MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETER STRUCTURES FOR THE WISHART DISTRIBUTION USING CONSTRAINTS

H.F. Strydom[1]

Department of Statistics, University of Pretoria, Pretoria, 0002, South Africa.

e-mail: *nina.strydom@up.ac.za*

N.A.S. Crowther

Department of Statistics, University of Pretoria, Pretoria, 0002, South Africa.

e-mail: *nico.crowther@up.ac.za*

Keywords: Maximum likelihood estimation under constraints, Wishart distribution, patterned covariance matrices, covariance matrix with zeros, homogeneity of covariance matrices.

Summary: Maximum likelihood estimation under constraints for estimation in the Wishart class of distributions, is considered. It provides a unified approach to estimation in a variety of problems concerning covariance matrices. Virtually all covariance structures can be translated to constraints on the covariances. This includes covariance matrices with given structure such as linearly patterned covariance matrices, covariance matrices with zeros, independent covariance matrices and structurally dependent covariance matrices. The methodology followed in this paper provides a useful and simple approach to directly obtain the exact maximum likelihood estimates. These maximum likelihood estimates are obtained via an estimation procedure for the exponential class using constraints.

[1] Corresponding author

1 Introduction

Maximum likelihood estimation of parameters and parametric functions in the case of even standard distributions, belonging to the exponential class, is in certain cases difficult to perform or leads to intractible procedures which often imply that the likelihood function has to be solved numerically. In this paper, a maximum likelihood procedure which was developed by Matthews & Crowther (1995), is extended by using the canonical form of the Wishart distribution directly in order to obtain maximum likelihood estimates (mle's) of parameter structures in a simple and unified way. It is done by considering the likelihood function as a restricted case of a broader class of likelihood functions of which the mle's are readily available. Parameter structures can be translated to constraints on the parameters. These parameters are the elements of the covariance matrix of the original multivariate normal distribution.

In this paper the focus is specifically on equality constraints on parameters. This does not relate to the inequality constraints on parameters and parameter matrices usually required in the case of maximum likelihood estimation of multivariate variance components as, for example, considered by Calvin & Dykstra (1995).

Estimation of covariance matrices in a variety of settings where the computational aspects become problematic, is considered. This includes estimation of patterned covariance matrices, estimation of a covariance matrix with zeros and testing the homogeneity of the covariance matrices of dependent multivariate normals. Constraints are implied by the model under consideration. In particular, the expected value of a Wishart matrix can be estimated enabling applications to patterned covariance matrices where the pattern is not necessarily linear. For a covariance matrix, $\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$, the variances σ_{11} and σ_{22} may be required to be nonlinearly related, e.g. $7\sigma_{11}^2 = \sigma_{22}$. In stead of reparameterizing the covariance matrix by substituting σ_{22} by $7\sigma_{11}^2$ and deriving the likelihood equations accordingly, the relation is brought into the estimation process by directly imposing the constraint that $7\sigma_{11}^2 - \sigma_{22} = 0$ on the elements of the covariance matrix $\mathbf{\Sigma}$ following the methodology described in Section 2. This procedure also lends itself naturally to imposing the constraint that certain covariances are zero. The extension of this procedure to groups of variables is straightforward. Suppose $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$. In this case, structure is imposed on submatrices by direct constraints, e.g. $\mathbf{\Sigma}_{12} = \mathbf{0}$ for independence or $\mathbf{\Sigma}_{11} = \mathbf{\Sigma}_{22}$ for testing homogeneity - in the case of independence or dependence. This enables inference regarding covariance matrices of dependent or independent multivariate normals to be accomplished in a straightforward manner.

The approach considered in this paper largely simplifies the philosophy, the modelling and computational aspects involved in the case of maximum likelihood estimation of covariance matrices since no derivation of likelihood equations is required. Covariance matrices can be modelled by direct imposition of constraints on specific elements of the covariance matrix. In Section 2 the general theory for maximum likelihood estimation under constraints is presented for the exponential family. Estimation of covariance matrices in terms of constraints, as a special case of the general procedure, is considered in Section 3. A simple introductory example is given. In Section 4 covariance matrices with given structure, e.g. patterned covariance matrices and covariance matrices with zeros, are discussed and illustrated with numerical examples. In Section 5 the problem of testing for the homogeneity of covariance matrices of dependent multivariate normal samples is considered. The theoretical basis is given and then illustrated with a numerical example. The test for independence and a test for the homogeneity of covariance matrices of several dependent multivariate normals, are illustrated. It is shown how the latter can be adjusted to make provision for patterns in the covariance matrices at the same time. Concluding remarks are made in Section 6.

