the estimates of the variances of the principal components being $\hat{\Lambda}'_j S \hat{\Lambda}_j / \hat{\Lambda}'_j \hat{\Lambda}_j$ for $j = 1, \ldots, r$.

Verification

Does the representation of θ given in (11.4.22) and (11.4.23) satisfy the likelihood Eq. (11.4.14)? Since θ is estimated through Λ_j for each $j = 1, \ldots, p$, we may replace θ in (11.4.14) by θ_j and insert the summation symbol. Equation (11.4.14) will then be

$$n \sum_j \frac{1}{\theta_j} - n \sum_j \frac{\Lambda'_j \Lambda_j}{1 + \theta_j \Lambda'_j \Lambda_j} - \mathrm{tr}(S)$$

$$+ 2 \sum_j \theta_j \frac{\Lambda'_j S \Lambda_j}{1 + \theta_j \Lambda'_j \Lambda_j} - \sum_j \theta_j^2 \frac{\Lambda'_j \Lambda_j (\Lambda'_j S \Lambda_j)}{(1 + \theta_j \Lambda'_j \Lambda_j)^2} = 0. \qquad (11.4.25)$$

Now, substituting the value of θ_j specified in (11.4.23) into (11.4.14), the left-hand side of (11.4.14) reduces to the following:

$$n \sum_j \frac{[\Lambda'_j S \Lambda_j - n(\Lambda'_j \Lambda_j)^2]}{n \Lambda'_j \Lambda_j} - n \sum_j \frac{\Lambda'_j \Lambda_j}{\Lambda'_j S \Lambda_j}[\Lambda'_j S \Lambda_j - n(\Lambda'_j \Lambda_j)^2] - \mathrm{tr}(S)$$

$$+ 2 \sum_j n \Lambda'_j \Lambda_j - \sum_j \frac{\Lambda'_j \Lambda_j}{\Lambda_j S \Lambda_j}(n \Lambda'_j \Lambda_j)^2$$

$$= \sum_j \frac{\Lambda'_j S \Lambda_j}{\Lambda'_j \Lambda_j} - \mathrm{tr}(S) = 0,$$

owing to Theorem 11.4.2. Hence, Eq. (11.4.14) holds for the value of θ given in (11.4.23) and the value of Λ_j specified in (11.4.20).

Since the basic estimating equation for $\hat{\theta}$ arises from (11.4.23) as

$$\theta[\Lambda'_j S \Lambda_j - n(\Lambda'_j \Lambda_j)^2] = n \Lambda'_j \Lambda_j, \qquad (v)$$

a combined estimate for θ can be secured. On dividing both sides of *(v)* by $\Lambda'_j \Lambda_j$ and summing up over j, $j = 1, \ldots, p$, it is seen that the resulting estimate of θ agrees with that given in (11.4.22).

11.4.4. Maximum value of the exponent

We have the estimate $\hat{\theta}$ of θ provided in (11.4.23) at the estimated value $\hat{\Lambda}_j$ of Λ_j for each j, where $\hat{\Lambda}_j$ is an eigenvector of S resulting from (11.4.20). The exponent of the likelihood function is $-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}S)$ and, in the current context, $\Sigma = \Lambda \Phi \Lambda' + \Psi$, the identification conditions being that $\Phi = I_p$ and $\Lambda' \Psi^{-1} \Lambda$ be a diagonal matrix with positive diagonal elements. Under these conditions and for the special case $\Psi = \sigma^2 I_p$ with $\sigma^{-2} = \theta$, we have shown that the exponent in the log-likelihood function reduces to $-\frac{1}{2}\theta \, \mathrm{tr}(S) + \frac{1}{2}\theta^2 \sum_{j=1}^r \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j}$. Now, consider $\theta \, \mathrm{tr}(S) - \sum_{j=1}^r \theta^2 \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} \equiv \delta$. Then,

$$\delta = \theta \operatorname{tr}(S) - \theta \sum_{j=1}^{r} \theta \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j}$$

$$= \theta \left[\operatorname{tr}(S) - \sum_{j=1}^{p} \theta \frac{\Lambda'_j S \Lambda_j}{1 + \theta \Lambda'_j \Lambda_j} \right]$$

$$= \theta \left[\operatorname{tr}(S) - \sum_{j=1}^{p} n \Lambda'_j \Lambda_j \right] \text{ from (11.4.19)}$$

$$= np \text{ from (11.4.22).}$$

Hence the result.

Example 11.4.1. Tests are conducted to evaluate x_1 : verbal-linguistic skills, x_2 : spatial visualization ability, and x_3 : mathematical abilities. Test scores on x_1, x_2, and x_3 are available. It is known that these abilities are governed by two intellectual faculties that will be identified as f_1 and f_2, and that linear functions of f_1 and f_2 are contributing to x_1, x_2, x_3. These coefficients in the linear functions, known as factor loadings, are unknown. Let $\Lambda = (\lambda_{ij})$ be the matrix of factor loadings. Then, we have the model

$$x_1 = \lambda_{11} f_1 + \lambda_{12} f_2 + \mu_1 + e_1$$
$$x_2 = \lambda_{21} f_1 + \lambda_{22} f_2 + \mu_2 + e_2$$
$$x_3 = \lambda_{31} f_1 + \lambda_{32} f_2 + \mu_3 + e_3.$$

