Examensarbete

The unweighted mean estimator in a Growth Curve model

Emil Karlsson

LiTH - MAT - EX - - 2016 / 03 - - SE
The unweighted mean estimator in a Growth Curve model

Mathematical statistics, Linköpings universitet

Emil Karlsson

LiTH - MAT - EX - - 2016 / 03 - - SE

Examensarbete: 30 hp

Level: A

Supervisor: Martin Singull,
Mathematical statistics, Linköpings universitet

Examiner: Martin Singull,
Mathematical statistics, Linköpings universitet

Linköping: April 2016
Abstract

The field of statistics is becoming increasingly more important as the amount of data in the world grows. This thesis studies the Growth Curve model in multivariate statistics which is a model that is not widely used. One difference compared with the linear model is that the Maximum Likelihood Estimators are more complicated. That makes it more difficult to use and to interpret which may be a reason for its not so widespread use.

From this perspective this thesis will compare the traditional mean estimator for the Growth Curve model with the unweighted mean estimator. The unweighted mean estimator is simpler than the regular MLE. It will be proven that the unweighted estimator is in fact the MLE under certain conditions and examples when this occurs will be discussed. In a more general setting this thesis will present conditions when the unweighted estimator has a smaller covariance matrix than the MLEs and also present confidence intervals and hypothesis testing based on these inequalities.

Keywords: Growth Curve model, maximum likelihood, eigenvalue inequality, unweighted mean estimator, covariance matrix, circular symmetric Toeplitz, intraclass, generalized intraclass

URL for electronic version:

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-131043
Acknowledgements

First I would like to thank my supervisor Martin Singull for introducing me to the field of multivariate statistics. I would also like to thank him for all the support in my studies overall and in this thesis specifically. It has been a pleasure discussing and working with you. I want to thank my opponent Jonas Granholm for reading my work and giving valuable remarks.

I would like to give my deepest thanks to my Ghazaleh for walking through life together with me and supporting my endeavor to study. I love you.
Nomenclature

The reoccurring abbreviations and symbols are described here.

Notation

Throughout this thesis matrices and vectors will be denoted by boldface transcription. Matrices will for the most part be denoted by capital letters and vectors by small letters of Latin or Greek alphabets. Scalars and matrix elements will be denoted by ordinary letters of Latin or Greek alphabets. Random variables will be denoted by capital letters from the end of the Latin alphabet. The end of proofs are marked by □.

LIST OF NOTATION

\( \mathbf{A}_{m,n} \) - matrix of size \( m \times n \)
\( M_{m,n} \) - the set of all matrices of size \( m \times n \)
\( a_{ij} \) - matrix element of the \( i \)-th row and \( j \)-th column
\( \boldsymbol{a}_n \) - vector of size \( n \)
\( c \) - scalar
\( \mathbf{A}' \) - transposed matrix \( \mathbf{A} \)
\( \mathbf{I}_n \) - identity matrix of size \( n \)
\( |\mathbf{A}| \) - determinant of \( \mathbf{A} \)
\( \text{rank}(\mathbf{A}) \) - rank of \( \mathbf{A} \)
\( \text{tr}(\mathbf{A}) \) - trace of \( \mathbf{A} \)
\( \mathbf{X} \) - random matrix
\( \mathbf{x} \) - random vector
\( \mathbf{X} \) - random variable
\( \mathbb{E}[\mathbf{X}] \) - expectation
\( \text{var}(\mathbf{X}) \) - variance
\( \text{cov}(\mathbf{X}, \mathbf{Y}) \) - covariance of \( \mathbf{X} \) and \( \mathbf{Y} \)
\( \mathbf{S} \) - sample dispersion matrix
\( \mathcal{N}_p(\mu, \Sigma) \) - multivariate normal distribution
\( \mathcal{N}_{p,n}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Psi}) \) - matrix normal distribution
MLE - Maximum likelihood estimator
List of Figures

2.1 Plot over the Reciprocal Condition Number of $V$ . . . . . . . . . . . . 12

5.1 Histogram of the first anti-eigenvalue when $\Sigma$ is symmetric circular Toeplitz . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

5.2 Histogram of the first anti-eigenvalue when $\Sigma$ has no structure . . . 27
## Contents

1 Introduction ............................................. 1
   1.1 Chapter outline .................................... 2

2 Mathematical background ................................. 3
   2.1 Linear algebra ..................................... 3
       2.1.1 General definitions and theorems .............. 3
       2.1.2 Generalized inverse .......................... 4
       2.1.3 Anti-eigenvalues .............................. 5
       2.1.4 Kronecker product ......................... 6
       2.1.5 Matrix derivatives ......................... 6
   2.2 Statistics .......................................... 7
       2.2.1 General concepts ............................. 7
       2.2.2 The Maximum Likelihood Estimator ........... 8
       2.2.3 The matrix normal distribution ............... 8
       2.2.4 Growth Curve model ......................... 9
       2.2.5 Wishart distribution .......................... 11

3 Unweighted estimator for Growth Curve model under \( \mathcal{R}(\Sigma^{-1}B) = \mathcal{R}(B) \) ........ 13
   3.1 Unweighted vs weighted mean estimator ........... 13
   3.2 Previous results ................................... 14
   3.3 Maximum likelihood estimator ..................... 15
   3.4 Intraclass model .................................. 16
   3.5 Generalized intraclass model ..................... 17

4 Unweighted estimator for Growth Curve model without assumptions on \( B \) .......... 19
   4.1 Background and previous results ................. 19
   4.2 Eigenvalues and coefficients .................... 19
   4.3 Unweighted vs weighted estimator with known and fixed eigenvectors ....... 20
       4.3.1 Distributions of eigenvalues ............... 20
       4.3.2 Hypothesis testing and confidence interval .. 21
       4.3.3 Symmetric circular Toeplitz structured covariance matrix ........ 22

5 Simulations .............................................. 25
   5.1 Distribution of the first anti-eigenvalue .......... 25
       5.1.1 Symmetric circular Toeplitz covariance matrix .. 25
       5.1.2 Unknown covariance matrix .................. 26
   5.2 Eigenvalue confidence interval .................... 26
   5.3 Hypothesis testing ................................ 28

Karlsson, 2016. xiii
6 Further research 31
Chapter 1

Introduction

In this thesis the unweighted mean estimator for a Growth Curve model is studied. The Growth Curve model was introduced by [Potthoff and Roy, 1964] when the observation matrix is normally distributed with an unknown covariance matrix. The Growth Curve model belongs to the curved exponential family, since it is a generalized multivariate analysis of variance model (GMANOVA). The difference of the structure for the Growth Curve model in comparison to the ordinary multivariate linear model (MANOVA) is that the Growth Curve model is bilinear. There are two design matrices involved for a bilinear model, in contrary to the MANOVA, where there is one design matrix. For more details and references about the Growth Curve model see Chapter 2.

For the Growth Curve model, originally [Potthoff and Roy, 1964] derived a class of weighted estimators for the mean parameter matrix. These results were later extended by [Khatri, 1966], who showed that the standard Maximum Likelihood Estimators (MLEs) also belong to this class. Under some conditions, that will be studied in detail in Chapter 3, [Reinsel, 1982] have shown that the unweighted estimator for a Growth Curve model is the MLE. The proof was done by using that the unweighted estimator is the least square estimator and under these conditions the least square estimators coincide with the MLE.

The studies of patterned covariance matrices started with [Wilks, 1946], when Wilks published a paper dealing with measurements on $k$-equivalent psychological tests using a MANOVA model. He used the intraclass model, where the variance components are equal and the covariance between them are equal as well. Over the years different directions and special kinds of structures has been studied. A common one is the Toeplitz covariance matrix which arise in time series analysis, signal and image processing, Markov chains and among other fields. During the last years there has been some extra attention for the intraclass covariance structure for a Growth Curve model, originating from [Khatri, 1973]. E.g., [Zežula, 2006] derived some simple explicit estimators of the variance and the correlation given the intraclass covariance structure, [Ye and Wang, 2009] and [Klein and Zežula, 2010] developed estimators with the unbiased estimating equations using orthogonal complement. Recently [Srivastava and Singull, 2016a, Srivastava and Singull, 2016b] has studied the problem of testing sphericity and intraclass structure.
Chapter 1. Introduction

1.1 Chapter outline

Chapter 2: Introduces required concepts and results in mathematics, mainly in linear algebra and statistics, that are used in this thesis.

Chapter 3: Presents conditions when the unweighted mean estimator for Growth Curve model aligns with the MLE. The example cases of intraclass and generalized intraclass are studied in detail.

