Approximation of the Distribution of the Location Parameter in the Growth Curve Model

Tõnu Kollo, Anu Roos and Dietrich von Rosen

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2005:8
ISSN 1651-8543

Approximation of the Distribution of the Location Parameter in the Growth Curve Model

Tõnu Kollo[1], Anu Roos

Institute of Mathematical Statistics
University of Tartu
J. Liivi 2-513, 50409 Tartu, Estonia

Dietrich von Rosen

Centre of Biostochastics
Swedish University of Agricultural Sciences
Box 7013, SE-750 07 Uppsala, Sweden

Abstract

In the paper an Edgeworth-type approximation to the density of the estimator of the location parameter in the Growth Curve model has been found. The approximation is a mixture of a normal and a Kotz-type distribution, thus being an elliptical distribution. Shape and properties of the distribution are examined. Finally a small example is given to demonstrate an application of the approximation.

Key words: Elliptical distribution, Growth Curve model, Kotz-type distribution, location parameter, mixture distribution

[1]E-mail address to the correspondence author: tonu.kollo@ut.ee

1 Motivation and overview of the problem

Over the years linear models have been one of the most widely used tools in statistics. Best linear unbiased estimators exist for the parameters describing the mean and if we assume an underlying normal distribution, these linear estimators are the best unbiased estimators. The linear model can be written in the form

$$\boldsymbol{x}' = \boldsymbol{\beta}'\boldsymbol{C} + \boldsymbol{e}',$$

where $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, $\boldsymbol{\beta}'$: $1 \times k$ is an unknown parameter matrix, $\boldsymbol{C}$: $k \times n$ is a known design matrix of full rank and σ^2 is the variance. The estimator of the parameter $\boldsymbol{\beta}$ is given by the equality

$$\widehat{\boldsymbol{\beta}}' = \boldsymbol{x}'\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}.$$

Here, $\widehat{\boldsymbol{\beta}} \sim N_k(\boldsymbol{\beta}, \sigma^2(\boldsymbol{C}\boldsymbol{C}')^{-1})$, i.e. $\widehat{\boldsymbol{\beta}}$ is multivariate normal with mean $\boldsymbol{\beta}$ and dispersion matrix $\sigma^2(\boldsymbol{C}\boldsymbol{C}')^{-1}$.

There exists an immediate extension to a multivariate linear model, i.e. a MANOVA model. Let

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{C} + \boldsymbol{E},$$

where $\boldsymbol{X}$: $p \times n$ is the data matrix, $\boldsymbol{B}$: $p \times k$ is a matrix of unknown parameters, $\boldsymbol{C}$: $k \times n$ is the same design matrix as in the univariate linear model, and the error matrix $\boldsymbol{E} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)$, where $\boldsymbol{e}_i \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ are i.i.d. and $\boldsymbol{\Sigma}$ is an unknown positive definite parameter matrix. The maximum likelihood estimator (MLE) of $\boldsymbol{B}$ is given by

$$\widehat{\boldsymbol{B}} = \boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \sim N_{p,k}(\boldsymbol{B}, (\boldsymbol{C}\boldsymbol{C}')^{-1}, \boldsymbol{\Sigma}), \tag{1.1}$$

where $N_{p,k}(\bullet, \bullet, \bullet)$ represents the matrix normal distribution (Kollo & von Rosen, 2005, pp. 191-193).

In the model, for each experimental unit, we have p correlated measurements. There is no functional relationship between the mean parameters within experimental units. However, if we suppose a linear functional relationship, we have the following extension of the MANOVA model

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{C} + \boldsymbol{E}, \tag{1.2}$$

where $\boldsymbol{X}$, $\boldsymbol{C}$ and $\boldsymbol{E}$ are as in the MANOVA model but $\boldsymbol{A}$: $p \times q$, $r(\boldsymbol{A}) = q$, is a within individuals design matrix, and $\boldsymbol{B}$ is of size $q \times k$. This is the Growth Curve model which was first introduced by Potthoff & Roy (1964), although some other authors had earlier worked with a similar model. For a review of the model see Woolson & Leeper (1980), von Rosen (1991) or Srivastava & von Rosen (1999). Kshirsagar & Smith (1995) have written a book on the model and for a recent contribution see Kollo & von Rosen (2005, Chapter 4), where the model and some extensions are presented. The MLE of $\boldsymbol{B}$ (e.g. see Khatri, 1966 or Kollo & von Rosen, 2005) equals

$$\widehat{\boldsymbol{B}} = (\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}, \tag{1.3}$$

where

$$\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{C})\boldsymbol{X}'. \tag{1.4}$$

Unlike the MANOVA model, $\widehat{\boldsymbol{B}}$ in (1.3) is a non-linear estimator. Note that if $\boldsymbol{A} = \boldsymbol{I}$, the estimator $\widehat{\boldsymbol{B}}$ in (1.3) is identical to $\widehat{\boldsymbol{B}}$ in (1.1). Unfortunately, $\widehat{\boldsymbol{B}}$ in (1.3) is not normally distributed. This observation is the starting point of our paper.

Several authors have described the distribution of $\widehat{\boldsymbol{B}}$ in (1.3). Gleser & Olkin (1970) were the first to derive the distribution in a canonical form of the model. Later Kabe (1975), using a different technique, considered the model in the set-up given in (1.2). Kenward (1986) expressed the density of $\widehat{\boldsymbol{B}}$ with the help of hypergeometric functions. All these results are difficult to apply. Fujikoshi (1985, 1987) derived asymptotic expansions with upper error bounds for the density of linear combinations of the elements in $\widehat{\boldsymbol{B}}$. Fujikoshi's results are based on an interesting matrix identity which leads to a decomposition into a sum of two independently distributed random variables, where one is normally distributed.