Let

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad M = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \text{ and } F = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix},$$

where M is some general effect, ϵ is the error vector or the sum total of the contributions from unknown factors, F represents the vector of contributing factors and Λ, the levels of the contributions. Let $\operatorname{Cov}(X) = \Sigma$, $\operatorname{Cov}(F) = \Phi$ and $\operatorname{Cov}(\epsilon) = \Psi$. Under the assumptions $\Phi = I$, $\Psi = \sigma^2 I$ where σ^2 is a real scalar quantity, and I is the identity matrix, and $\Lambda' \Psi^{-1} \Lambda$ is diagonal, estimate the factor loadings λ_{ij}'s and σ^2 in Ψ. A battery of tests are conducted on a random sample of six individuals and the following are the data, where our notations are **X**: the matrix of sample values, $\bar{X}$: the sample average, $\bar{\mathbf{X}}$: the matrix of sample averages, and S: the sample sum of products matrix. So, letting

$$\mathbf{X} = [X_1, X_2, \ldots, X_6] = \begin{bmatrix} 4 & 2 & 4 & 2 & 4 & 2 \\ 4 & 2 & 1 & 2 & 1 & 2 \\ 2 & 5 & 3 & 3 & 3 & 2 \end{bmatrix},$$

estimate the factor loadings λ_{ij}'s and the variance σ^2.

Solution 11.4.1. We begin with the computations of the various quantities required to arrive at a solution. Observe that since the matrix of factor loadings Λ is 3×2 and a random sample of 6 observation vectors is available, $n = 6$, $p = 3$ and $r = 2$ in our notation.

In this case,

$$\bar{X} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, \quad \bar{\mathbf{X}} = [\bar{X}, \bar{X}, \dots, \bar{X}],$$

$$\mathbf{X} - \bar{\mathbf{X}} = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 \\ 2 & 0 & -1 & 0 & -1 & 0 \\ -1 & 2 & 0 & 0 & 0 & -1 \end{bmatrix},$$

$$[\mathbf{X} - \bar{\mathbf{X}}][\mathbf{X} - \bar{\mathbf{X}}]' = S = \begin{bmatrix} 6 & 0 & -2 \\ 0 & 6 & -2 \\ -2 & -2 & 6 \end{bmatrix}.$$

An estimator/estimate of Σ is $\hat{\Sigma} = \frac{S}{n}$ whereas an unbiased estimator/estimate of Σ is $\frac{S}{n-1}$. An eigenvalue of $\frac{S}{\alpha}$ is $\frac{1}{\alpha}$ times the corresponding eigenvalue of S. Moreover, constant multiples of eigenvectors are also eigenvectors for a given eigenvalue. Accordingly, we will work with S instead of $\hat{\Sigma}$ or an unbiased estimate of Σ. Since

$$\begin{vmatrix} 6 - \lambda & 0 & -2 \\ 0 & 6 - \lambda & -2 \\ -2 & -2 & 6 - \lambda \end{vmatrix} = 0 \Rightarrow (6 - \lambda)(\lambda^2 - 12\lambda + 28) = 0,$$

the eigenvalues are $\lambda_1 = 6 + \sqrt{8}$, $\lambda_2 = 6$, $\lambda_3 = 6 - \sqrt{8}$, the two largest ones being $\lambda_1 = 6 + \sqrt{8}$ and $\lambda_2 = 6$. Let us evaluate the eigenvectors U_1, U_2 and U_3 associated with these three eigenvalues. An eigenvector U_1 corresponding to $\lambda_1 = 6 + \sqrt{8}$ will be a solution of the equation

$$\begin{bmatrix} -\sqrt{8} & 0 & -2 \\ 0 & -\sqrt{8} & -2 \\ -2 & -2 & -\sqrt{8} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow U_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 6$, the equation to be solved is

$$\begin{bmatrix} 0 & 0 & -2 \\ 0 & 0 & -2 \\ -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow U_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

As for the eigenvalue $\lambda_3 = 6 - \sqrt{8}$, it is seen from the derivation of U_1 that $U_3' = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1]$. Let us now examine the denominator of $\hat{\theta}$ in (11.4.22). Observe that $U_1'U_1 = 2$, $U_2'U_2 = 2$, $U_3'U_3 = 2$ and $n\sum_{j=1}^{p} U_j'U_j = 6(2+2+2) = 36$. However, $\text{tr}(S) = (6 + \sqrt{8}) + (6) + (6 - \sqrt{8}) = 18$. So, let us multiply each vector by $\frac{1}{\sqrt{3}}$ so that $n(\sum_{j=1}^{p} U_j'U_j = 6(\frac{2}{3} + \frac{2}{3} + \frac{2}{3}) = 12$ and $\text{tr}(S) - n\sum_{j=1}^{p} U_j'U_j = 18 - 12 > 0$. Thus, the estimate of θ is given by

$$\hat{\theta} = \frac{np}{\text{tr}(S) - n\sum_{j=1}^{p} U_j'U_j} = \frac{(6)(3)}{18 - 12} = 3 \Rightarrow \hat{\sigma}^2 = \frac{1}{3}$$

In light of (11.4.20), the factor loadings are estimated by U_1 and U_2 scaled by $\frac{1}{\sqrt{3}}$. Hence, the estimates of the factor loadings, denoted with a hat, are the following: $\hat{\lambda}_{11} = (\frac{1}{\sqrt{3}})(-\frac{1}{\sqrt{2}}) = -\frac{1}{\sqrt{6}}$, $\hat{\lambda}_{21} = (\frac{1}{\sqrt{3}})(-\frac{1}{\sqrt{2}}) = -\frac{1}{\sqrt{6}}$, $\hat{\lambda}_{31} = (\frac{1}{\sqrt{3}})(1) = \frac{1}{\sqrt{3}}$, $\hat{\lambda}_{12} = (\frac{1}{\sqrt{3}})(1) = \frac{1}{\sqrt{3}}$, $\hat{\lambda}_{22} = (\frac{1}{\sqrt{3}})(-1) = -\frac{1}{\sqrt{3}}$, $\hat{\lambda}_{32} = 0$.