Chapter 4: Compares the unweighted mean estimator with the MLE for a Growth Curve model and presents a test to determine if it has better properties than the MLE for some observations. The case when the covariance matrix has a circular Toeplitz structure is studied in detail.

Chapter 5: Simulations based on the results from the previous chapters.

Chapter 6: Mentions improvements and directions of research which are interesting with regard to the area.
Chapter 2

Mathematical background

This chapter will introduce some of the mathematics that are needed to understand this thesis. Some of the material can be new to the reader and others not, but the aim of this chapter is to give a person studying a master program in mathematics enough background to understand this thesis. The theorems will be given without proofs but the proofs can be accessed through the referenced sources of each section.

2.1 Linear algebra

Since this thesis makes extensive use of linear algebra and matrices some common definitions and theorems will be introduced here.

2.1.1 General definitions and theorems

In the first part of this section follows some general definitions and theorems in real linear algebra. This section presents some notation and expressions that are required later. These results will be given without proofs but the proofs can be found in [Horn and Johnson, 2012][Kollo and von Rosen, 2011][Bernstein, 2009].

Definition 1. The set of all matrices with $m$ rows and $n$ columns is denoted as $M_{m,n}$ and similarly the set of all vectors with $m$ rows will be denoted as $M_{m}$.

Definition 2. A matrix $A \in M_{n,n}$ is called the identity matrix of size $n$, denoted with $I_n$, if the diagonal elements are 1 and the off-diagonal elements are 0.

Definition 3. A vector $1_n \in M_{n}$ is called the one-vector of size $n$, i.e.,

$$1_n = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}',$$

where $1'$ denotes the transpose of $1$.

Definition 4. A matrix $A \in M_{m,n}$ is called the zero-matrix, denoted as $0_{m,n}$, if all matrix elements of $A$ equals 0.

Definition 5. The range of a matrix $A \in M_{m,n}$, denoted as $\mathcal{R}(A)$, is defined by

$$\mathcal{R}(A) = \{Ax : x \in M_n\}.$$
Definition 6. The rank of a matrix $A \in M_{m,n}$, denoted as \text{rank}(A), is defined as the number of linearly independent columns (or rows) of the matrix.

Definition 7. A matrix $A \in M_{m,n}$ is called symmetric if $A' = A$.

Definition 8. A matrix $A \in M_{m,n}$ is called normal if $AA' = A'A$.

Definition 9. A matrix $A \in M_{m,n}$ is called orthogonal if $A'A = AA' = I_n$.

Definition 10. A symmetric square matrix $A \in M_{n,n}$ is positive (semi-)definite if $x'Ax \geq 0$ for any vector $x \neq 0$.

Definition 11. The values $\lambda_i$ that satisfy $Ax_i = \lambda_i x_i$, are called eigenvalues of the matrix $A$. The vector $x_i$, that corresponds to the eigenvalue $\lambda_i$, is called eigenvector of $A$ corresponding to $\lambda_i$.

And lastly two theorems regarding matrix decomposition, called the spectral theorem and the singular value decomposition theorem.

Theorem 1 (Spectral decomposition theorem). Any normal matrix $A \in M_{n,n}$ has an orthonormal basis of eigenvectors. In other words, any normal matrix $A$ can be represented as

$$A = UDU',$$

where $U \in M_{n,n}$ is an orthogonal matrix and $D \in M_{n,n}$ is a diagonal matrix with the eigenvalues of $A$ on the diagonal.

Here follows another decomposition theorem called the singular value decomposition theorem.

Theorem 2 (Singular value decomposition). Let $A \in M_{m,n}$, assume that $A$ is nonzero, let $r = \text{rank}(A)$, and define $B = \text{diag}([\sigma_1(A), \ldots, \sigma_r(A)])$, where $\sigma_i$ are the singular values of $A$. Then, there exist orthogonal matrices $S_1 \in M_{n,n}$ and $S_2 \in M_{m,m}$ such that

$$A = S_1 \begin{pmatrix} B & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{pmatrix} S_2.$$

Furthermore, each column of $S_1$ is an eigenvector of $AA'$, while each column of $S_2'$ is an eigenvector of $A'A$.

2.1.2 Generalized inverse

This section introduces the concept of generalized inverse, that is an extension of the regular inverse. These results can be found in [Bernstein, 2009].

Definition 12. Let $A \in M_{m,n}$. If $A$ is nonzero, then, by the singular value decomposition theorem, there exist orthogonal matrices $S_1 \in M_{n,n}$ and $S_2 \in M_{m,m}$ such that

$$A = S_1 \begin{pmatrix} B & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{pmatrix} S_2,$$

where $B = \text{diag}([\sigma_1(A), \ldots, \sigma_r(A)])$, $r = \text{rank}(A)$, and $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_r(A) > 0$ are the positive singular values of $A$. Then, the (Moore-Penrose) generalized inverse $A^+$ of $A$ is the $m \times n$ matrix

$$A^+ = S_2' \begin{pmatrix} B^{-1} & 0_{r,n-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{pmatrix} S_1'.$$

If $A = 0_{n,m}$, then $A^+ = 0_{m,n}$, although if $m = n$ and $\det(A) \neq 0$, then $A^+ = A^{-1}$. 
Lemma 1. The matrix $A^+$ in Definition 12 above has the following properties.

(i) $AA^+ A = A$.
(ii) $A^+ AA^+ = A^+$.
(iii) $(AA^+)^t = AA^+$.
(iv) $(A^+ A)^t = A^+ A$.

The only generalized inverse $A^+$ that satisfies the above properties is the Moore-Penrose generalized inverse from Definition 12.

Here follows some more properties regarding the Moore-Penrose generalized inverse.

Proposition 1. Let $A \in M_{m,n}$. Then the following statements hold:

(i) $AA^+$ is the projector onto $R(A)$.
(ii) $A^+ A$ is the projector onto $R(A')$.
(iii) $A^+ = A'(A'A)^+ = (A'A)^+ A'$.
(iv) $A^{+t} = (A'A)^+ A = A(A'A)^+$.

2.1.3 Anti-eigenvalues

The information in this section and more information about anti-eigenvalues can for example be found in [Rao, 2005].

Let $A \in M_{p,p}$ be a positive definite matrix. Then the cosine of the angle $\theta$ between a vector $x$ and $Ax$ is

$$
\cos \theta = \frac{x'Ax}{\sqrt{(x'x)(x'A^+x)}},
$$

which has the value 1 if $x$ is an eigenvector of $A$, i.e., $Ax = \lambda x$ for some $\lambda$. The concept of anti-eigenvalue arise when the following question is raised: For what vector $x$, does $\cos \theta$ takes the minimum value or the angle of separation between $x$ and $Ax$ is a maximum. The minimum value is attained at

$$
x = \frac{\sqrt{\lambda_p}x_1 \pm \sqrt{\lambda_1}x_p}{\sqrt{\lambda_1 + \lambda_p}},
$$

and the minimum value equals

$$
\mu_1 = \frac{2\sqrt{\lambda_1 \lambda_p}}{\lambda_1 + \lambda_p},
$$

that can be seen in [Rao, 2005]. The number $\mu_1$ above is called the first anti-eigenvalue of $A$ and for $\mu_1$ there exist a pair of vectors, $u_1$ and $u_2$, that is called the first anti-eigenvectors of $A$. It is possible to continue to define more anti-eigenvalues but only the first one is of interest in this thesis. The above can be summarized in the following definition and theorem.
Definition 13. Let $A \in M_{p,p}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p > 0$. The first anti-eigenvalue of $A$ is defined as

$$\mu_1 = \frac{2\sqrt{\lambda_1 \lambda_p}}{\lambda_1 + \lambda_p},$$

with the corresponding first anti-eigenvectors

$$x = \frac{\sqrt{\lambda_p} x_1 \pm \sqrt{\lambda_1} x_p}{\sqrt{\lambda_1 + \lambda_p}} = (u_1, u_2)$$

where $x_1, x_p$ are the corresponding eigenvectors of $\lambda_1$ and $\lambda_p$.