General Edgeworth type expansions for multivariate density functions were presented by Kollo & von Rosen (1998) and if using them, Fujikoshi's results can be generalized so that a density approximation for the matrix $\widehat{\boldsymbol{B}}$ with an upper error bound is obtained. We will observe the remarkable fact that the density approximation is a density itself which is unusual when performing Edgeworth expansions. Furthermore, it is noticed that a random variable with the approximating density has the same mean and dispersion matrix as $\widehat{\boldsymbol{B}}$ and that the approximating density is a mixture of two elliptical distributions: a normal and a Kotz-type distribution. The results can be used when constructing confidence sets. In Section 2 we give necessary notation and results on the multivariate density approximation of the distribution of the parameter matrix estimate $\widehat{\boldsymbol{B}}$. Section 3 deals with properties of the density approximation for $\widehat{\boldsymbol{B}}$, while in Section 4 we examine our approximation in the Growth Curve model. Also a small simulation study is included to show the properties of the approximating distribution.

2 Formal expansion of the density of $\widehat{\boldsymbol{B}}$

Several strategies could be followed when approximating the distribution of $\widehat{\boldsymbol{B}}$ in (1.3). The asymptotic normal distribution would be the most common approximation to look for. Asymptotic normality of $\widehat{\boldsymbol{B}}$ is based on the convergence of the matrix $\boldsymbol{S}$, given in (1.4), if $n \to \infty$. It follows in the same way as for the sample covariance matrix in multivariate analysis (see Anderson (2003), p. 86, for example) that if $n \to \infty$,

$$\frac{1}{n-k}\boldsymbol{S} \xrightarrow{P} \boldsymbol{\Sigma}, \tag{2.1}$$

where $\xrightarrow{P}$ denotes the convergence in probability.

After replacing $\boldsymbol{S}^{-1}$ by $\boldsymbol{\Sigma}^{-1}$ in (1.3), we get a natural approximation for $\widehat{\boldsymbol{B}}$:

$$\boldsymbol{B}_N = (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}, \tag{2.2}$$

which is normally distributed because $\boldsymbol{X}$ is matrix normally distributed, $\boldsymbol{X}_{p,n} \sim N(\boldsymbol{ABC}, \boldsymbol{\Sigma}, \boldsymbol{I})$. Therefore $\boldsymbol{B}_N \sim N_{q,k}(\boldsymbol{B}, (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}, (\boldsymbol{C}\boldsymbol{C}')^{-1})$. However, it appears that a specific elliptical distribution gives a much better approximation. Clearly both $\widehat{\boldsymbol{B}}$ and $\boldsymbol{B}_N$ give unbiased estimators:

$$E[\widehat{\boldsymbol{B}}] = E[\boldsymbol{B}_N] = \boldsymbol{B}. \tag{2.3}$$

In Kollo & von Rosen (2005, §4.2) the dispersion matrix $D[\widehat{\boldsymbol{B}}]$ is found:

$$D[\widehat{\boldsymbol{B}}] = \frac{n-k-1}{n-k-p+q-1}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}. \tag{2.4}$$

The dispersion matrix of $\boldsymbol{B}_N$ is given by

$$D[\boldsymbol{B}_N] = (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}. \tag{2.5}$$

From here:

$$D[\widehat{\boldsymbol{B}}] - D[\boldsymbol{B}_N] = \frac{p-q}{n-k-p+q-1}(\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1},$$

which is positive definite. Thus $\boldsymbol{B}_N$ underestimates the variation of $\widehat{\boldsymbol{B}}$. This can be expected as the random matrix $\boldsymbol{S}$ in $\widehat{\boldsymbol{B}}$ has been replaced by $\boldsymbol{\Sigma}$ in $\boldsymbol{B}_N$. Let us rewrite $\widehat{\boldsymbol{B}}$ in the following form:

$$\begin{aligned}\widehat{\boldsymbol{B}} =&(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \\ =&(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \\ &+(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}(\boldsymbol{I}-\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1})\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}.\end{aligned} \tag{2.6}$$

The two terms in (2.6) are independent, what follows from the next result (Kollo & von Rosen, 2005, p. 196).

Lemma 2.1. *Let $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$, $\boldsymbol{Y} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\Psi})$ and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{K}$ and $\boldsymbol{L}$ be non-random matrices of proper sizes. Then,*

(i) *$\boldsymbol{A}\boldsymbol{X}\boldsymbol{K}$ is independent of $\boldsymbol{C}\boldsymbol{X}\boldsymbol{L}$ for all constant matrices $\boldsymbol{K}$ and $\boldsymbol{L}$ if and only if $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{C}' = \boldsymbol{0}$;*

(ii) *$\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}'$ is independent of $\boldsymbol{Y}\boldsymbol{B}$ if and only if $\boldsymbol{B}'\boldsymbol{\Psi}\boldsymbol{A}'\boldsymbol{\Psi} = \boldsymbol{0}$ and $\boldsymbol{B}'\boldsymbol{\Psi}\boldsymbol{A}\boldsymbol{\Psi} = \boldsymbol{0}$.*

Lemma 2.1 (i) yields that

$$(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}$$

and

$$(\boldsymbol{I}-\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1})\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}$$

are independently distributed. Moreover, by Lemma 2.1 (ii) the sums of square matrices $\boldsymbol{S}$ and $\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1}$ are independent.

Therefore

$$\boldsymbol{U} = (\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{S}^{-1}(\boldsymbol{I}-\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Sigma}^{-1})\boldsymbol{X}\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{C}')^{-1} \tag{2.7}$$

is independent of $\boldsymbol{B}_N$.

Thus we can represent $\widehat{\boldsymbol{B}}$ as a sum $\widehat{\boldsymbol{B}} = \boldsymbol{B}_N + \boldsymbol{U}$, where $\boldsymbol{B}_N$ and $\boldsymbol{U}$ are independent. In Kollo & von Rosen (1998) a general relation between two multivariate density functions was given. We formulate this in the next theorem. When the approximating density is $N_{pq}(\boldsymbol{M}, \boldsymbol{\Sigma})$, the obtained density expansion is given in the next theorem.