11.5. General Case

Let

$$\delta_j = \begin{bmatrix} \theta_1\lambda_{1j} \\ \theta_2\lambda_{2j} \\ \vdots \\ \theta_p\lambda_{pj} \end{bmatrix}, \quad \delta_j'\delta_j = \Lambda_j'\Theta^2\Lambda_j, \quad \Lambda_j = \begin{bmatrix} \lambda_{1j} \\ \lambda_{2j} \\ \vdots \\ \lambda_{pj} \end{bmatrix},$$

$$\Psi^{-1} = \Theta^2 = \begin{bmatrix} \theta_1^2 & 0 & \cdots & 0 \\ 0 & \theta_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta_p^2 \end{bmatrix} \quad \text{and} \quad |I + \Lambda'\Theta^2\Lambda| = \prod_{j=1}^{r}(1 + \delta_j'\delta_j).$$

We will take δ_j, $j = 1, \ldots, r$, and $\Theta = \text{diag}(\theta_1, \theta_2, \ldots, \theta_p)$ as the parameters. Expressed in terms of the δ_j's and Θ, the log-likelihood function is the following:

$$\ln L = -\frac{np}{2}\ln(2\pi) + n\sum_{j=1}^{p}\ln\theta_j - \frac{n}{2}\sum_{j=1}^{r}\ln(1 + \delta_j'\delta_j)$$

$$- \frac{1}{2}\text{tr}(\Theta^2 S) + \frac{1}{2}\sum_{j=1}^{r}\frac{1}{1 + \delta_j'\delta_j}(\delta_j'\Theta S\Theta\,\delta_j). \tag{11.5.1}$$

Let us take $\delta_j,\ j = 1, \ldots, r$, and Θ as the parameters. Differentiating $\ln L$ partially with respect to the vector δ_j, for a specific j, and equating the result to a null vector, we have the following (referring to Chap. 1 for vector/matrix derivatives):

$$-\frac{n}{2}\frac{2\delta_j}{1+\delta_j'\delta_j} - \frac{1}{2}(\delta_j'\Theta S\Theta\,\delta_j)\frac{2\delta_j}{(1+\delta_j'\delta_j)^2} + \frac{1}{2}\frac{2\,\Theta S\Theta\,\delta_j}{1+\delta_j'\delta_j} = O, \tag{i}$$

which multiplied by $1 + \delta_j'\delta_j > 0$, yields

$$-n\delta_j - \frac{(\delta_j'\Theta S\Theta\delta_j)}{1+\delta_j'\delta_j}\delta_j + (\Theta S\Theta)\delta_j = O. \tag{11.5.2}$$

On premultiplying (11.5.2) by δ_j' and then by $1 + \delta_j'\delta_j$, and simplifying, we then have

$$-n(1 + \delta_j'\delta_j)(\delta_j'\delta_j) + (\delta_j'\Theta S\Theta\delta_j) = O \Rightarrow$$
$$\frac{\delta_j'\Theta S\Theta\delta_j}{1+\delta_j'\delta_j} = n\delta_j'\delta_j. \tag{11.5.3}$$

Let us differentiate $\ln L$ as given in (11.5.1) partially with respect to θ_j for a specific j such as $j = 1$. Then,

$$\frac{n}{\theta_j} - \frac{1}{2}2\theta_j s_{jj} + \frac{1}{2}\sum_{j=1}^{r}\frac{1}{1+\delta_j'\delta_j}\frac{\partial}{\partial\theta_j}(\delta_j'\Theta S\Theta\delta_j),$$

where

$$\frac{\partial}{\partial\theta_j}(\delta_j'\Theta S\Theta\delta_j) = \delta_j'\Big[\frac{\partial}{\partial\theta_j}\Theta\Big]S\Theta\delta_j + \delta_j'\Theta S\Big[\frac{\partial}{\partial\theta_j}\Theta\Big]\delta_j,$$

$$\frac{\partial}{\partial\theta_1}\Theta = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix} \Rightarrow \theta_1\frac{\partial}{\partial\theta_1}\Theta = \begin{bmatrix} \theta_1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0, \end{bmatrix},$$

so that

$$\Big[\theta_1\frac{\partial}{\partial\theta_1} + \cdots + \theta_p\frac{\partial}{\partial\theta_p}\Big]\Theta = \Theta. \tag{ii}$$

Hence,

$$\Big[\sum_{j=1}^{p}\theta_j\frac{\partial}{\partial\theta_j}\Big]\sum_{j=1}^{r}\frac{\delta_j'\Theta S\Theta\delta_j}{1+\delta_j'\delta_j} = 2\sum_{j=1}^{r}\frac{\delta_j'\Theta S\Theta\delta_j}{1+\delta_j'\delta_j},$$

and then,

$$\left[\sum_{j=1}^{p}\theta_j\frac{\partial}{\partial\theta_j}\right]L=0 \Rightarrow$$

$$np-\sum_{j=1}^{p}\theta_j^2 s_{jj}+\sum_{j=1}^{r}\frac{\delta_j'\Theta S\Theta\delta_j}{1+\delta_j'\delta_j}=0. \tag{11.5.4}$$