Theorem 3. Let $A \in M_{p,p}$ be a positive definite matrix with the first anti-eigenvalue $\mu_1$. Then

$$\cos \theta = \frac{x'Ax}{\sqrt{(x'x)(x^2A^2x)}} \geq \mu_1.$$

2.1.4 Kronecker product

Definition 14. Let $A = (a_{ij}) \in M_{p,q}$ and $B = (b_{ij}) \in M_{r,s}$, Then the $pr \times qs$-matrix $A \otimes B$ is called a direct product (Kronecker product) of the matrices $A$ and $B$, if

$$A \otimes B = [a_{ij}B], i = 1, \ldots, p; j = 1, \ldots, q$$

where

$$a_{ij}B = \begin{pmatrix} a_{ij}b_{11} & \cdots & a_{ij}b_{1s} \\ \vdots & \ddots & \vdots \\ a_{ij}b_{r1} & \cdots & a_{ij}b_{rs} \end{pmatrix}.$$

Definition 15. Let $A = (a_1, \ldots, a_q) \in M_{p,q}$, where $a_i, i = 1, \ldots, q$, is the $i$-th column vector. The vectorization operator vec$(\cdot)$ is an operator from $\mathbb{R}^{p \times q}$ to $\mathbb{R}^{pq}$, with

$$\text{vec}(A) = \begin{pmatrix} a_1 \\ \vdots \\ a_q \end{pmatrix}.$$

Proposition 2. Let $A \in M_{p,q}$, $B \in M_{q,r}$ and $C \in M_{r,s}$. Then

$$\text{vec}ABC = (C' \otimes A)\text{vec}B.$$

2.1.5 Matrix derivatives

In this section some elementary matrix derivatives will be presented. These results can be found in [Kollo and von Rosen, 2011].

Proposition 3 (Chain rule). Let $Z \in M_{m,n}$ be a function of $Y$ and $Y$ be a function of $X$. Then

$$\frac{dZ}{dX} = \frac{dY}{dX} \frac{dZ}{dY}.$$
Proposition 4. Let $A$ and $B$ be constant matrices of proper sizes. Then
\[ \frac{d(AXB)}{dX} = B \otimes A'. \]

Proposition 5. Let all matrices be of proper sizes and the elements of $X$ mathematically independent and variable. Then
\[ \frac{d\text{tr}(AXBX')}{dX} = \text{vec} A'XB' + \text{vec}AXB. \]

Remark 1. In order to define matrix derivate some restrictions are made. When a matrix derivate is a derivate of a matrix $Y$ by another matrix $X$ a common assumption is that the matrix $X$ is mathematically independent and variable. It means that,

1. the elements of $X$ are non-constant,
2. no two or more elements are functionally dependent.

The restrictiveness exist because it excludes certain classes of matrices such as symmetric matrices, diagonal matrices, triangular matrices among others from $X$.

2.2 Statistics

This section presents some general definitions in statistics as well as some results in multivariate statistics and an introduction to the Growth Curve model. The ambition is to give enough substance to establish a basis for the understanding of this thesis. The results will be given without proof but the proofs can be found in the references for each subsection.

2.2.1 General concepts

Here follows some definitions and properties regarding estimation in statistics. These definitions and theorems comes from [Casella, 2002].

At first we present the definition of an unbiased estimator.

Definition 16. The bias of a point estimator $\hat{\theta}$ of a parameter $\theta$ is the difference between the expected value of $\hat{\theta}$ and the true value $\theta$; that is, $\text{Bias}_\theta(\hat{\theta}) = E[\hat{\theta} - \theta]$. An estimator whose bias is identically (in $\theta$) equal to 0 is called unbiased and satisfies $E[\hat{\theta}] = \theta$ for all $\theta$.

Here follows a definition of efficiency between two unbiased estimators.

Definition 17. Let $\hat{T}_1$ and $\hat{T}_2$ be two unbiased estimators of the parameter $T$. Then $\hat{T}_1$ is multivariate more efficient then $\hat{T}_2$ if the covariance matrix $\text{cov}(\hat{T}_1) \leq \text{cov}(\hat{T}_2)$ for all possible values of $T$.

Note 1. In the definition above $\text{cov}(\hat{T}_1) \leq \text{cov}(\hat{T}_2)$ means that the matrix $\text{cov}(\hat{T}_1)$ is smaller than $\text{cov}(\hat{T}_2)$ with regard to Loewner’s partial ordering. It means that the matrix $\text{cov}(\hat{T}_2) - \text{cov}(\hat{T}_1)$ is positive semi-definite. If $\text{cov}(\hat{T}_1) < \text{cov}(\hat{T}_2)$ it means that the matrix $\text{cov}(\hat{T}_2) - \text{cov}(\hat{T}_1)$ instead is positive definite.
2.2.2 The Maximum Likelihood Estimator

In this section the estimating procedure, called maximum likelihood, is examined closer and some results regarding properties of these estimators are presented. These results can be found in [Casella, 2002].

First we present a formal definition of the maximum likelihood estimator (MLE) and the likelihood function.

**Definition 18.** A likelihood function $L(\theta)$ is the probability density for the occurrence of a sample configuration $x_1, \ldots, x_n$ given that probability density $f(x; \theta)$ with regard to parameter $\theta$ is known,

$$L(\theta) = f(x_1; \theta) \ldots f(x_n; \theta).$$

**Definition 19.** For each sample point $x$, let $\hat{\theta}(x)$ be a parameter value at which the likelihood function, denoted $L(\theta)$, attains its maximum as a function of $\theta$, with $x$ held fixed. A maximum likelihood estimator (MLE) of the parameter $\theta$ based on a sample $X$ is $\hat{\theta}(X)$.

The main reasons why the MLEs are popular are the following theorems regarding their asymptotic properties.

**Theorem 4.** Let $X_1, X_2, \ldots$ be identically independent distributed $f(x; \theta)$, and let $L(\theta|x) = \prod_{i=1}^{n} f(x_i; \theta)$ be the likelihood function. Let $\hat{\theta}$ denote the MLE of $\theta$. Let $\tau(\theta)$ be a continuous function of $\theta$. Under some regularity conditions, that are described in the note below, on $f(x; \theta)$ and hence, $L(\theta|x)$, for every $\epsilon > 0$ and every $\theta \in \Theta$

$$\lim_{n \to \infty} P_\theta(|\tau(\hat{\theta}) - \tau(\theta)| \geq \epsilon) = 0.$$ 

That is, $\tau(\hat{\theta})$ is a consistent estimator $\tau(\theta)$.

**Theorem 5.** Let $X_1, X_2, \ldots$ be identically independent distributed $f(x; \theta)$, let $\hat{\theta}$ denote the MLE of $\theta$, and let $\tau(\theta)$ be a continuous function of $\theta$. Under some regularity conditions, that are described in the note below, on $f(x; \theta)$ and hence $L(\theta|x)$,

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \to N(0, \text{var}_{LB}(\hat{\theta})),$$

where $\text{var}_{LB}(\hat{\theta})$ is the Cramér-Rao Lower Bound. That is, $\tau(\hat{\theta})$ is a consistent, asymptotically unbiased and asymptotically efficient estimator of $\tau(\theta)$.

**Note 2.** The regularity conditions in the two previous theorems are minor restrictions on the probability density function. Since the underlying distribution in this thesis is the normal distribution, the regularity conditions poses no restriction in the development of results in this thesis. The specifics can be found in Miscellanea 10.6.2 in [Casella, 2002].

2.2.3 The matrix normal distribution

This section presents some definitions and estimators for the normal distribution generalized into the multivariate and matrix normal distribution. These results can be found in [Casella, 2002] [Kollo and von Rosen, 2011] [McCulloch et al., 2008] [Srivastava and Khatri, 1979] [Muirhead, 2005].

Here follows the definition of a univariate normal distribution.
Definition 20. A random variable \(X\) is said to have a univariate normal distribution, i.e., \(X \sim N(\mu, \sigma^2)\), if the probability density function of \(X\) is,

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).
\]

The univariate model above is often generalized in a multivariate setting. That leads to the following definition of a multivariate (vector-) normal distribution.

Definition 21. A random vector \(x_m\) is said to have a \(m\)-variate normal distribution if, for every \(a \in \mathbb{R}^m\), the distribution of \(a^\prime x\) is univariate normal distributed.

Theorem 6. If \(x\) has an \(m\)-variate normal distribution then both \(\mu = E[x]\) and \(\Sigma = \text{cov}(x)\) exist and the distribution of \(x\) is determined by \(\mu\) and \(\Sigma\).

The following estimators are frequently used for a multivariate distribution.

Theorem 7. Let \(x_1, \ldots, x_n\) be random sample from a multivariate normal population with mean \(\mu\) and covariance \(\Sigma\). Then,

\[
\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^\prime,
\]

are the MLE respective corrected MLE of \(\mu\) and \(\Sigma\).

Theorem 8. The estimators in [Theorem 7] are sufficient, consistent and unbiased with respect to the multivariate normal distribution.

The matrix normal distribution adds another dimension of dependence in the data and is often used to model multivariate datasets. Here follows its definition.