Theorem 2.2. *If $\boldsymbol{Y}$ is a random $p \times q$-matrix with finite first four moments, then the density $f_{\boldsymbol{Y}}(\boldsymbol{X})$ can be presented through the density $f_N(\boldsymbol{X})$ of the distribution $N_{pq}(\boldsymbol{M}, \boldsymbol{\Sigma})$ by the following formal matrix Edgeworth type expansion:*

$$\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{X}) = f_N(\boldsymbol{X}) \Big\{ & 1 + E[\mathrm{vec}(\boldsymbol{Y} - \boldsymbol{M})]' \mathrm{vec}\boldsymbol{H}_1(\mathrm{vec}\boldsymbol{X}, \mathrm{vec}\boldsymbol{M}, \boldsymbol{\Sigma}) \\
& + \frac{1}{2} \mathrm{vec}' \{ D[\mathrm{vec}\boldsymbol{Y}] - \boldsymbol{\Sigma} + E[\mathrm{vec}(\boldsymbol{Y} - \boldsymbol{M})] E[\mathrm{vec}(\boldsymbol{Y} - \boldsymbol{M})]' \} \\
& \times \mathrm{vec}\boldsymbol{H}_2(\mathrm{vec}\boldsymbol{X}, \mathrm{vec}\boldsymbol{M}, \boldsymbol{\Sigma}) \\
& + \frac{1}{6} \Big(\mathrm{vec}'(c_3[\boldsymbol{Y}] + 3\mathrm{vec}'(D[\mathrm{vec}\boldsymbol{Y}] - \boldsymbol{\Sigma}) \\
& \otimes E[\mathrm{vec}'(\boldsymbol{Y} - \boldsymbol{M})] + E[\mathrm{vec}'(\boldsymbol{Y} - \boldsymbol{M})]^{\otimes 3} \Big) \\
& \times \mathrm{vec}\boldsymbol{H}_3(\mathrm{vec}\boldsymbol{X}, \mathrm{vec}\boldsymbol{M}, \boldsymbol{\Sigma}) + \cdots \Big\}.
\end{aligned} \tag{2.8}$$

In the above density expansion the first three multivariate Hermite polynomials appear:

$$\begin{aligned}
& \boldsymbol{H}_1(\mathrm{vec}\boldsymbol{X}, \mathrm{vec}\boldsymbol{M}, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1} \mathrm{vec}(\boldsymbol{X} - \boldsymbol{M}); \\
& \boldsymbol{H}_2(\mathrm{vec}\boldsymbol{X}, \mathrm{vec}\boldsymbol{M}, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1} \mathrm{vec}(\boldsymbol{X} - \boldsymbol{M}) \mathrm{vec}'(\boldsymbol{X} - \boldsymbol{M}) \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}; \\
& \boldsymbol{H}_3(\mathrm{vec}\boldsymbol{X}, \mathrm{vec}\boldsymbol{M}, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1} \mathrm{vec}(\boldsymbol{X} - \boldsymbol{M}) (\mathrm{vec}'(\boldsymbol{X} - \boldsymbol{M}) \boldsymbol{\Sigma}^{-1})^{\otimes 2} \\
& - \boldsymbol{\Sigma}^{-1} \mathrm{vec}(\boldsymbol{X} - \boldsymbol{M}) \mathrm{vec}' \boldsymbol{\Sigma}^{-1} - \{ \mathrm{vec}'(\boldsymbol{X} - \boldsymbol{M}) \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} \} \\
& - \boldsymbol{\Sigma}^{-1} \otimes \{ \mathrm{vec}'(\boldsymbol{X} - \boldsymbol{M}) \boldsymbol{\Sigma}^{-1} \}.
\end{aligned} \tag{2.9}$$

For the definition and derivation of multivariate Hermite polynomials the interested reader is referred to Kollo & von Rosen (2005, §2.2.4).

From Theorem 2.2 we get the following formal approximation for the density of $\widehat{\boldsymbol{B}}$.

Theorem 2.3. *Let $\widehat{\boldsymbol{B}}$, $\boldsymbol{B}_N$ and $\boldsymbol{U}$ be given by* (1.3), (2.2) *and* (2.7), *respectively. Then an Edgeworth type expansion of the density of $\widehat{\boldsymbol{B}}$ equals*

$$f_{\widehat{\boldsymbol{B}}}(\boldsymbol{B}_0) = f_{\boldsymbol{B}_E}(\boldsymbol{B}_0) + \cdots ,$$

where

$$f_{\boldsymbol{B}_E}(\boldsymbol{B}_0) = f_{\boldsymbol{B}_N}(\boldsymbol{B}_0) \Big\{ 1 + \frac{1}{2} s \left(\mathrm{tr}\{ \boldsymbol{A}' \boldsymbol{\Sigma}^{-1} \boldsymbol{A} (\boldsymbol{B}_0 - \boldsymbol{B}) \boldsymbol{C} \boldsymbol{C}' (\boldsymbol{B}_0 - \boldsymbol{B})' \} - kq \right) \Big\}, \tag{2.10}$$

with

$$s = \frac{p-q}{n-k-p+q-1}. \tag{2.11}$$

Proof. To prove the theorem we have to plug the expressions of the first two moments of $\boldsymbol{B}_N$ and $\widehat{\boldsymbol{B}}$, given by equalities (2.3)-(2.5), into the formal approximation in Theorem 2.2.

From (2.8) and (2.9) we get

$$\begin{aligned} f_{\widehat{\boldsymbol{B}}}(\boldsymbol{B}_0) = f_{\boldsymbol{B}_N}(\boldsymbol{B}_0)\Big\{1 + \frac{1}{2}\mathrm{vec}'\{D[\widehat{\boldsymbol{B}}] - (\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\} \\ \times \mathrm{vec}\big\{(\boldsymbol{C}\boldsymbol{C}') \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})\mathrm{vec}(\boldsymbol{B}_0 - \boldsymbol{B})\mathrm{vec}'(\boldsymbol{B}_0 - \boldsymbol{B}) \\ \times (\boldsymbol{C}\boldsymbol{C}') \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}) - (\boldsymbol{C}\boldsymbol{C}') \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})\big\} + \ldots\Big\}. \end{aligned} \tag{2.12}$$