However, given (11.5.3), we have $n\delta_j'\delta_j = \delta_j'\Theta S\Theta\delta_j/(1+\delta_j'\delta_j)$, $j=1,\ldots,r$, and therefore (11.5.4) can be expressed as

$$np-\sum_{j=1}^{p}\theta_j^2 s_{jj}+\sum_{j=1}^{r}n\delta_j'\delta_j=0. \tag{iii}$$

Letting

$$c=\frac{1}{p}\sum_{j=1}^{r}\delta_j'\delta_j, \tag{11.5.5}$$

equation *(iii)* can be written as

$$\sum_{j=1}^{p}[n(1+c)-\theta_j^2 s_{jj}]=0, \quad c=0, \; j=r+1,\ldots,p,$$

so that a solution for θ_j is

$$\hat{\theta}_j^2=\frac{n(1+c)}{s_{jj}} \quad \text{or} \quad \hat{\sigma}_j^2=\frac{s_{jj}}{n(1+c)}, \tag{11.5.6}$$

with the proviso that $c=0$ for $\hat{\theta}_j^2$ and $\hat{\sigma}_j^2$, $j=r+1,\ldots,p$. Then, an estimate of $\Theta S\Theta$ is given by

$$\hat{\Theta} S \hat{\Theta}=n(1+c)\begin{bmatrix}\frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}}\end{bmatrix}S\begin{bmatrix}\frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}}\end{bmatrix}$$

$$=n(1+c)R \tag{11.5.7}$$

where R is the sample correlation matrix, and on applying the identities (11.5.3) and (11.5.7), (11.5.2) becomes

$$-n\delta_j - n(\delta_j' \delta_j)\delta_j + (n(1+c)R)\delta_j = 0 \Rightarrow$$

$$\left[R - \frac{1 + \hat{\delta}_j' \hat{\delta}_j}{1+c} I \right] \delta_j = O. \tag{11.5.8}$$

This shows that δ_j is an eigenvector of R. If ν_j is an eigenvalue of R, then the largest r eigenvalues are of the form

$$\nu_j = \frac{1 + \hat{\delta}_j' \hat{\delta}_j}{1+c}, \ \ j = 1, \dots, r, \tag{11.5.9}$$

and the remaining ones are $\nu_{r+1}, \dots, \nu_p$, where c is as specified in (11.5.5). Thus, the procedure is the following: Compute the eigenvalues ν_j, $j = 1, \dots, p$, of R and determine the corresponding eigenvectors, denoted by δ_j, $j = 1, \dots, p$. The first r of them which correspond to the r largest ν_j's, are $\hat{\delta}_j = \hat{\Theta} \hat{\Lambda}_j \Rightarrow \hat{\Lambda}_j = \hat{\Theta}^{-1}\hat{\delta}_j$, $j = 1, \dots, r$. Let

$$\hat{\delta}_j = \begin{bmatrix} \hat{\delta}_{1j} \\ \hat{\delta}_{2j} \\ \vdots \\ \hat{\delta}_{pj} \end{bmatrix}, \ \hat{\Lambda}_j = \begin{bmatrix} \hat{\lambda}_{1j} \\ \hat{\lambda}_{2j} \\ \vdots \\ \hat{\lambda}_{pj} \end{bmatrix} \text{ and } \hat{\Theta}^2 = \begin{bmatrix} \frac{n(1+c)}{s_{11}} & 0 & \cdots & 0 \\ 0 & \frac{n(1+c)}{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n(1+c)}{s_{pp}} \end{bmatrix}. \tag{11.5.10}$$

Then,

$$\hat{\lambda}_{ij} = \frac{\sqrt{s_{jj}}}{\sqrt{n(1+c)}} \hat{\delta}_{ij}, \ i = 1, \dots, p, \ j = 1, \dots, r, \tag{11.5.11}$$

and $\hat{\Theta}$ is available from (11.5.10). All the model parameters have now been estimated.

11.5.1. The exponent in the likelihood function

Given the MLE's of the parameters which are available from (11.5.6), (11.5.8), (11.5.10) and (11.5.11), what will be the maximum value of the likelihood function? Let us examine its exponent:

$$-\frac{1}{2}\text{tr}(\hat{\Theta}^2 S) + \frac{n}{2} \sum_{j=1}^{r} \hat{\delta}_j' \hat{\delta}_j = -\frac{1}{2}n(1+c)\text{tr}(R) + \frac{n}{2}pc$$

$$= -\frac{1}{2}n(1+c)p + \frac{1}{2}npc = -\frac{np}{2},$$

that is, the same value of the exponent that is obtained under a general Σ. The estimates were derived under the assumptions that $\Sigma = \Psi + \Lambda\Lambda'$, $\Lambda'\Psi^{-1}\Lambda$ is diagonal with the diagonal elements $\delta'_j\delta_j$, $j = 1, \ldots, r$, and $\Psi = \Theta^{-2}$.

Example 11.5.1. Using the data set provided in Example 11.4.1, estimate the factor loadings and the diagonal elements of $\text{Cov}(\epsilon) = \Psi = \text{diag}(\psi_{11}, \ldots, \psi_{pp})$. In this example, $p = 3$, $n = 6$, $r = 2$.