Definition 22. Let \(\Sigma = \tau \tau^\prime\) and \(\Psi = \gamma \gamma^\prime\), where \(\tau \in M_{p,r}\) and \(\gamma \in M_{n,s}\). A matrix \(X \in M_{p,n}\) is said to be matrix normally distributed with parameters \(M, \Sigma\) and \(\Psi\), if it has the same distribution as,

\[
M + \tau U \gamma^\prime,
\]

where \(M \in M_{p,n}\) is non-random and \(U \in M_{r,s}\) consists of \(s\) independent and identically distributed (i.i.d.) \(N_r(0, I)\) vectors \(u_i, i = 1, 2, \ldots, s\). If \(X \in M_{p,n}\) is matrix normally distributed, this will be denoted \(X \sim N_{p,n}(M, \Sigma, \Psi)\).

The matrix normal distribution is nothing more than a multivariate normal distribution with some additional structure on the covariance matrix, i.e., let \(X \sim N_{p,n}(M, \Sigma, \Psi)\), then \(\text{vec}X = x \sim N_{pn}(\text{vec}M, \Omega)\), where \(\Omega = \Psi \otimes \Sigma\).

2.2.4 Growth Curve model

In this subsection The Growth Curve model will be presented. It was introduced by [Potthoff and Roy, 1964] and is defined as follows.

Definition 23. Let \(X \in M_{p,N}\) and \(\xi \in M_{q,m}\) be the observation and unknown parameter matrices, respectively, and let \(B \in M_{p,q}\) and \(A \in M_{m,N}\) be the known within and between individual design matrices, respectively. Suppose that \(q \leq p\) and \(r + p \leq N\), where \(r = \text{rank}(A)\). The Growth Curve model is given by

\[
X = B\xi A + \varepsilon,
\]
where the columns of $\varepsilon$ are assumed to be independently $p$-variate normally distributed with mean zero and an unknown positive definite covariance matrix $\Sigma$, i.e.,

$$\varepsilon \sim N_p(0, \Sigma, I_N).$$

The MLEs for the parameters $\xi$ and $\Sigma$ can be found in multiple places in the literature but were first presented by [Khatri, 1966]. Here follows the estimators as well as the covariance matrix of $\hat{\xi}$ which can be found in [Kollo and von Rosen, 2011].

**Theorem 9.** The MLEs of The Growth Curve model are

$$\hat{\xi}_{MLE} = (B'V^{-1}B)^{-1}B'V^{-1}XA'(AA')^{-1}$$

and

$$N\hat{\Sigma}_{MLE} = (X - B\hat{\xi}_{MLE}A)(X - B\hat{\xi}_{MLE}A)'.$$

where $V = X(I - A'(AA')^{-1}A)X'$, $A$ and $B$ have full rank. The covariance matrix of $\hat{\xi}_{MLE}$ is given by

$$\text{cov}(\hat{\xi}_{MLE}) = \frac{n - 1}{n - 1 - (p - q)}(AA')^{-1} \otimes (B'\Sigma^{-1}B)^{-1}.$$

In [Potthoff and Roy, 1964] the following example of data is presented that the authors modeled as a Growth Curve model.

**Example 1.** A certain measurement in a dental study was made on each of 11 girls and 16 boys at the ages 8, 10, 12 and 14. We will assume that the covariance matrix of the 4 correlated observations is the same for all 27 individuals.
This is a typical application for the Growth Curve model. We have a relation between observations of certain individuals, the difference in age between the observations, as well as properties of the individuals themselves.

**Example 2.** Since the inverse of \( V \) is used when estimating the MLE of \( \xi \) problems can arise when the dimension of \( p \) is close to \( N \). When \( p \) gets closer to \( N \), the condition number decrease and the numerical properties of the inverse of \( V \) becomes increasingly worse. This also affect the covariance matrix of the estimator of \( \xi \), that increase when \( p \) is close to \( N \) and the estimator loses precision. In Figure 2.1 is a plot containing the reciprocal condition number of \( V \) for a Growth Curve model. We assume that \( X = B\xi A + E \sim N_p(B\xi A, \Sigma, I_N) \) where \( N = 100, m = 20, q = 20 \) and \( p \) ranging from 20 – 80. It can be seen as \( p \) gets closer to \( N \) the condition number decrease. For the problem were \( \Sigma \) and \( B \) generated for each sample since their dimension changes with \( p \). The matrices \( \xi \) and \( A \) were constant without any structure.

Since \( \xi \) is the mean parameters of the Growth Curve model it is possible to find multiple unbiased estimators. One in particular that will be studied further is the unweighted estimator of \( \xi \). It will be shown that this estimator does not require the inverse of \( V \) hence its numerical properties are better when \( p \) is close to \( N \).

### 2.2.5 Wishart distribution

In this subsection the Wishart distribution and some useful theorems that will be used in this thesis will be introduced. These results can be found in [Kollo and von Rosen, 2011].

**Definition 24.** An random matrix \( W \in M_{p,p} \) is said to be Wishart distributed if and only if \( W = XX' \) for some matrix \( X \), where \( X \sim N_{p,n}(M, \Sigma, I) \), \( \Sigma \geq 0 \). If \( M = 0 \), it is called a central Wishart distribution that will be denoted \( W \sim W_p(\Sigma, n) \).
and if $M \neq 0$ it is called a non-central Wishart distribution which will be denoted $W_p(\Sigma, n, Q)$, where $Q = MM'$.

The Wishart distribution is an generalization to multiple dimensions of the $\chi^2$-distribution. These distributions are useful when estimating covariance matrices in multivariate statistics. The Wishart distribution stays Wishart distributed under linear transformations, a property that is summarized in the following Theorem.

**Theorem 10.** Let $W \sim W_p(\Sigma, n, Q)$ and $A \in M_{q,p}$. Then

$$AWA' \sim W_q(A\Sigma A', n, AQA).$$

From the Theorem above it is possible to see the connection to a $\chi^2$-distribution for the following special case.

**Corollary 1.** Let $W \sim W_p(\Sigma, n)$ and $a \in M_{p,1}$. Then

$$a'Wa \sim \chi^2(n).$$

A combination of independent $\chi^2$-distributions is $F$-distributed which can be seen below.

**Definition 25.** Let $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$ be independent random variables. Then $\frac{X}{m}$ and $\frac{Y}{n}$ are independent random variables. Then $\frac{X}{m}$ and $\frac{Y}{n}$ are independent random variables. Then $\frac{X}{m} \sim F(m, n)$. 
Chapter 3

Unweighted estimator for Growth Curve model under
\[ R(\Sigma^{-1}B) = R(B) \]

This chapter studies the unweighted estimator for a Growth Curve model under certain conditions of \( \Sigma \) and \( B \) studied by [Srivastava and Singull, 2016a, Srivastava and Singull, 2016b]. It presents a new proof of the MLE under these conditions and introduces two useful cases with intraclass respective generalized intraclass where this condition occurs in practice.

3.1 Unweighted vs weighted mean estimator

Characteristic for the Growth Curve model is the post-matrix \( A \) that describes the covariance between rows of the result matrix. This type of modeling assumes that the observed objects have equal covariance between their observations. One common way to use this is to estimate the covariance matrix for the relationship between the result over time. Growth Curves for children can for example be represented in this fashion and a model that can estimate the length of a child, with certain attributes, over time can be presented. The General Linear model only has the relationship between columns, in the previous example that is the relationship between the children’s attributes. The design matrix \( B \) describes the relationship between rows. But this extra ability needs to be addressed when estimating the mean. This special property is the reason that the MLE is a weighted estimator.

The MLEs for the parameters \( \xi \) and \( \Sigma \) can be found in multiple places in the literature but were first presented by [Khatri, 1966] and are described in Chapter 2. The estimator \( \hat{\xi}_{MLE} \) is called weighted because it contains the matrix \( V^{-1} \). This estimator and the covariance estimator that follows from it are complex expressions since it contains an inverse that is unstable when \( p \) is close to \( n \), this makes the covariance matrix for the estimator larger with respect to Loewner’s partial ordering. Problems can arise since the variance of the mean can be extremely high which render the estimations useless. Especially when it is compared with the mean estimator for a General Linear model. [Srivastava and Singull, 2016a, Srivastava and Singull, 2016b] has shown that an unweighted version of the estimators above can be motivated for tests.
and estimations. This estimator has better properties when $V$ is close to singular.

**Theorem 11.** The unweighted estimator $\hat{\xi}_{UW}$ is an unbiased estimator of $\xi$ and is defined as

$$\hat{\xi}_{UW} = (B'B)^{-1}B'XA'(AA')^{-1}$$

and

$$N\hat{\Sigma}_{UW} = (X - B\hat{\xi}_{UW}A)(X - B\hat{\xi}_{UW}A)'.$$  

The covariance matrix for the unweighted mean estimator is given by

$$\text{cov}(\hat{\xi}_{UW}) = (AA')^{-1} \otimes (B'B)^{-1}B'\Sigma B(B'B)^{-1}.$$  

The difference between the MLE and the unweighted estimator can be seen in the lack of $V$ or its inverse that heavily simplifies the expressions.