2 Maximum Likelihood Estimation under Constraints

The random vector $\boldsymbol{t} : k \times 1$ belongs to the canonical exponential family if its probability density function is of the form (Barndorff-Nielsen (1982) or Brown (1986))

$$p(\boldsymbol{t}, \boldsymbol{\theta}) = a(\boldsymbol{\theta})b(\boldsymbol{t})exp(\boldsymbol{\theta}'\boldsymbol{t}) = b(\boldsymbol{t})exp\left\{\boldsymbol{\theta}'\boldsymbol{t} - \kappa(\boldsymbol{\theta})\right\}, \qquad \boldsymbol{t} \in R^k, \qquad \boldsymbol{\theta} \in \aleph \tag{1}$$

where $\boldsymbol{\theta} : k \times 1$ the canonical (natural) parameter, $\boldsymbol{t} : k \times 1$ the canonical statistic and $\aleph$ the natural parameter space for the canonical parameter $\boldsymbol{\theta}$. The function $\kappa(\boldsymbol{\theta}) = -\ell n\ a(\boldsymbol{\theta})$ is referred to as the cumulant generating function or the log Laplace transform. It is important to note that the statistic $\boldsymbol{t}$ in (1) is a canonical statistic. If $\boldsymbol{t}$ is a sufficient statistic in the regular exponential class, it can be transformed to canonical form.

The mean vector and covariance matrix of $\boldsymbol{t}$ are given by $E(\boldsymbol{t}) = \frac{\partial}{\partial \boldsymbol{\theta}}\kappa(\boldsymbol{\theta}) = \boldsymbol{m}$ and $Cov(\boldsymbol{t}) = \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\kappa(\boldsymbol{\theta}) = \boldsymbol{V}$. In this case the mle of $\boldsymbol{m}$ without any constraints is $\widehat{\boldsymbol{m}} = \boldsymbol{t}$.

The mle's of $\boldsymbol{m}$ may in general not exist, except under constraints which are implied by a particular model. In this paper, the mle's are not obtained from maximizing a likelihood function in terms of parameters. The mle's $\widehat{\boldsymbol{m}}$ of $\boldsymbol{m}$ are obtained under a set of constraints, $\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{0}$, imposed on the expected values of the canonical statistics (cf. Matthews & Crowther, 1995, 1998). The function $\boldsymbol{g}(\boldsymbol{m})$ is a continuous vector valued function of $\boldsymbol{m}$, for which the first order partial derivatives exist. Then $\widehat{\boldsymbol{m}}$ may be obtained from

$$\widehat{\boldsymbol{m}} = \boldsymbol{t} - (\boldsymbol{G}_m \boldsymbol{V})' \left(\boldsymbol{G}_t \boldsymbol{V} \boldsymbol{G}_m'\right)^* \boldsymbol{g}(\boldsymbol{t}) \tag{2}$$

where $\boldsymbol{G}_m = \frac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}}$, $\boldsymbol{G}_t = \left.\frac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}=\boldsymbol{t}}$ and $\left(\boldsymbol{G}_t \boldsymbol{V} \boldsymbol{G}_m'\right)^*$ a generalized inverse of $\left(\boldsymbol{G}_t \boldsymbol{V} \boldsymbol{G}_m'\right)$. The expression is invariant with respect to the choice of the generalized inverse. Many seemingly complicated hypotheses regarding $\boldsymbol{m}$ may be imbedded into this framework in order to obtain the mle of $\boldsymbol{m}$ in a simple way.

In general, the iterative procedure implies a double iteration over $\boldsymbol{t}$ and $\boldsymbol{m}$. The first iteration stems from the Taylor series linearization of $\boldsymbol{g}(\boldsymbol{t})$ and the second from the fact that $\boldsymbol{V}$ may be a function of $\boldsymbol{m}$. The usual conditions for convergence, similar to those for Newton-Raphson, hold in both the iterations. If $\boldsymbol{g}(\boldsymbol{t})$ is a linear function, no first iteration is necessary. Similarly, if $\boldsymbol{V}$ and $\boldsymbol{G}_m$ are not functions of $\boldsymbol{m}$, no second iteration over $\boldsymbol{m}$ is necessary.

The procedure is initialized with the observed canonical statistic as the starting value for both $\boldsymbol{t}$ and $\boldsymbol{m}$. Convergence is attained first over $\boldsymbol{t}$ (which is immediate in the case of linear constraints on $\boldsymbol{m}$) and then over $\boldsymbol{m}$. The converged value of $\boldsymbol{t}$ is used as the next approximation for $\boldsymbol{m}$, with iteration over $\boldsymbol{m}$ starting at the observed $\boldsymbol{t}$. The covariance matrix $\boldsymbol{V}$ may be a function of $\boldsymbol{m}$, in which case it is recalculated for each new value of $\boldsymbol{m}$ in the iterative procedure. Convergence over $\boldsymbol{m}$ yields $\widehat{\boldsymbol{m}}$, the mle of $\boldsymbol{m}$ under the constraints. The estimation process is summarized by an algorithm given by Strydom & Crowther (2012) which is given here to elucidate the estimation process:

Algorithm 1 Obtaining mle's under constraints.