Solution 11.5.1. We will adopt the same notations and make use of some of the computational results already obtained in the previous solution. First, we need to compute the eigenvalues of the sample correlation matrix R. The sample sum of products matrix S is given by

$$S = \begin{bmatrix} 6 & 0 & -2 \\ 0 & 6 & -2 \\ -2 & -2 & 6 \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & 0 & -\frac{2}{6} \\ 0 & 1 & -\frac{2}{6} \\ -\frac{2}{6} & -\frac{2}{6} & 1 \end{bmatrix} = \frac{1}{6}S.$$

Hence, the eigenvalues of R are $\frac{1}{6}$ times the eigenvalues of S, that is, $\nu_1 = \frac{1}{6}(6 + \sqrt{8}) = 1 + \frac{\sqrt{2}}{3}$, $\nu_2 = \frac{1}{6}(6) = 1$ and $\nu_3 = 1 - \frac{\sqrt{2}}{3}$. Since $\frac{1}{6}$ will be canceled when determining the eigenvectors, the eigenvectors of S will coincide with those of R. They are the following, denoted again by δ_j, $j = 1, 2, 3$:

$$\delta_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}, \quad \delta_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \delta_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 1 \end{bmatrix}.$$

Therefore, $\delta'_1\delta_1 = 2$, $\delta'_2\delta_2 = 2$, $\delta'_3\delta_3 = 2$ and c as defined in (11.5.5) is $c = \frac{1}{3}(2 + 2) = \frac{4}{3} \Rightarrow n(1 + c) = 6(1 + \frac{4}{3}) = 14$. Then, in light of (11.5.10), the estimates of ψ_{jj}, $j = 1, 2, 3$, are available as $\hat{\psi}_{jj} = \hat{\theta}_j^{-2} = \frac{s_{jj}}{n(1+c)}$ or $\hat{\psi}_{11} = \hat{\theta}_1^{-2} = \frac{6}{14} = \frac{3}{7} = \hat{\psi}_{22} = \hat{\psi}_{33}$, denoting the estimates with a hat. Therefore, the diagonal matrix $\hat{\Psi} = \text{diag}(\frac{3}{7}, \frac{3}{7}, \frac{3}{7})$. Hence, the matrix $\Theta^{-1} = \Psi^{\frac{1}{2}}$ is estimated by $\hat{\Theta}^{-1} = \text{diag}(\frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}})$. From (11.5.11), $\hat{\Lambda}_j = \hat{\Theta}^{-1}\delta_j = \text{diag}(\frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}})\delta_j$, that is, δ_j is pre-multiplied by $\frac{\sqrt{3}}{\sqrt{7}}$. Therefore, the estimates of the factor loadings are: $\hat{\lambda}_{11} = -(\frac{\sqrt{3}}{\sqrt{7}})(\frac{1}{\sqrt{2}}) = \hat{\lambda}_{21} = -\hat{\lambda}_{13} = -\hat{\lambda}_{23}$, $\hat{\lambda}_{31} = \frac{\sqrt{3}}{\sqrt{7}} = \hat{\lambda}_{33} = \hat{\lambda}_{12} = -\hat{\lambda}_{22}$, and $\hat{\lambda}_{32} = 0$.

11.6. Tests of Hypotheses

The usual test in connection with the current topic consists of assessing identifiability, that is, testing the hypothesis H_0 that the population covariance matrix $\Sigma > O$ can be represented as $\Sigma = \Lambda \Phi \Lambda' + \Psi$ when $\Phi = I$, $\Lambda' \Psi^{-1} \Lambda$ is a diagonal matrix with positive diagonal elements, $\Psi > O$ is a diagonal matrix and $\Lambda = (\lambda_{ij})$ is a $p \times r$, $r \leq p$, matrix of full rank r, whose elements are the factor loadings. That is,

$$H_0 : \ \Sigma = \Lambda \Lambda' + \Psi. \tag{11.6.1}$$

In this instance, a crucial aspect of the hypothesis H_o consisting of determining whether "the model fits", is that the number r be designated since the other quantities n, the sample size, and p, the order of the observation vector, are preassigned. Thus, the phrase "model fits" means that for a given r, Σ can be expressed in the form $\Sigma = \Psi + \Lambda \Lambda'$, in addition to satisfying the identification conditions. The assumed model has the representation: $X = M + \Lambda F + \epsilon$ where $X' = (x_1, \ldots, x_p)$ stands for the $p \times 1$ vector of observed scores on p tests or p batteries of tests, M is a $p \times 1$ vector of general effect, F is an $r \times 1$ vector of unknown factors, $\Lambda = (\lambda_{ij})$ is the unknown $p \times r$ matrix of factor loadings and ϵ is the $p \times 1$ error vector. When ϵ and F are uncorrelated, the covariance matrix of X is given by

$$\Sigma = \Lambda \Phi \Lambda' + \Psi$$

where $\Phi = \text{Cov}(F) > O$ and $\Psi = \text{Cov}(\epsilon) > O$ with Φ being $r \times r$ and Ψ being $p \times p$ and diagonal. A simple random sample from X will be taken to mean a sample of independently and identically distributed (iid) $p \times 1$ vectors $X'_j = (x_{1j}, x_{2j}, \ldots, x_{pj})$, $j = 1, \ldots, n$, with n denoting the sample size. The sample sum of products matrix or "corrected" sample sum of squares and cross products matrix is $S = (s_{ij})$, $s_{ij} = \sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$, where, for example, the average of the x_i's comprising the i-th row of $\mathbf{X} = [X_1, \ldots, X_n]$, namely, $\bar{x}_i$, is $\bar{x}_i = \sum_{k=1}^{n} x_{ik}/n$. If ϵ and F are independently normally distributed, then the likelihood ratio criterion or λ-criterion is