This chapter will follow onto the lines of thought in [Srivastava and Singull, 2016a] and view some cases where it at least is performing equally well before showing, in the next chapter, conditions for when it performs better.

### 3.2 Previous results

In order to compare the efficiency of the unweighted and the weighted mean estimator, their respective covariance matrices must be studied. In [Baksalary and Puntanen, 1991] the following inequality is presented which is a generalization of the inequality published in [Khatri, 1966; Gaffke and Krafft, 1977]. This inequality can be used to compare the covariance matrices of the estimators of interest but also see when they coincide.

**Theorem 12.** Let $\Sigma \in M_{p,p}$ be a positive definite matrix and $B \in M_{p,q}$ be of full column rank $q \leq p$. Then

$$(B'\Sigma^{-1}B)^{-1} \leq (B'B)^{-1}B'\Sigma B(B'B)^{-1},$$

where $\leq$ is regard to Loewner's partial ordering, with equality if and only if $\mathcal{R}((\Sigma^{-1}B) = \mathcal{R}(B)$.

The inequality above can also be rewritten in the following way using Proposition 1

$$(B'\Sigma^{-1}B)^{-1} \leq (B'B)^{-1}B'\Sigma B(B'B)^{-1}$$

$$\iff (B'\Sigma^{-1}B)^{-1} \leq B'\Sigma B'B$$

In order to establish the condition $\mathcal{R}((\Sigma^{-1}B) = \mathcal{R}(B)$ assumptions need to be made on $\Sigma$ and $B$. In practice this essentially means that columns spanning $\Sigma$, except for the diagonal matrix part of $\Sigma$, needs to be part of $B$. This is certainly an restriction but specific vectors such as 1 often occurs in the design matrix $B$. This means that there are common applications where this is applicable. Some examples will be presented in following sections. There have been some comparisons between models where un-weighted respective weighted estimators has been used with different types of models. But in [Srivastava and Singull, 2016a] those ideas were applied to a Growth Curve model.
3.3 Maximum likelihood estimator

The MLEs for a Growth Curve model under some special conditions has been proposed by [Khatri, 1966, Reinsel, 1982]. The proofs have been done using least square estimators and the requirements has not been clearly specified. In this section follows an clear presentation and proof for the condition where the MLE is the unweighted estimator. The approach of the proof is based on the usual Maximum Likelihood method.

**Theorem 13.** The MLEs for $\xi$ and $\Sigma$ in a Growth Curve model with condition $\mathcal{R}(\Sigma^{-1}B) = \mathcal{R}(B)$ are

$$
\hat{\xi}_{MLE} = (B'B)^{-1}B'XA'(AA')^{-1}
$$

$$
\hat{\Sigma}_{MLE} = (X - B\hat{\xi}UW)(X - B\hat{\xi}UW)'
$$

**Proof.** For the Growth Curve model the likelihood function with respect to $\xi$ and $\Sigma$ can be obtained using the the density of the matrix normal distribution that yields

$$
L(\xi, \Sigma) = (2\pi)^{-\frac{1}{2}np} |\Sigma|^{-\frac{1}{2}n} \exp(-\frac{1}{2} \text{tr}(\Sigma^{-1}(X - B\xi A)(X - B\xi A)'))).
$$

Since $\exp$ is a monotone function the only part of $L$ that depends on $\xi$ is $\text{tr}(\Sigma^{-1}(X - B\xi A)(X - B\xi A)')$. Thus by maximizing that expression with regard to $\xi$ will maximize the likelihood function. If you derive $\text{tr}(\Sigma^{-1}(X - B\xi A)(X - B\xi A)')$ with regard to the matrix $\xi$ with the chain rule, see Proposition 3 you reach

$$
\frac{d(X - B\xi A)}{d\xi} \cdot \frac{d(\text{tr}(\Sigma^{-1}(X - B\xi A)(X - B\xi A)'))}{d(X - B\xi A)}.
$$

Using the rules of matrix derivation of Proposition 4 and Proposition 5 it equals

$$
(A \otimes B') \left( \text{vec}(\Sigma^{-1}(X - B\xi A)) + \text{vec}(\Sigma^{-1}(X - B\xi A)) \right) = 2(A \otimes B')\text{vec}(\Sigma^{-1}(X - B\xi A)) = 2\text{vec}(B'^{-1}(X - B\xi A)A')
$$

where Proposition 2 is used for the last equality.

Equating this expression to zero yield the linear equation in $\xi$

$$
B'^{-1}(X - B\xi A)A' = 0
\quad \iff \quad B\Sigma^{-1}XA' = B\Sigma^{-1}B\xi AA'.
$$

From this equation the solution for $\xi$ can be simplified using the fact the when $\mathcal{R}(\Sigma^{-1}B) = \mathcal{R}(B)$ the following inverse $(B'^{-1}B)^{-1} = B'^{-1}B^+$. 

$$
B\Sigma^{-1}XA' = B'\Sigma^{-1}B\xi AA'
\quad \iff \quad B'^{-1}B\Sigma^{-1}B'\Sigma^{-1}XA' = \xi AA'.
$$

The expression $B'^{-1}B'$ is a projector onto $C(B)$ it is equivalent to $\Sigma^{-1}B\Sigma^{-1}B = \Sigma^{-1}BB'^{-1} \Sigma$ and it is possible to rewrite it as

$$
B^+\Sigma^{-1}BB^+\Sigma^{-1}XA' = \xi AA'.
$$
with the help of the properties of the Moore-Penrose inverse, see Proposition 1. The MLE for $\xi$ can now be reached with

$$\hat{\xi}_{\text{MLE}} = B^+ X A' (A A')^+.$$  

The MLE for $\Sigma$ will then simply be the following

$$\Sigma_{\text{MLE}} = (X - B \hat{\xi}_{\text{MLE}} A)(X - B \hat{\xi}_{\text{MLE}} A)' = V + (I - AA^+) V (I - AA^+)$$

where $V = X (I - A (A A')^+) A' X'$ in accordance with the standard MLE. In order to conclude the proof it needs to be shown that the estimators actually achieve a maximum. This can be done in a similar fashion as for the usual MLE in [Kollo and von Rosen, 2011] and will not be done here.

\[ \square \]

### 3.4 Intraclass model

A common vector in the design matrix $B$ is the 1-vector. This comes from the fact that statistic models often has an offset from origo as a static parameter. The interpretation of this parameter is the average of all samples if all other factors are set to zero.

In this section we will see that if a covariance matrix of an Growth Curve model has an intraclass structure and the design matrix $B$ contains an 1-vector then $R(\Sigma^{-1} B) = R(B)$ holds.

An intraclass structure on the covariance matrix arise in different applications in statistics, e.g., cluster randomized studies, and is defined in the following way

**Definition 26.** An covariance matrix $\Sigma$ is said to be of intraclass if it can be written as

$$\Sigma_{IC} = \sigma^2 \left( (1 - \rho) I + \rho 11' \right)$$

where $-\frac{1}{p-1} < \rho < 1$ and $\sigma > 0$.

The inverse of an covariance matrix with intraclass structure is also an intraclass which is summarized in the following lemma that will be used in the proof below.

**Lemma 2.** For an intraclass covariance matrix $\Sigma_{IC}$ defined above the inverse can be written as

$$\Sigma_{IC}^{-1} = \frac{1}{\sigma^2 (1 - \rho)} \left( I - \frac{\rho}{1 + (n-1) \rho} 11' \right)$$

which is also an intraclass matrix.

The result that the condition $R(\Sigma^{-1} B) = R(B)$ holds for an intraclass covariance matrix where an 1 is included in the design matrix is presented in the following theorem

**Theorem 14.** Assume that the matrix $B$ contains an 1-vector and $\Sigma_{IC}$ is intraclass then $R(\Sigma_{IC}^{-1} B) = R(B)$. 

3.5. Generalized intraclass model

Proof. Without loss of generality the problem can be rearranged such that the design matrix \( B \) can be written as \( B = [1 \ B_0] \). The inverse of \( \Sigma_{IC} \) can, with the added assumptions on \( B \), be simplified to

\[
\Sigma_{IC}^{-1} = \frac{1}{\sigma^2(1 - \rho)} \left( I - \frac{\rho}{1 + (n - 1)\rho} \B_0 \right) \B_0
\]

where \( \sigma \) is a known positive definite matrix and \( w \) is a known vector.