Using a property of the vec-operator

$$\mathrm{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\mathrm{vec}\boldsymbol{B}$$

and taking into account (2.4) and (2.5) we get

$$\begin{aligned} f_{\widehat{\boldsymbol{B}}}(\boldsymbol{B}_0) = f_{\boldsymbol{B}_N}(\boldsymbol{B}_0)\Big\{1 + \frac{1}{2}s\,\mathrm{vec}'\big((\boldsymbol{C}\boldsymbol{C}')^{-1} \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}\big) \\ \times \mathrm{vec}\big\{(\boldsymbol{C}\boldsymbol{C}') \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})\mathrm{vec}(\boldsymbol{B}_0 - \boldsymbol{B})\mathrm{vec}'(\boldsymbol{B}_0 - \boldsymbol{B}) \\ \times (\boldsymbol{C}\boldsymbol{C}') \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}) - (\boldsymbol{C}\boldsymbol{C}') \otimes (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})\big\} + \ldots\Big\}. \end{aligned} \tag{2.13}$$

Using the property of the trace function

$$\mathrm{tr}(\boldsymbol{A}'\boldsymbol{B}) = \mathrm{vec}'\boldsymbol{A}\mathrm{vec}\boldsymbol{B}$$

we have

$$\begin{aligned} f_{\widehat{\boldsymbol{B}}}(\boldsymbol{B}_0) &= f_{\boldsymbol{B}_N}(\boldsymbol{B}_0)\Big\{1 + \frac{1}{2}s\,\big(\mathrm{tr}\{\mathrm{vec}(\boldsymbol{B}_0 - \boldsymbol{B}) \\ &\quad \times \mathrm{vec}'(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{B}_0 - \boldsymbol{B})\boldsymbol{C}\boldsymbol{C}')\} - \boldsymbol{I}_k \otimes \boldsymbol{I}_q\big) + \ldots\Big\} \\ &= f_{\boldsymbol{B}_N}(\boldsymbol{B}_0)\Big\{1 + \frac{1}{2}s\,\big(\mathrm{tr}\{\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{B}_0 - \boldsymbol{B})\boldsymbol{C}\boldsymbol{C}'(\boldsymbol{B}_0 - \boldsymbol{B})'\} - kq\big) \\ &\quad + \ldots\Big\} \end{aligned} \tag{2.14}$$

what completes the proof. □

Following the ideas of Fujikoshi (1987), Kollo & von Rosen (2005, §4.3.2) have shown that $f_{\boldsymbol{B}_E}(\boldsymbol{B}_0)$ is a good approximation to the density $f_{\widehat{\boldsymbol{B}}}(\boldsymbol{B}_0)$:

Corollary 2.3.1.

$$\left| f_{\boldsymbol{B}_E}(\boldsymbol{B}_0) - f_{\widehat{\boldsymbol{B}}}(\boldsymbol{B}_0) \right| = O(n^{-2}).$$

3 Mixture of normal and Kotz distribution

Before presenting properties of the approximation for density (2.10), we give some basic facts about the Kotz-type distributions (see for example, Fang et al., 1990).

Definition 3.1. Let $\boldsymbol{x} = (X_1, \ldots, X_p)'$ be a random $p-$vector. The vector $\boldsymbol{x}$ has a *Kotz-type distribution* with the parameters $\boldsymbol{\mu}$, $\boldsymbol{V}$, N, s, r, if the density $f_{\boldsymbol{x}}(\boldsymbol{x})$ of $\boldsymbol{x}$ has a form:

$$\begin{aligned} f_{\boldsymbol{x}}(\boldsymbol{x}) = C_p |\boldsymbol{V}|^{-\frac{1}{2}} [(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^{N-1} \\ \times \exp\{(-r[(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})]^s)\}, \quad r,s>0,\ 2N+p>2, \end{aligned} \tag{3.1}$$

where C_p is a normalizing constant:

$$C_p = \frac{s\Gamma(p/2)}{\pi^{p/2}\Gamma(2N+p-2/2s)} r^{(2N+p-2/2s)}.$$

If we look upon the density function as a univariate function of the quadratic form $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, we get a function $g(u)$, which is called a *density generator*:

$$g(u) = C_p u^{N-1} \exp(-ru^s), \quad r,s>0,\ 2N+p>2. \tag{3.2}$$

We write $\boldsymbol{x} \sim K_p(\mu, \boldsymbol{V}, N, s, r)$ if $\boldsymbol{x}$ has the density function (3.1). When $N=1$, $s=1$ and $r=\frac{1}{2}$, we get the multivariate normal distribution. When $N=2$, $s=1$ and $r=\frac{1}{2}$, we get the distribution which we call Kotz distribution and denote $\boldsymbol{x} \sim K_p(\mu, \boldsymbol{V})$. Now and later on we shall follow this distinction also in the matrix case: the general distribution is called Kotz-type distribution and the special case for $N=2$, $s=1$ and $r=\frac{1}{2}$ is called the Kotz distribution.

Definition 3.2. We say that $\boldsymbol{Y} : p \times n$ has *the matrix Kotz-type distribution* with parameters $\boldsymbol{M}$, $\boldsymbol{V}$, $\boldsymbol{W}$, N, s and r, $\boldsymbol{Y} \sim K_{p,n}(\boldsymbol{M}, \boldsymbol{V}, \boldsymbol{W}, N, s, r)$ if

$$\boldsymbol{Y} = \boldsymbol{M} + \boldsymbol{\delta} \boldsymbol{X} \boldsymbol{\gamma}',$$

where $\boldsymbol{M} : p \times n$ is a constant matrix, matrices $\boldsymbol{\delta} : p \times p$, $\boldsymbol{V} = \boldsymbol{\delta}\boldsymbol{\delta}'$, $\boldsymbol{\gamma} : n \times n$, $\boldsymbol{W} = \boldsymbol{\gamma}\boldsymbol{\gamma}'$ are full rank and $\boldsymbol{X} : p \times n$ is a random matrix, so that vec$\boldsymbol{X} \sim K_{p,n}(\boldsymbol{0}, \boldsymbol{I}_{pn}, N, s, r)$. We say that $\boldsymbol{X} : p \times n$ has *the matrix spherical Kotz-type distribution* $K_{p,n}(\boldsymbol{0}, \boldsymbol{I}_p, \boldsymbol{I}_n, N, s, r)$.

We say that the matrix $\boldsymbol{Y}$ is Kotz distributed if $N = 2$, $s = 1$ and $r = \frac{1}{2}$ and denote $\boldsymbol{Y} \sim K_{p,n}(\boldsymbol{M}, \boldsymbol{V}, \boldsymbol{W})$.