$$\lambda = \frac{\max_{H_0} L}{\max L} = \frac{|\hat{\Sigma}|^{\frac{n}{2}}}{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|^{\frac{n}{2}}} \ \Rightarrow \ w = \lambda^{\frac{2}{n}} = \frac{|\hat{\Sigma}|}{|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|} \tag{11.6.2}$$

where $\hat{\Sigma} = \frac{1}{n} S$ and the covariance matrix $\Sigma = \Lambda \Lambda' + \Psi$ under H_0, with $\Phi = \text{Cov}(F)$ assumed to be an identity matrix and the $r \times r$ matrix $\Lambda' \Psi^{-1} \Lambda = \text{diag}(\delta'_1 \delta_1, \ldots, \delta'_r \delta_r)$ having positive diagonal elements $\delta'_j \delta_j$, $j = 1, \ldots, r$. Referring to Sect. 11.4.2, we have

$$|\Lambda \Lambda' + \Psi| = |\Psi| \, |\Lambda' \Psi^{-1} \Lambda + I| \tag{11.6.3}$$

and $1 + \delta_j' \delta_j = 1 + \Lambda_j' \Psi^{-1} \Lambda_j = 1 + \Lambda_j' \Theta^2 \Lambda_j$ where $\delta_j = \Psi^{-\frac{1}{2}} \Lambda_j = \Theta \Lambda_j$ and Λ_j is the j-th column of Λ for $j = 1, \ldots, r$. It was shown in (11.5.8) that δ_j is an eigenvector of the sample correlation matrix R and

$$\prod_{j=1}^{r} (1 + \delta_j' \delta_j) = |\Lambda' \Psi^{-1} \Lambda + I|.$$

However, in view of the discussion following (11.5.8), an eigenvalue of R is of the form $v_j = \frac{1 + \delta_j' \delta_j}{(1+c)}$, $j = 1, \ldots, r$. Let $v_1, \ldots v_p$ be the eigenvalues of R and let the largest r of them be $v_1, \ldots, v_r$. It also follows from (11.5.8) that $\hat{\Theta}^2 = n(1+c)\mathrm{diag}(\frac{1}{s_{11}}, \ldots, \frac{1}{s_{pp}})$ with $\Theta = \Psi^{-\frac{1}{2}}$. Thus,

$$\frac{|\hat{\Sigma}|}{|\hat{\Psi}|} = \left| \hat{\Theta} \left(\frac{S}{n} \right) \hat{\Theta} \right| = |(1+c)R| = \left\{ \prod_{j=1}^{r} (1+c)v_j \right\} (1+0)^{p-r} v_{r+1} \cdots v_p$$

$$= \left\{ \prod_{j=1}^{r} (1 + \hat{\delta}_j' \hat{\delta}_j) \right\} v_{r+1} \cdots v_p$$

$$\Rightarrow \frac{|\frac{1}{n} S|}{|\hat{\Psi}| \, |\hat{\Lambda}' \hat{\Psi}^{-1} \hat{\Lambda} + I|} = \frac{\{\prod_{j=1}^{r} (1 + \hat{\delta}_j' \hat{\delta}_j)\} v_{r+1} \cdots v_p}{\{\prod_{j=1}^{r} (1 + \hat{\delta}_j' \hat{\delta}_j)\}}$$

$$= v_{r+1} \cdots v_p = w = \lambda^{\frac{2}{n}}. \tag{11.6.4}$$

Hence, we reject the null hypothesis for small values of the product $v_{r+1} \cdots v_p$, that is, the product of the smallest $p - r$ eigenvalues of the sample correlation matrix R. In order to evaluate critical points, one would require the null distribution of the product of the eigenvalues, $v_{r+1} \cdots v_p$, which is difficult to determine for a general p. How can rejecting the null hypothesis that the "model fits" be interpreted? Since, in the whole structure, the decisive quantity is r, we are actually rejecting the hypothesis that a given r is the number of main factors contributing to the observations. Hence, we may seek a larger or smaller r, keeping the structure unchanged and testing the same hypothesis again until the hypothesis is not rejected. We may then assume that the r specified at the last stage is the number of main factors contributing to the observation or we may assert that, with that particular r, there is evidence that the model fits.

We will now determine conditions ensuring that the likelihood ratio criterion λ be less than or equal to one. While, assuming that $\Lambda' \Psi^{-1} \Lambda$ is diagonal, the left-hand side of the deciding equation, $\Sigma = \Lambda \Lambda' + \Psi$, has $p(p+1)/2$ parameters, there are $pr + p - r(r-1)/2$

conditions on the right-hand side where $r(r-1)/2$ arises from the diagonality condition. The difference is then

$$\frac{p(p+1)}{2} - \left[pr + p - \frac{r(r-1)}{2}\right] = \frac{1}{2}[(p-r)^2 - (p+r)] \equiv \rho. \qquad (11.6.5)$$

This ρ depends upon the parameters p and r, whereas λ depends upon p, r and c. Thus, λ may not be ≤ 1. In order to make $\lambda \le 1$, we can make c close to 0 by multiplying the $\hat{\delta}_j$'s by a constant, observing that this is always possible because the $\hat{\delta}_j$'s are the eigenvectors of R. By selecting a constant m and taking the new $\hat{\delta}_j$ as $\frac{1}{\sqrt{m}}\delta_j$, c can be made close to 0 and λ will be ≤ 1, so that rejecting the null hypothesis for small values of λ will make sense. It may so happen that there will not be any parameter left to be restricted by the hypothesis H_o that "model fits". The quantity ρ appearing in (11.6.5) could then be ≤ 0, and in such an instance, the hypothesis would not make sense and could not be tested.