The inverse of an generalized intraclass covariance matrix is also generalized intraclass which is summarized in the following lemma and will be used in the proof below.

Lemma 3. For an generalized intraclass covariance matrix \( \Sigma_{GIC} \) defined above its inverse can be written as

\[
\Sigma_{GIC}^{-1} = \frac{1}{\sigma^2(1 - \rho)} \left( G^{-1} - \frac{\rho}{1 + (n - 1)\rho} w w' \right)
\]

which is also an generalized intraclass matrix.

The proof follows directly since the matrix \( G \) and vector \( W \) are known.

The result that the condition \( \mathcal{R}(\Sigma^{-1}B) = \mathcal{R}(B) \) holds for an intraclass covariance matrix where an \( w \) is included in the design matrix is presented in the following theorem.

3.5 Generalized intraclass model

In this section it will be shown that the results regarding intraclass can be extended to a covariance matrix of an Growth Curve model with an generalized intraclass structure. This means that instead of assuming that the one-vector \( 1 \) exist in the design matrix \( B \) a known general vector \( w \) from the design matrix \( B \). Hence this is a generalization of the model in the previous section.

The generalized intraclass covariance matrix can be defined in the following way.

Definition 27. An covariance matrix \( \Sigma \) is said to be of generalized intraclass if it can be written as

\[
\Sigma_{GIC} = \sigma^2((1 - \rho)G + \rho w w')
\]

where \( -\frac{1}{p-1} \leq \rho < 1 \), \( \sigma > 0 \) and \( G \) is a known positive definite matrix and \( w \) is a known vector.

These results can be extended for a combination of known matrices and corresponding \( w \)-vectors.
Theorem 15. Assume that the design matrix $B$ contains a known vector $w$ and $\Sigma_{GIC}$ is generalized intraclass with a known positive definite matrix $G$ and vector $w$, then $\mathcal{R}(\Sigma_{GIC}^{-1}B) = \mathcal{R}(B)$.

Proof. Without loss of generality the problem can be rearranged such that the design matrix $B$ can be written as $B = [w \ B_0]$. The inverse of $\Sigma_{GIC}$ can with the added assumptions on $B$ be simplified to

$$\Sigma_{GIC}^{-1}B = \frac{1}{\sigma^2(1 - \rho)} \left( G^{-1} - \frac{\rho}{1 + (n - 1)\rho}ww' \right) [w \ B_0]$$

$$= \frac{1}{\sigma^2(1 - \rho)} \left[ \left( G^{-1} - \frac{(p-1)\rho}{1 + (p-1)\rho} \right)w \ G^{-1}B_0 - \frac{\rho}{1 + (p-1)\rho}ww'B_0 \right].$$

Since $C = \frac{\rho}{1 + (p-1)\rho}ww'B_0$ is of rank 1 there exist a $\alpha_i \in \mathbb{R}$ such that $c_i = \alpha_i(G^{-1} - \frac{(p-1)\rho}{1 + (p-1)\rho})w$ for all $i$, where $c_i$ is a row in $C$. Thus

$$\mathcal{R}(\Sigma_{GIC}^{-1}B) = \mathcal{R}\left( \frac{1}{\sigma^2(1 - \rho)} \left[ \left( G^{-1} - \frac{(p-1)\rho}{1 + (p-1)\rho} \right)w \ G^{-1}B_0 - C \right] \right)$$

$$= \mathcal{R}([w \ B_0]) = \mathcal{R}(B).$$

$\square$
Chapter 4

Unweighted estimator for Growth Curve model without assumptions on $B$

4.1 Background and previous results

In the previous chapter we developed some conditions when the unweighted estimator is the MLE for a Growth Curve model. It is also possible to compare the unweighted estimator with the weighted estimator of $\xi$ without any assumptions on $B$. This has been done for a singular linear model in [Liu and Neudecker, 1997] but not for a Growth Curve model. It will be shown that the loss of precision will depend on the eigenvalues of $\Sigma$ as well as the dimensions of the problem and in plenty of cases outside $R(\Sigma^{-1}B) = R(B)$, the unweighted estimator of $\xi$ will still be preferred.

An hypothesis testing procedure for this is also presented when the eigenvector of the covariance matrix is known. This is applicable when the covariance matrix, for example, have a circular Toeplitz structure.

Below follows an linear algebra inequality that is a matrix generalization of Kantorovich inequality. It is presented in [Liu and Neudecker, 1997] and presents an upper bound on the equality in Theorem 12 with regard to $\Sigma$'s eigenvalues and it will serve as a basis for the comparisons between the estimators in this chapter.

**Theorem 16.** Let $\Sigma \in M_{p,p}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p > 0$ and $U \in M_{p,q}$ with full column rank $q \leq p$. Then

$$U^+\Sigma U^+ \leq \frac{1}{\mu_1} \left(U^T \Sigma^{-1} U\right)^{-1}$$

where $\mu_1$ is the first anti-eigenvalue of $\Sigma$ given by $\mu_1 = \frac{2\sqrt{\lambda_1 \lambda_p}}{\lambda_1 + \lambda_p}$ or its reciprocal.

4.2 Eigenvalues and coefficients

As can be seen in the previous section an important part of the comparisons between the estimators are done using the first anti-eigenvalue or indirect the smallest and largest
eigenvalue. Combining the results of Theorem 9, Theorem 11, and Theorem 16 the following theorem is yielded for when the unweighted estimators are preferred.

**Theorem 17.** If

\[
\frac{(\lambda_1 + \lambda_p)^2}{4\lambda_1 \lambda_p} \leq \frac{n - 1}{n - 1 - (p - q)},
\]

(4.1)

then the unweighted estimator of \( B \) has a smaller covariance matrix, with accordance to Loewner's partial ordering, than the MLE.

For a large number of observations \( N \) the weighted mean estimator, the MLE, of \( \xi \) has a smaller covariance matrix. This is expected since it is the MLE and with regard to Theorem 5 it should be the unbiased estimator with the smallest covariance matrix when \( N \) tends to infinity. However, when \( p \) is close to \( N \) the factor \( \frac{(n-1)}{(n-1)-(p-q)} \) increases and the covariance matrix can be larger for the weighted estimator compared to the unweighted estimator with regard to Loewner’s partial ordering.

### 4.3 Unweighted vs weighted estimator with known and fixed eigenvectors

There exist many covariance matrices that the conditions \( R(\Sigma^{-1} B) = R(B) \) are not fulfilled for. Since the condition heavily restricts \( \Sigma \)'s span it also restricts its eigenspace. Therefore untrivial eigenspaces are difficult to handle from the perspective of Theorem 12.

In this section we will see that it is possible to establish distributions for eigenvalues and ratios of them when the covariance matrix has fixed eigenvectors. These results can be used together with Theorem 17 in order to determine if the unweighted respective weighted estimator performs better. This section will use the notation \( \Sigma_{FE} \) for a covariance matrix with known and fixed eigenvectors. We will denote the known and fixed eigenvectors with \( \gamma_k \) with the corresponding eigenvalues denoted \( \lambda_k \).

#### 4.3.1 Distributions of eigenvalues

In matrix theory, eigenvalues and their eigenvectors play an important role for the analysis of matrices. In statistics there exist some general results regarding eigenvalues and eigenvectors. The usual problem is that they require asymptotic properties to be useful or their exact expressions are complicated. But in this section it will be shown that under some common structures for a covariance matrix useful theorems regarding eigenvalues can be established. These theorems can be used to differentiate between the unweighted and weighted estimators in a Growth Curve model and make inference about the first anti-eigenvalue.

Some progress can be made if the estimator of a covariance matrix follow a Wishart distribution. In Theorem 3.3.12 in Srivastava and Khatri, 1979, the relationship between the Wishart and the \( \chi^2 \)-distribution can be seen. The following theorem is a special case where the Wishart distribution is central.

**Theorem 18.** Let \( W \sim W_p(D, n) \), where \( W = (W_{ij}) \) and \( D = (D_{ii}) \) is a diagonal matrix. Then \( W_{ii}/D_{ii} \) are independently distributed, each having the distribution of a \( \chi^2 \) random variable with \( n \) degrees of freedom.
When the eigenvectors are fixed, the following estimator is proposed for estimating eigenvalues.

**Proposition 6.** Let $W \sim W_p(\Sigma_{FE}, n)$, where $\Sigma_{FE}$ is a covariance matrix with known and fixed eigenvectors. Let $\lambda_1 \geq \cdots \geq \lambda_p$ and $v_1, \ldots, v_p$ be the eigenvalues and corresponding eigenvectors of $\Sigma$. Then the proposed estimator is $\hat{\lambda}_i = v_i^T W v_i$.