Using Definition 3.2 and some properties of the vec-operator we get the density function for the matrix Kotz-distribution:

$$f_{\boldsymbol{Y}}(\boldsymbol{Y}) = |\boldsymbol{V}|^{-\frac{n}{2}} |\boldsymbol{W}|^{-\frac{p}{2}} g(\operatorname{tr}\{\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{M})\boldsymbol{W}^{-1}(\boldsymbol{Y} - \boldsymbol{M})'\}),$$

where $g(\cdot)$ is the density generating function of Kotz-type distributions, given in (3.2).

Now we start to examine the approximation $f_{\boldsymbol{B}_E}$.

Theorem 3.1. *The function* $f_{\boldsymbol{B}_E}$ *in* (2.10) *is a density function, if*

$$0 < 1 - \frac{1}{2} skq < 1. \tag{3.3}$$

Proof. From the assumption (3.3) $f_{\boldsymbol{B}_E}(\boldsymbol{B}) \geq 0$ for any $\boldsymbol{B}$. Let us integrate the function $f_{\boldsymbol{B}_E}$:

$$\begin{aligned}
&\int_{R^{q \times k}} f_{\boldsymbol{B}_E}(\boldsymbol{B}_0) d\boldsymbol{B}_0 = \int_{R^{q \times k}} \Big\{ 1 + \frac{1}{2} s \ \big(\operatorname{tr}\{\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{B}_0 - \boldsymbol{B})\boldsymbol{C}\boldsymbol{C}' \\
&\times\ (\boldsymbol{B}_0 - \boldsymbol{B})'\} - kq\big) \Big\} f_{\boldsymbol{B}_N}(\boldsymbol{B}_0) d\boldsymbol{B}_0 = (1 - \frac{1}{2} skq) \int_{R^{q \times k}} f_{\boldsymbol{B}_N}(\boldsymbol{B}_0) d\boldsymbol{B}_0 \\
&+ \frac{1}{2} skq \int_{R^{q \times k}} \frac{1}{kq} \big(\operatorname{tr}\{\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{B}_0 - \boldsymbol{B})\boldsymbol{C}\boldsymbol{C}'(\boldsymbol{B}_0 - \boldsymbol{B})'\}\big) f_{\boldsymbol{B}_N}(\boldsymbol{B}_0) d\boldsymbol{B}_0 \\
&= 1,
\end{aligned}$$

as obviously

$$\int_{R^{q \times k}} f_{\boldsymbol{B}_N}(\boldsymbol{B}_0) d\boldsymbol{B}_0 = 1,$$

and

$$\int_{R^{q\times k}} \frac{1}{kq}\left(\mathrm{tr}\{\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{B}_0-\boldsymbol{B})\boldsymbol{C}\boldsymbol{C}'(\boldsymbol{B}_0-\boldsymbol{B})'\}\right) f_{\boldsymbol{B}_N}(\boldsymbol{B}_0)d\boldsymbol{B}_0 = 1,$$

because we integrate the density function of the matrix Kotz distribution $K_{q,k}(\boldsymbol{B},(\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1},(\boldsymbol{C}\boldsymbol{C}')^{-1})$. □

Corollary 3.1.1. *The distribution of $\boldsymbol{B}_E$ given in* (2.10) *is a mixture of a normal distribution and a Kotz distribution with weights $1-\frac{1}{2}skq$ and $\frac{1}{2}skq$ respectively, if the assumption* (3.3) *holds, where s is defined in* (2.11)*. The dimensions k and q are fixed in the model* (1.2)*.*

Proof. The statement follows straightforwardly from the proof of Theorem 3.1. □

The Kotz distribution is a multimodal distribution, see Kollo & Roos (2005), for example. The mixture of the normal distribution and the Kotz distribution can be unimodal if the weight of the normal distribution is large enough. The problem is studied in the next theorem.

Theorem 3.2. *Let the distribution of a random matrix $\boldsymbol{X}$ be the mixture of the matrix spherical Kotz distribution $K_{p,n}(\boldsymbol{0},\boldsymbol{I}_p,\boldsymbol{I}_n)$ and standard matrix normal distribution $N_{p,n}(\boldsymbol{0},\boldsymbol{I}_p,\boldsymbol{I}_n)$ with positive weights λ and $1-\lambda$. The density function of the matrix $\boldsymbol{X}$ is unimodal if and only if*

$$\lambda \le \frac{np}{2+np}.$$

Proof. Let us denote $\boldsymbol{x} = \mathrm{vec}\boldsymbol{X}$. Then we can write the density function of the random matrix $\boldsymbol{X}$ as follows:

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{np}{2}}}(1-\lambda+\frac{\lambda}{np}\boldsymbol{x}'\boldsymbol{x})\exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right). \tag{3.4}$$

We need to find extrema of the function $f(\boldsymbol{x})$. For that we find the derivative of the function (3.4).

Differentiating the product in the density expression we get:

$$\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = \frac{1}{(2\pi)^{\frac{np}{2}}}\left(\frac{\lambda}{np}\frac{d\boldsymbol{x}'\boldsymbol{x}}{d\boldsymbol{x}}\exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right) + \frac{d\exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right)}{d\boldsymbol{x}}(1-\lambda+\frac{\lambda}{np}\boldsymbol{x}'\boldsymbol{x})\right).$$

Now we calculate the needed derivatives:

$$\frac{d\boldsymbol{x}'\boldsymbol{x}}{d\boldsymbol{x}} = 2\boldsymbol{x};$$
$$\frac{d\exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right)}{d\boldsymbol{x}} = \frac{d\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right)}{d\boldsymbol{x}}\frac{d\exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right)}{d\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right)} = -\exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right)\boldsymbol{x}.$$

The desired derivative has the following form:

$$\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = \frac{1}{(2\pi)^{\frac{np}{2}}}\exp\left(-\frac{1}{2}\boldsymbol{x}'\boldsymbol{x}\right)\left(\frac{2\lambda}{np} - 1 + \lambda - \frac{\lambda}{np}\boldsymbol{x}'\boldsymbol{x}\right)\boldsymbol{x}.$$