The density of the sample correlation matrix R is provided in Example 1.25 of Mathai (1997, p. 58). Denoting this density by $f(R)$, it is the following for the population covariance matrix Σ in a parent $N_p(\mu, \Sigma)$ population with Σ being a positive definite diagonal matrix, as was assumed in Sect. 11.6:

$$f(R) = \frac{[\Gamma(\frac{m}{2})]^p}{\Gamma_p(\frac{m}{2})}|R|^{\frac{m}{2}-\frac{p+1}{2}}, \ R > O, \ m = n-1, \ n > p, \qquad (11.6.6)$$

and zero elsewhere, where n is the sample size.

11.6.1. Asymptotic distribution of the likelihood ratio statistic

For a large sample size n, $-2\ln\lambda$ is approximately distributed as a chisquare random variable having k degrees of freedom where λ is the likelihood ratio criterion and k is the number of parameters restricted by the hypothesis H_0. This approximation holds whenever the sample size n is large and $k \ge 1$. With ρ as defined in (11.6.5), we have

$$k = \rho = \frac{1}{2}[(p-r)^2 - (p+r)]. \qquad (11.6.7)$$

However, $p-r = 1$ and $p+r = 5$ in the illustrative example, so that $k = -2$. Accordingly, even if the sample size n were large, this asymptotic result would not be applicable.

11.6.2. How to decide on the number r of main factors?

The structure of the population covariance matrix Σ under the model $X_j = M + \Lambda F + E_j$, $j = 1, \ldots, n$, is

$$\Sigma = \Lambda\Phi\Lambda' + \Psi \ \Rightarrow \ \Sigma = \Lambda\Lambda' + \Psi \ \text{for} \ \Phi = I, \qquad (11.6.8)$$

where it is assumed that E_j and F are uncorrelated, $\Sigma = \mathrm{Cov}(X_j) > O$ is $p \times p$, $\Phi = \mathrm{Cov}(F) = I$, the $r \times r$ identity matrix, $\Psi = \mathrm{Cov}(E_j)$ is a $p \times p$ diagonal matrix and $\Lambda = (\lambda_{ij})$ is a full rank $p \times r$, $r \leq p$, matrix whose elements are the factor loadings. Under the orthogonal factor model, $\Phi = I$. Moreover, to ensure the identification of the model, we assume that $\Lambda'\Psi^{-1}\Lambda$ is a diagonal matrix. Before initiating any data analysis, we have to assign a value to r on the basis of the data set at hand in order to set up the model. Thus, the matter of initially setting the number of main factors has to be addressed. Given

$$R = \Lambda\Lambda' + \Psi \qquad (11.6.9)$$

where $\Lambda = (\lambda_{ij})$ is a $p \times r$ matrix and Ψ is a $p \times p$ diagonal matrix, does a solution that is expressible in terms of the elements of R (or those of S if S is used), exist for all λ_{ij}'s and ψ_{jj}'s? In general

$$R = \lambda_1 U_1 U_1' + \cdots + \lambda_p U_p U_p' \qquad (11.6.10)$$

where the λ_j's are the eigenvalues of R and the U_j's are the corresponding normalized eigenvectors. Observe that $U_j U_j'$ is $p \times p$ whereas $U_j' U_j = 1$, $j = 1, \ldots, p$. If $r = p$, then a solution always exists for (11.6.9). When taking $\Psi = O$, we can always let $R = BB'$ for some $p \times p$ matrix B, which can be achieved for example via a triangular decomposition. Accordingly, the relevant aspects are $r < p$ and the diagonal elements in Ψ, namely, the ψ_{jj}'s being positive. Can we then solve for all the λ_{ij}'s and ψ_{jj}'s involved in (11.6.9) in terms of the elements in R? The answer is that a solution exists, but only when certain conditions are satisfied. Our objective is to select a value of r that is as small as possible and then, to obtain a solution to (11.6.9) in terms of the elements in R.

The analysis is to be carried out by utilizing either the sample sum of products matrix S or the sample correlation matrix R. The following are some of the guidelines for selecting r in order to set up the model.

(i): Compute all the eigenvalues of R (or S). Let r be the number of eigenvalues ≥ 1 if the sample correlation matrix R is used. If S is used, then determine all the eigenvalues, calculate the average of these eigenvalues, and count the number of eigenvalues that are greater than or equal to this average. Take that number to be r.