Since the eigenvalue $\lambda_i$ of a matrix $\Sigma$ can be defined as $v_i^T \Sigma v_i$, this is a logical estimator. The usual problem is that the estimators of the eigenvectors $v_i$ themselves are random variables. This makes it difficult to derive something useful from the estimator $\hat{\lambda}_i$ above. But if the covariance matrix has known and fixed eigenvectors the following theorem for distribution of the proposed estimator can be established.

**Theorem 19.** The estimators $\hat{\lambda}_i$ from [Proposition 6](#) are independent and follow a $\chi^2$-distribution with $n$ degrees of freedom.

**Proof.** Let $v_1, \ldots, v_p$ denote the known eigenvectors for the covariance matrix $\Sigma_{FE}$. Since every covariance matrix has a spectral decomposition, see [Theorem 1](#) it can be rewritten as $\Sigma = VDV'$, where $V = (v_1, \ldots, v_p)'$ and $D$ is a diagonal matrix composed of $\Sigma$'s eigenvalues. The matrix $\hat{L} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_p) = V'WV$. According to [Theorem 10](#) it is Wishart distributed with $V'\Sigma V = V'VDVV' = D = \text{diag}(\lambda_1, \ldots, \lambda_p)$ and $n$ degrees of freedom. In [Theorem 18](#) it can be seen that the estimators $\hat{\lambda}_i \sim \chi^2(n)$ are independently distributed.

**Lemma 4.** For two $\hat{\lambda}_i$ and $\hat{\lambda}_j$ with $i \neq j$ from [Theorem 19](#) the random variable $\frac{\hat{\lambda}_i}{\hat{\lambda}_j}$ is $F$-distributed with $(n, n)$ degrees of freedom.

**Proof.** Since $\lambda_i$ and $\lambda_j$ are independent and follow a $\chi^2$-distribution with $n$ degrees of freedom. It follows from [Definition 25](#) that the expression $\frac{\lambda_i}{\lambda_j}$ is $F$-distributed with $(n, n)$ degrees of freedom.

### 4.3.2 Hypothesis testing and confidence interval

Some theorems regarding eigenvalue estimation as well as their relationship and distribution have now been presented. This leads back to the starting point of this discussion. When does the weighted respective unweighted mean estimator do better in a Growth Curve model? In [Theorem 16](#) the answer to this question depend on the first anti-eigenvalue of the covariance matrix, or indirectly the largest and smallest eigenvalue. In order to make inference regarding which estimator performs best we need to make inference about the first anti-eigenvalue. However, the theory of anti-eigenvalues, or specifically the first anti-eigenvalue, has not been studied much in statistics. In order to better understand this some simulations will be done in Chapter 5. From the definition of the first anti-eigenvalue, we can use the smallest and largest eigenvalue to make inference regarding it. From equation (4.1) follows a condition when the unweighted estimator is preferred. Now the question regarding which estimator performs best can be answered.

The condition $(\lambda_1 + \lambda_p)^2 \leq \frac{n - 1}{n - 1 - (p - q)}$ can be rewritten in the following way,

$$\frac{\lambda_1}{4\lambda_p} + \frac{\lambda_p}{4\lambda_1} + \frac{1}{2} \leq \frac{n - 1}{n - 1 - (p - q)} \iff l + \frac{1}{l} \leq \frac{4(n - 1)}{n - 1 - (p - q) - 2},$$
where $l = \frac{\lambda_p}{\lambda_1}$ is the quotient of the largest and smallest eigenvalue. Hence, it is possible to determine which estimator performs best with help of the ratio between the largest and smallest eigenvalue. The distribution of the ratio is F-distributed when the eigenvectors are fixed and known from Lemma 4. Since the distribution is known, it is possible to calculate a confidence interval as well as hypothesis testing if wanted.

The main equation for this inference is

$$\frac{\hat{\lambda}_p/\lambda_p}{\hat{\lambda}_1/\lambda_1} = \frac{\lambda_p}{\hat{\lambda}_1} \sim F(n, n),$$

where $n$ is the degrees of freedom of the covariance matrix.

Example 3. Construct the test statistic according to the following

$$T = \left(\frac{\hat{\lambda}_1 + \hat{\lambda}_p}{4\lambda_1}\right)^2 = \frac{\lambda_1}{4\lambda_p} + \frac{\lambda_p}{2\lambda_1} + \frac{1}{2}$$

that is, the first anti-eigenvalue squared. Then the hypothesis if the unweighted or weighted mean estimator performs best can be derived from equation (4.1).

$$H_0 : \left(\frac{\lambda_1 + \lambda_p}{4\lambda_1}\right)^2 = \frac{n-k}{n-k-(p-q)} \quad \text{vs} \quad H_1 : \left(\frac{\lambda_1 + \lambda_p}{4\lambda_1}\right)^2 \neq \frac{n-k}{n-k-(p-q)}$$

The limits of the test can be derived as follows: Assume that $P(a \leq \frac{\lambda_p}{\lambda_1} \leq b) = P(\frac{1}{b} \leq \lambda_1/\lambda_p \leq \frac{1}{a}) = 1 - \alpha$. Combining these will result in $T_{CI} = P \left( \frac{\alpha}{4} + \frac{1}{4b} + \frac{1}{2} \leq T \leq \frac{b}{4} + \frac{1}{4a} + \frac{1}{2} \right) = 1 - \alpha$. Hence if

$$T \leq \frac{a}{4} + \frac{1}{4b} + \frac{1}{2}$$

the unweighted estimator performs better than the weighted. If instead

$$T \geq \frac{b}{4} + \frac{1}{4a} + \frac{1}{2}$$

the weighted estimator performs better.

### 4.3.3 Symmetric circular Toeplitz structured covariance matrix

A test has now been proposed where the covariance matrix has known and fixed eigenvectors. This property arises in different structures. One of the covariance matrices carrying this structure is the symmetric circular Toeplitz matrix. This kind of matrix has a known structure with eigenvectors that only depend on the size of the matrix, not the parameters. Therefore it is especially suitable for the conditions given in Theorem 16.

Here follows the definition and expressions for eigenvalues and eigenvectors for a symmetric circular Toeplitz matrix that can, for example, be found in [Olkin and Press, 1969].

**Definition 28.** A covariance matrix $T \in M_{p,p}$ is called symmetric circular Toeplitz when

$$T = \begin{pmatrix} t_0 & t_1 & t_2 & \ldots & t_1 \\ t_1 & t_0 & t_1 & \ldots & t_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_1 & \ldots & \ldots & \ldots & t_1 \end{pmatrix}$$
where \( t_{ij} = t_{[|i-j|]} \). The matrix \( T \) depends on \( \frac{p+3}{2} \) parameters when \( p \) is odd and \( \frac{p+2}{2} \) when \( p \) is even.

**Theorem 20.** A symmetric positive definite circular Toeplitz matrix has eigenvalues

\[
\lambda_k = \sum_{i=1}^{p} t_k \cos(2\pi p^{-1}(k-1)(p-i+1))
\]

and corresponding eigenvectors \( \gamma_k \) defined by

\[
\gamma_{ik} = p^{-\frac{1}{2}} \left[ \cos(2\pi p^{-1}(i-1)(k-1)) + \sin(2\pi p^{-1}(i-1)(k-1)) \right]
\]

where \( i = 1, \ldots, p \) and \( k = 1, \ldots, p \)

Since the eigenvectors \( \gamma_k \) does not depend on the parameters of the matrix it is possible to use the established theorems to make statistical inference.

Two other covariance matrices that has known and fixed eigenvectors are the intraclass and generalized intraclass covariance matrix described in the previous chapter. These are actually specific examples of the symmetric circular Toeplitz covariance matrix.
Chapter 4. Unweighted estimator for Growth Curve model without assumptions on $B$
Chapter 5

Simulations

In this chapter some simulations relevant to this thesis will be presented. There will be some simulations regarding the distribution of anti-eigenvalues as well as most relevant theorems from this thesis. The purpose is to simulate the effects of the statements and give an idea of the performance and significance of them.

5.1 Distribution of the first anti-eigenvalue

In this section some histograms regarding the distribution of the first anti-eigenvalue defined in [Definition 13] will be presented. First an simulation will be made when the covariance matrix has a symmetric circular Toeplitz structure, [Definition 28], after that an unstructured covariance matrix will be studied.