One extremum is always obtained at $\boldsymbol{x} = \boldsymbol{0}$. Additionally, the derivative can be zero if

$$\frac{2\lambda}{np} - 1 + \lambda - \frac{\lambda}{np}\boldsymbol{x}'\boldsymbol{x} = 0 \quad \Longleftrightarrow \quad \boldsymbol{x}'\boldsymbol{x} = \frac{2\lambda + \lambda np - np}{\lambda},$$

which has no solution, if

$$\frac{2\lambda + \lambda np - np}{\lambda} < 0 \quad \Longleftrightarrow \quad \lambda \leq \frac{np}{2+np}.$$

This proves the theorem. □

Considering the non-spherical mixture of the distributions $N_{p,n}(\boldsymbol{M}, \boldsymbol{V}, \boldsymbol{W})$ and $K_{p,n}(\boldsymbol{M}, \boldsymbol{V}, \boldsymbol{W})$, the condition for unimodality remains the same, because the mean and dispersion of the distribution do not have an effect on modality.

Corollary 3.2.1. *Consider the model* (1.2). *The density function of $\boldsymbol{B}_E$ given in* (2.10) *is unimodal if and only if $s < 2(2+kq)^{-1}$.*

Proof. From Corollary 3.1.1 we know that the distribution of the estimator $\boldsymbol{B}_E$ is a mixture of a Kotz distribution and a normal distribution with weights $\frac{1}{2}skq$ and $1 - \frac{1}{2}skq$, respectively. The dimension of the matrix $\boldsymbol{B}_E$ is $q \times k$. From Theorem 3.2 the following relation must hold for unimodality:

$$\frac{1}{2}skq < \frac{kq}{2+kq} \Longrightarrow s < 2(2+kq)^{-1}.$$

□

Next we consider the marginal distribution of the mixture of a Kotz distribution and a normal distribution. We shall see that when considering the marginals of the mixture, the distribution of the marginals is a similar mixture with a smaller weight for Kotz distribution. So when we reduce the dimensions of the random matrix the similarity to the normal distribution increases. This result shows, that in low-dimensional cases the approximation by normal distribution may be appropriate but in higher dimensions the approximation is not so good anymore. We also make use of this theorem in Section 4 in a simulation study.

Theorem 3.3. *Let a random matrix $\boldsymbol{X}$ be a mixture of random matrices $\boldsymbol{X}_N$ and $\boldsymbol{X}_K$ with weights $1-\lambda$ and λ respectively, where the matrix $\boldsymbol{X}_N \sim N_{p,n}(\boldsymbol{M}, \boldsymbol{V}, \boldsymbol{W})$ and the matrix $\boldsymbol{X}_K \sim K_{p,n}(\boldsymbol{M}, \boldsymbol{V}, \boldsymbol{W})$. Let*

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{X}_2 \\ \tilde{\boldsymbol{x}}_1 & \end{pmatrix},$$

where $\boldsymbol{x}_1 : p_1 \times 1$, $\tilde{\boldsymbol{x}}_1 : (p-p_1) \times 1$ and $\boldsymbol{X}_2 : p \times (n-1)$, let

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{\mu}_1 & \boldsymbol{M}_2 \\ \tilde{\boldsymbol{\mu}}_1 & \end{pmatrix}$$

be with the same structure as $\boldsymbol{X}$ and the matrix

$$\boldsymbol{\Sigma} = \boldsymbol{V} \otimes \boldsymbol{W} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Then the vector $\boldsymbol{x}_1$ is a mixture of random vectors $\boldsymbol{x}_{1N}$ and $\boldsymbol{x}_{1K}$ with weights $1-\lambda p_1/(np)$ and $\lambda p_1/(np)$ respectively, where $\boldsymbol{x}_{1N} \sim N_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{x}_{1K} \sim K_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

Proof. Let us denote $\boldsymbol{x} = \mathrm{vec}\boldsymbol{X}$, $\boldsymbol{x}_2' = (\tilde{\boldsymbol{x}}_1', \mathrm{vec}'\boldsymbol{X})$, $\boldsymbol{\mu} = \mathrm{vec}\boldsymbol{M}$, $\boldsymbol{\mu}_2' = (\tilde{\boldsymbol{\mu}}_1', \mathrm{vec}'\boldsymbol{M}_2)$ and $p_2 = np - p_1$. The density of $\boldsymbol{x}$ has the form:

$$f_{\boldsymbol{x}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{np}{2}}} |\boldsymbol{\Sigma}|^{-1} \exp(-\frac{1}{2}a) \left(\frac{\lambda a}{np} + 1 - \lambda \right),$$

where

$$a = (\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}).$$

Let us denote $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{\mu}$, $\boldsymbol{z}_1 = \boldsymbol{x}_1 - \boldsymbol{\mu}_1$ and $\boldsymbol{z}_2 = \boldsymbol{x}_2 - \boldsymbol{\mu}_2$.

To obtain the density function for $\boldsymbol{x}_1$ we need to integrate the density function of the vector $\boldsymbol{x}$ by the vector $\boldsymbol{x}_2$. Next we change the form of some expressions in the density of $\boldsymbol{x}$, to make the integration easier.

We shall make use of the representation of the inverse matrix (Horn & Johnson, 1990, p 31):

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}\Rightarrow \boldsymbol{A}^{-1}=\begin{pmatrix}\boldsymbol{A}_{11}^{-1}+\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22\cdot 1}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22\cdot 1}^{-1}\\ -\boldsymbol{A}_{22\cdot 1}^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{A}_{22\cdot 1}^{-1}\end{pmatrix},$$

where

$$\boldsymbol{A}_{22\cdot 1}=\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}.$$

Now we get:

$$\boldsymbol{\Sigma}^{-1}=\begin{pmatrix}\boldsymbol{\Sigma}_{11}^{-1}+\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\\ -\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{\Sigma}_{22\cdot 1}^{-1}\end{pmatrix},$$