(2): Carry out a Principal Component Analysis on R (or S). If S is used, ensure that the units of measurements are not creating discrepancies. Compute the variances of these Principal Components, which are the eigenvalues of R (or S). Let λ_j, $j = 1, \ldots, p$, denote these eigenvalues. Compute the ratios

$$\frac{\lambda_1 + \cdots + \lambda_m}{\lambda_1 + \cdots + \lambda_p}, \quad m = 1, 2, \ldots,$$

and stop with that m for which the desired fraction of the total variation in the data is accounted for. Take that m as r. When implementing the principal component approach, the factor loadings λ_{ij}'s and the ψ_{jj}'s can be estimated as follows: From (11.6.10), write

$$R = A + B \text{ with } A = \sum_{j=1}^{r} \lambda_j U_j U'_j \text{ and } B = \sum_{j=r+1}^{p} \lambda_j U_j U'_j,$$

where A can be expressed as VV' with $V = [\sqrt{\lambda_j} U_1, \ldots, \sqrt{\lambda_j} U_r]$. Then, V is taken as an approximate estimate of Λ or as $\hat{\Lambda}$. Observe that $\lambda_j > 0, \ j = 1, \ldots, p$. The sum over j of the i-th diagonal elements of $\lambda_j U_j U'_j, \ j = r+1, \ldots, p$, will provide an estimate for $\psi_{ii}, \ i = 1, \ldots, p$. These estimates can also be obtained as follows: Consider the estimate of σ_{ii} denoted by $\hat{\sigma}_{ii}$ which is equal to the sum of all the i-th diagonal elements in $A + B$; it will be 1 if R is used and $\hat{\sigma}_{ii}$ if S is utilized in the analysis; then, $\hat{\psi}_{ii} = \hat{\sigma}_{ii} - \sum_{j=1}^{r} \hat{\lambda}_{ij}$ and $\hat{\psi}_{ii}$ is now the sum of the i-th diagonal elements in B.

(iii): Consider the individual correlations in the sample correlation matrix R. Identify the largest ones in absolute value. If the largest ones occur at the (1,3)-th and (2,3)-th positions, then the factor f_3 will be deemed influential. Start with $r = 1$ (factor f_3) and carry out the analysis. Then, assess the proportion of the total variation accounted for by $\hat{\sigma}_{33}$. Should the proportion not be satisfactory, we may continue with $r = 2$. If the (2,3)-th position value is larger in absolute value than the value at the (1,3)-th position, then f_2 may be the next significant factor. Compute $\hat{\sigma}_{33} + \hat{\sigma}_{22}$ and determine the proportion to the total variation. If the resulting model is rejected, then take $r = 3$, and continue in this fashion until an acceptable proportion of the total variation is accounted for.

(iv): The maximum likelihood method. With this approach, we begin with a preselected r and test the hypothesis that, when comprising r factors, the model fits. If the hypothesis is rejected, then we let number of influential factors be $r-1$ or $r+1$ and continue the process of testing and deciding until the hypothesis is not rejected. That final r is to be taken as the number of main factors contributing towards the observations. The initial value of r may be determined by employing one of the methods described in *(i)* or *(ii)* or *(iii)*.

Exercises

11.1. For the following data, where the 6 columns of the matrix represent the 6 observation vectors, verify whether $r = 2$ provides a good fit to the data. The proposed model is the Factor Analysis model $X = M + \Lambda F + \epsilon, \ F' = (f_1, f_2), \ \Lambda$ is 3×2 and of rank 2, $\text{Cov}(\epsilon) = \Psi$ is diagonal, $\text{Cov}(F) = \Phi = I,$ and $\text{Cov}(X) = \Sigma > O$. The data set is

$$\begin{bmatrix} 0 & 1 & -1 & 0 & 1 & -1 \\ 1 & 1 & 0 & 2 & 2 & 0 \\ -1 & -1 & 0 & -1 & -1 & -2 \end{bmatrix}.$$

11.2. For the model $X = M + \Lambda F + \epsilon$ with the conditions as specified in Exercise 11.1, verify whether the model with $r = 2$ or $r = 3$ gives a good fit on the basis of the following data, where the columns in the matrix represent five observation vectors:

$$\begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ -1 & 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 0 \end{bmatrix}.$$

11.3. Do a Principal Component Analysis in Exercise 11.1 to assess what percentage of the total variation in the data is accounted for by $r = 2$.

11.4. Do a Principal Component Analysis in Exercise 11.2 to determine what percentages of the total variation in the data are accounted for by $r = 2$ and $r = 3$.

11.5. Even though the sample sizes are not large, perform tests based on the asymptotic chisquare to assess whether the two tests there agree with the findings in Exercises 11.1 and 11.2.

11.6. Four model identification conditions are stated at the end of Sect 11.3.1. Develop λ-criteria under the conditions stated in *(i)*: case (2); *(ii)*: case (3), selecting your own B_1; *(iii)*: case (4).

References

E. S. Allman, C. Matias and J. A. Rhodes (2009): Identifiability of parameters in latent structure models with many observed variables, *The Annals of Statistics*, **37**, 3099–3132

T. W. Anderson (2003): *An Introduction to Multivariate Statistical Analysis*, 3rd ed., Wiley-Interscience, New Jersey.

T. W. Anderson and H. Rubin (1956): Statistical inference in factor analysis, Proc. 3rd Berkeley Symposium on Mathematical Statistics and Probability, Vol. **5**, 111–150.

Y. Chen, X. Li, and S. Zhang (2020): Structured latent factor analysis for large-scale data: identifiability, estimability, and their implications, *Journal of the American Statistical Association*, **115**, 1756–1770.

M. Ihara and Y. Kano (1995): Identifiability of full, marginal, and conditional factor analysis models, *Statistics & Probability Letters*, **23**, 343–350.

A. M. Mathai (1997): *Jacobians of Matrix Transformations and Functions of Matrix Argument*, World Scientific Publishing, New York.

A.M. Mathai (2021): Factor analysis revisited, *C.R. Rao, 100th Birth Anniversary Volume, Statistical Theory and Practice*, https://doi.org/10.1007/s42519-020-00160-1

L. L. Wegge (1996): Local identifiability of the factor analysis and measurement error model parameter, *Journal of Econometrics*, **70**, 351–382.