5.1.1 Symmetric circular Toeplitz covariance matrix

In this simulation it is assumed that $x \sim N(1_8, \Sigma_{CT})$, where

$$
\Sigma_{CT} = \begin{pmatrix}
8 & 0 & 2 & 0 & 0 & 0 & 2 & 0 \\
0 & 8 & 0 & 2 & 0 & 0 & 0 & 2 \\
2 & 0 & 8 & 0 & 2 & 0 & 0 & 0 \\
0 & 2 & 0 & 8 & 0 & 2 & 0 & 0 \\
0 & 0 & 2 & 0 & 8 & 0 & 2 & 0 \\
0 & 0 & 0 & 2 & 0 & 8 & 0 & 2 \\
2 & 0 & 0 & 0 & 2 & 0 & 8 & 0 \\
0 & 2 & 0 & 0 & 0 & 2 & 0 & 8 \\
\end{pmatrix}
$$

has a symmetric circular Toeplitz structure. A simulation of 10 000 samples of the distribution of the first anti-eigenvalue of the estimated covariance matrix $\Sigma_{CT}$ were made. For each sample 80 observations were made as a basis for the estimation of the covariance matrix. This yielded the following result:

In the Figure 5.1 above it seems that the first anti-eigenvalue follows a symmetric distribution.
5.1.2 Unknown covariance matrix

In this simulation we assumed that $x \sim N(1_8, \Sigma)$ where

$$\Sigma = \begin{pmatrix}
1.70726 & 1.76325 & 1.89110 & 0.91624 & 1.74651 & 1.85767 & 1.45273 & 0.99676 \\
1.76325 & 2.51264 & 2.12305 & 1.03537 & 2.30951 & 2.20502 & 1.97452 & 2.03714 \\
1.89110 & 2.12305 & 3.11924 & 0.94676 & 2.11125 & 2.55778 & 1.57962 & 1.77444 \\
0.91624 & 1.03537 & 0.94676 & 0.63925 & 1.17584 & 0.83902 & 0.75082 & 0.80187 \\
1.74651 & 2.30951 & 2.11125 & 1.17584 & 3.14451 & 2.52562 & 1.67662 & 2.16732 \\
1.85767 & 2.20502 & 2.55778 & 0.83902 & 2.52562 & 3.19486 & 1.76564 & 1.68779 \\
1.45273 & 1.97452 & 1.57062 & 0.75082 & 1.67662 & 1.76564 & 1.78145 & 1.15548 \\
0.99676 & 2.03714 & 1.77444 & 0.80187 & 2.16732 & 1.68779 & 1.15548 & 2.77791
\end{pmatrix}.$$ 

A simulation of 10,000 samples of the distribution of the first anti-eigenvalue of the estimated covariance matrix $\Sigma$ were made. For each sample 80 observations were made as a basis for the estimation of the covariance matrix. This yielded the following result:

In Figure 5.2 it seems that the first anti-eigenvalue follows a symmetric distribution as in Figure 5.1 which is interesting since this indicates that there might exist more general theorems regarding the distribution of anti-eigenvalues.

5.2 Eigenvalue confidence interval

In this section confidence intervals for eigenvalues of the covariance matrix will be calculated based on the results in Chapter 4 as an example of what can be done. We assume that $X \sim N_5(n, \Sigma_{CT}, I_n)$ where $\Sigma_{CT}$ is a circular Toeplitz covariance matrix and $n$ are different number of observations where
5.2. Eigenvalue confidence interval

The true eigenvalues of $\Sigma_{CT}$ are

\[
e = \begin{pmatrix}
5.38197 \\
5.38197 \\
7.61803 \\
7.61803 \\
14
\end{pmatrix}.
\]

Figure 5.2: Histogram of the first anti-eigenvalue when $\Sigma$ has no structure.
This confirms the theoretical result from Proposition 6 and shows that the range of the confidence intervals are useful in practice.

### 5.3 Hypothesis testing

In this section we will simulate hypothesis testing based on the testing procedure in Chapter 4. The purpose of this simulation is to see how well the test distinguishes between the unweighted and the weighted mean estimator for a Growth Curve model in practice. We assume that \( X = B \xi A + E \sim N_{6,N}(B \xi A, \Sigma_{CT}, I_n) \) where \( \Sigma_{CT} \) is a circular Toeplitz covariance matrix and \( N \) are the number of observations. We model this as a Growth Curve model and make 10 000 tests with significance level \( \alpha = 0.05 \) per \( N \) to determine the power of the test. The true value of the \( H_1 \) is displayed as reference point. If it is greater than zero the unweighted estimator is preferred and below zero the standard MLE is preferred. For this test

\[
B = \begin{pmatrix}
0.24790 & 0.14523 & 0.17287 & 0.30586 \\
0.39350 & 0.83045 & 0.02769 & 0.38080 \\
0.82326 & 0.58860 & 0.06529 & 0.13457 \\
0.32888 & 0.62643 & 0.45441 & 0.18983 \\
0.82305 & 0.29646 & 0.12651 & 0.99140 \\
0.47549 & 0.64442 & 0.27167 & 0.62054
\end{pmatrix},
\]

\[
\xi = \begin{pmatrix}
1 & 2 & 3 & 4 \\
2 & 2.5 & 3 & 3.5 \\
1 & 3 & 5 & 7 \\
1 & 4 & 2 & 5
\end{pmatrix},
\]

and 

\[
\Sigma = \begin{pmatrix}
25 & 7 & 10 & 5 & 10 & 7 \\
7 & 25 & 10 & 5 & 10 & 7 \\
10 & 7 & 25 & 7 & 10 & 5 \\
5 & 10 & 7 & 25 & 7 & 10 \\
10 & 5 & 10 & 7 & 25 & 7 \\
7 & 10 & 5 & 10 & 7 & 25
\end{pmatrix}.
\]
Since the size of $A$ depends on $N$ it was randomized for each sample. Here follows the results

<table>
<thead>
<tr>
<th>$N$</th>
<th>Power of the test</th>
<th>Value of $H_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.9313</td>
<td>0.8750</td>
</tr>
<tr>
<td>6</td>
<td>0.8283</td>
<td>0.5417</td>
</tr>
<tr>
<td>7</td>
<td>0.7308</td>
<td>0.3750</td>
</tr>
<tr>
<td>8</td>
<td>0.6392</td>
<td>0.2750</td>
</tr>
<tr>
<td>9</td>
<td>0.5908</td>
<td>0.2083</td>
</tr>
<tr>
<td>10</td>
<td>0.5128</td>
<td>0.1607</td>
</tr>
<tr>
<td>11</td>
<td>0.4830</td>
<td>0.1250</td>
</tr>
<tr>
<td>12</td>
<td>0.4534</td>
<td>0.0972</td>
</tr>
<tr>
<td>13</td>
<td>0.4257</td>
<td>0.0750</td>
</tr>
<tr>
<td>14</td>
<td>0.3901</td>
<td>0.0568</td>
</tr>
<tr>
<td>15</td>
<td>0.3673</td>
<td>0.0417</td>
</tr>
<tr>
<td>16</td>
<td>0.3423</td>
<td>0.0288</td>
</tr>
<tr>
<td>17</td>
<td>0.3222</td>
<td>0.0179</td>
</tr>
<tr>
<td>18</td>
<td>0.3078</td>
<td>0.0083</td>
</tr>
<tr>
<td>20</td>
<td>0.2806</td>
<td>-0.0074</td>
</tr>
<tr>
<td>22</td>
<td>0.2501</td>
<td>-0.0197</td>
</tr>
<tr>
<td>50</td>
<td>0.1981</td>
<td>-0.0824</td>
</tr>
<tr>
<td>100</td>
<td>0.5124</td>
<td>-0.1044</td>
</tr>
<tr>
<td>200</td>
<td>0.9308</td>
<td>-0.1148</td>
</tr>
<tr>
<td>300</td>
<td>0.9961</td>
<td>-0.1183</td>
</tr>
<tr>
<td>500</td>
<td>1.0000</td>
<td>-0.1210</td>
</tr>
</tbody>
</table>

This shows that the tests perform well in practice and when $N$ goes to infinity the standard MLE is preferred as expected. But for small values of $N$ the unweighted mean estimator will be preferred.
Chapter 6

Further research

Further research topics in this area could be a continuation and exploration of anti-eigenvalues in statistics. Since it is a measure of the rotation of a matrix it would be interesting to see its place in multivariate statistics. Another area could be the unweighted estimators and find out more of how they compare against MLEs and other commonly used estimators, both for the Growth Curve model and other statistical models.
Bibliography


Karlsson, 2016.


Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years from the date of publication barring exceptional circumstances. The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility. According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement. For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: http://www.ep.liu.se/

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtida ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstå. Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art. Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsmann i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart. För yttligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

© 2016, Emil Karlsson