We calculate the expression for a by partitioned matrices:

$$\begin{aligned} a &= \boldsymbol{z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{z}\\ &= \boldsymbol{z}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1+(\boldsymbol{\Sigma}_{22\cdot 1}^{-\frac{1}{2}}\boldsymbol{z}_2-\boldsymbol{\Sigma}_{22\cdot 1}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1)'(\boldsymbol{\Sigma}_{22\cdot 1}^{-\frac{1}{2}}\boldsymbol{z}_2-\boldsymbol{\Sigma}_{22\cdot 1}^{-\frac{1}{2}}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1).\end{aligned}$$

The determinant of $\boldsymbol{\Sigma}$ can be found by the formula:

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_{11} & \boldsymbol{A}_{12}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{pmatrix}\Longrightarrow |\boldsymbol{A}|=|\boldsymbol{A}_{11}||\boldsymbol{A}_{22}-\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}|.$$

From here we get:

$$|\boldsymbol{\Sigma}|=|\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22\cdot 1}|.$$

After a change of variables in the original density expression of $\boldsymbol{x}$:

$$\begin{aligned}&\boldsymbol{z}_2=\boldsymbol{x}_2-\boldsymbol{\mu}_2;\\ &\boldsymbol{y}=\boldsymbol{z}_2-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1\Rightarrow \boldsymbol{z}_2=\boldsymbol{y}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1\Rightarrow d\boldsymbol{y}=d\boldsymbol{z}_2\Rightarrow J=1,\end{aligned}$$

we get:

$$\begin{aligned} f_{Z_1}(\boldsymbol{z}_1) &= \frac{1}{(2\pi)^{\frac{np}{2}}}|\boldsymbol{\Sigma}|^{-1}\exp(-\frac{1}{2}\boldsymbol{z}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1)\int_{\mathbb{R}^{p_2}}\exp(-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\boldsymbol{y})\\ &\quad\left(\frac{\lambda}{np}\boldsymbol{z}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1+\frac{\lambda}{np}\boldsymbol{y}'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\boldsymbol{y}+1-\lambda\right)d\boldsymbol{y}\\ &= \frac{1}{(2\pi)^{\frac{np}{2}}}|\boldsymbol{\Sigma}|^{-1}\exp(-\frac{1}{2}\boldsymbol{z}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1)\left[\left(\frac{\lambda}{np}\boldsymbol{z}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1+1-\lambda\right)\right.\\ &\quad\left.\times(2\pi)^{\frac{p_2}{2}}|\boldsymbol{\Sigma}_{22\cdot 1}|+\frac{\lambda}{p}(2\pi)^{\frac{p_2}{2}}|\boldsymbol{\Sigma}_{22\cdot 1}|E(\boldsymbol{y}'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\boldsymbol{y})\right].\end{aligned}$$

Applying the equalities $D\boldsymbol{y} = \boldsymbol{\Sigma}_{22\cdot 1}$ and $E\boldsymbol{y} = \boldsymbol{0}$, we get $E(\boldsymbol{y}'\boldsymbol{\Sigma}_{22\cdot 1}^{-1}\boldsymbol{y}) = p_2$. That implies:

$$f_{Z_1}(\boldsymbol{z}_1) = \; = \frac{1}{(2\pi)^{\frac{p_1}{2}}}|\boldsymbol{\Sigma}_{11}|^{-1}\exp(-\frac{1}{2}\boldsymbol{z}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1)\left(\frac{\lambda}{np}\boldsymbol{z}_1'\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{z}_1 + 1 - \frac{\lambda p_1}{np}\right).$$

Returning to the original variables $\boldsymbol{z}_1 = \boldsymbol{x}_1 - \boldsymbol{\mu}_1$, the statement is proved. $\square$

4 Simulation

Here we study a low-dimensional example. Consider two groups, both consisting of five objects. Let each object be measured three times. Let each object in the first group have a theoretical mean vector $\boldsymbol{\mu}_1 = (1,2,3)'$ and each object in the second group $\boldsymbol{\mu}_2 = (2,4,6)'$. Let the correlation between two consecutive observations be 0.5 and between the first and the third observation 0.25. Consider the data matrix $\boldsymbol{X} \sim N_{3,10}(\boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{I}_{10})$, where

$$\boldsymbol{M} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 & 2 & 4 & 4 & 4 & 4 & 4 \\ 3 & 3 & 3 & 3 & 3 & 6 & 6 & 6 & 6 & 6 \end{pmatrix}; \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix}.$$

The following model is assumed to hold:

$$\mu_{i,t} = \beta_{i,0} + \beta_{i,1}t,$$

where $i = 1,2$, $t = 1,2,3$. Then in matrix form, as given in (1.2), the design matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ are:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}; \quad \boldsymbol{C} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

In that case obviously the real value of $\boldsymbol{B}$ is known:

$$\boldsymbol{B} = \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix}.$$

The approximation for the distribution of the estimator of $\boldsymbol{B}$ in our example is a mixture of the normal distribution $N_{2,2}(\boldsymbol{B}, \boldsymbol{V}, \boldsymbol{W})$ and the Kotz

Table 4.1: Characteristics of the elements of the matrix $\widehat{\boldsymbol{B}}$

	mean	variance
β_{01}	0	0.42
β_{11}	1	0.075
β_{02}	0	0.42
β_{12}	2	0.075

distribution $K_{2,2}(\boldsymbol{B}, \boldsymbol{V}, \boldsymbol{W})$, where $\boldsymbol{V} = (\boldsymbol{A}'\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}$ and $\boldsymbol{W} = (\boldsymbol{C}\boldsymbol{C}')^{-1}$. These matrices equal:

$$\boldsymbol{V} = \begin{pmatrix} 2.1 & -0.75 \\ -0.75 & 0.375 \end{pmatrix}, \quad \boldsymbol{W} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}.$$

The weights of the mixture can be found using (2.11). By the formula we get 2/3 as the weight of the normal distribution and 1/3 to be the weight of the Kotz distribution.

To check the theory we simulated 100,000 samples (using the software package R), estimated the parameter matrix $\boldsymbol{B}$ from each sample and examined the obtained empirical distribution.

To study the distribution visually we use the marginals of the distribution of $\widehat{\boldsymbol{B}}$, which can be plotted in two dimensions. The parameters of the univariate marginals can be found using Theorem 3.3. The distribution of the vector $\mathrm{vec}\widehat{\boldsymbol{B}}$ is a mixture of $N_4(\mathrm{vec}\boldsymbol{B}, \boldsymbol{V}_v)$ and $K_4(\mathrm{vec}\boldsymbol{B}, \boldsymbol{V}_v)$, where

$$\boldsymbol{V}_v = \boldsymbol{V} \otimes \boldsymbol{W} = \begin{pmatrix} 0.42 & -0.15 & 0 & 0 \\ -0.15 & 0.075 & 0 & 0 \\ 0 & 0 & 0.42 & -0.15 \\ 0 & 0 & -0.15 & 0.075 \end{pmatrix},$$

with weights 2/3 and 1/3, respectively.

Using Theorem 3.3, the distribution of each element of the matrix $\widehat{\boldsymbol{B}}$ is a mixture of the Kotz distribution and the normal distribution with weights 1/12 and 11/12 respectively. The parameters of the described distributions are given in Table 4.1.

The density functions of one-dimensional marginals for the generated data and the theoretical density functions are given in Figure 4.1.

The data is presented in the histogram whereas the underlying theoretical density function follows the included curve. Clearly, the estimated density of

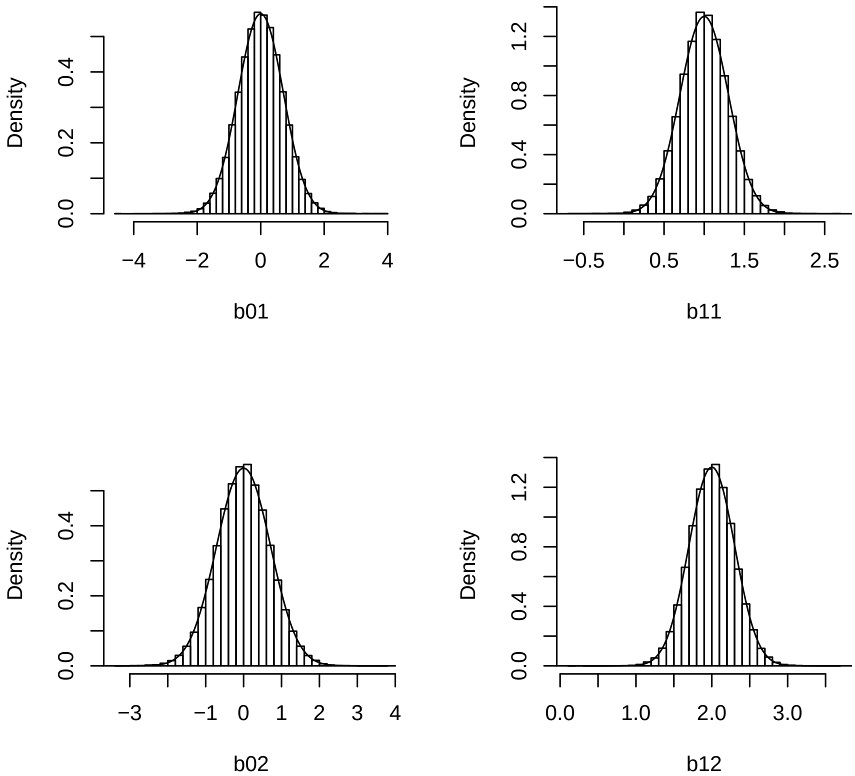

Figure 4.1: The marginals of the distribution of $\widehat{\boldsymbol{B}}$

the generated data follows the theoretical density well. Moreover we can see, that when we look at the one-dimensional marginals the true distribution is not very far from being normal, but this does not hold in higher dimensions.

Acknowledgement

T. Kollo and A. Roos are thankful to Estonian Science Foundation for financial support trough the grant GMTMS 5686 and the project TMTMS 1776.

References

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, Third edition. Wiley, New York.

Fang, K.-T., Kotz, S. & Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions.* Monographs on Statistics and Applied Probability, **36**, Chapman and Hall, London.

Fujikoshi, Y. (1985). An error bound for an asymptotic expansion of the distribution function of an estimate in a multivariate linear model. *Ann. Statist.* **13** 827–831.

Fujikoshi, Y. (1987). Error bounds for asymptotic expansions of the distribution of the MLE in a gmanova model. *Ann. Inst. Statist. Math.* **39** 153-161.

Gleser, L.J. & Olkin, I. (1970). Linear models in multivariate analysis. In: *Essays in Probability and Statistics.* University of North Carolina, Chapel Hill, 267–292.

Horn, R.A. & Johnson, C.R. (1990). *Matrix Analysis.* Cambridge University Press, Cambridge.

Kabe, D.G. (1975). Some results for the GMANOVA model. *Comm. Statist.* **4** 813–820.

Kenward, M.G. (1986). The distribution of a generalized least squares estimator with covariance adjustment. *J. Multivariate Anal.* **20** 244–250.

Khatri, C.G. (1966). A note on a Manova model applied to problems in growth curve. *Ann. Inst. Statist. Math.* **18** 75–86.

Kollo, T. & Roos, A. (2005) On Kotz-type elliptical distributions. In: *Contemporary Multivariate Analysis and Experimental Designs - In Celebration of Professor Kai-Tai Fang's 65th Birthday.* Eds. J. Fan and G. Li. World Scientific, New Jersey, 271–282.

Kollo, T. & von Rosen, D. (1998). A unified approach to the approximation of multivariate densities. *Scand. J. Statist.* **25** 93–109.

Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices.* Springer, Dordrecht.

Kshirsagar, A.M. & Smith, W.B. (1995). *Growth Curves.* Marcel Dekker, New York.

Potthoff, R.F. & Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51** 313–326.

von Rosen, D. (1991). The growth curve model: A review. *Comm. Statist. Theory Methods* **20** 2791–2822.

Srivastava, M.S. & von Rosen, D. (1999). Growth curve models. In: *Multivariate Analysis, Design of Experiments, and Survey Sampling*, Ed. S. Ghosh. Marcel Dekker, New York, 547–578.

Woolson, R.F. & Leeper, J.D. (1980). Growth curve analysis of complete and incomplete longitudinal data. *Comm. Statist. A, Theory Methods* **9** 1491–1513.