Step 1:	Specify $\boldsymbol{t}_0$, the vector of observed canonical statistics.
Step 2:	Let $\boldsymbol{t} = \boldsymbol{t}_0$.
Step 3A:	Let $\boldsymbol{m} = \boldsymbol{t}$, $\boldsymbol{t} = \boldsymbol{t}_0$.
	Calculate $\boldsymbol{G}_m$, $\boldsymbol{V}$.
Step 3B:	Let $\boldsymbol{t}_p = \boldsymbol{t}$.
	Calculate $\boldsymbol{g}(\boldsymbol{t})$, $\boldsymbol{G}_t$.
	Calculate $\boldsymbol{t} = \boldsymbol{t} - (\boldsymbol{G}_m\boldsymbol{V})'(\boldsymbol{G}_t\boldsymbol{V}\boldsymbol{G}_m')^*\boldsymbol{g}(\boldsymbol{t})$.
	If $(\boldsymbol{t} - \boldsymbol{t}_p)'(\boldsymbol{t} - \boldsymbol{t}_p) < \epsilon$ (a small positive number determining the accuracy)
	then go to Step 3A,
	else repeat Step 3B.
Step 4	If $(\boldsymbol{m} - \boldsymbol{t})'(\boldsymbol{m} - \boldsymbol{t}) < \epsilon$
	then convergence is attained.

The exact mle's are given by $\boldsymbol{m}$.

For the constrained model to be properly defined, the number of independent constraints, $\nu = rank(\boldsymbol{G}_m\boldsymbol{V}\boldsymbol{G}_m')$, should equal the difference between the number of elements of $\boldsymbol{m}$ and the number of unknown parameters of the constrained model. Constraints need not be independent and the expression is invariant with respect to the choice of the generalized inverse (Matthews & Crowther, 1998).

The asymptotic covariance matrix of $\widehat{\boldsymbol{m}}$ is given by

$$Cov(\widehat{\boldsymbol{m}}) = \boldsymbol{V} - (\boldsymbol{G}_m\boldsymbol{V})'\left(\boldsymbol{G}_m\boldsymbol{V}\boldsymbol{G}_m'\right)^*\boldsymbol{G}_m\boldsymbol{V} \tag{3}$$

which is estimated by replacing $\boldsymbol{m}$ with $\widehat{\boldsymbol{m}}$. The standard error of the estimates are given by $\boldsymbol{\sigma}(\widehat{\boldsymbol{m}})$, the square root of the vector of diagonal elements of $Cov(\widehat{\boldsymbol{m}})$.

The Wald statistic

$$W = \boldsymbol{g}'(\boldsymbol{t})(\boldsymbol{G}_t\boldsymbol{V}\boldsymbol{G}_t')^*\boldsymbol{g}(\boldsymbol{t}) \tag{4}$$

is a measure of goodness of fit under the hypothesis $\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{0}$. It is asymptotically χ^2 distributed with ν degrees of freedom.

In the next section it is shown how the theory discussed here for the exponential class, is specifically generalized to address estimation problems based on the Wishart distribution.

3 Estimation of Covariance Matrices in terms of Constraints

Suppose $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N$ is a random sample from a multivariate normal distribution, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\overline{\boldsymbol{y}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i$ denote the sample mean vector and $\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{N}(\boldsymbol{y}_i - \overline{\boldsymbol{y}})(\boldsymbol{y}_i - \overline{\boldsymbol{y}})'$ the sample covariance matrix with $n = N-1$. The sample covariance matrix $\boldsymbol{S}$, which is an unbiased estimate of $\boldsymbol{\Sigma}$ is $W_p(n, \boldsymbol{\Sigma}/n)$ distributed. The probability density function of $\boldsymbol{S}$ is

$$\begin{aligned} f(\boldsymbol{S}) &= \frac{det(\boldsymbol{S})^{\frac{n-p-1}{2}}\left(\frac{n}{2}\right)^{\frac{np}{2}}}{\Gamma_p(\frac{1}{2}n)} exp\left\{-\frac{n}{2}tr\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \frac{n}{2}\ell n[det(\boldsymbol{\Sigma})]\right\} \\ &= \frac{det(\boldsymbol{S})^{\frac{n-p-1}{2}}\left(\frac{n}{2}\right)^{\frac{np}{2}}}{\Gamma_p(\frac{1}{2}n)} exp\left\{-\frac{n}{2}\left(\frac{N}{n}\right) vec(\boldsymbol{\Sigma}^{-1})'vec(\frac{n}{N}\boldsymbol{S}) - \tfrac{n}{2}\ell n[det(\boldsymbol{\Sigma})]\right\} \\ &= b(\boldsymbol{t})exp\left[\boldsymbol{\theta}'\boldsymbol{t} - \kappa(\boldsymbol{\theta})\right] \end{aligned} \quad (5)$$

where $vec(\boldsymbol{S})$ denote the stacked columns of the $p \times p$ matrix $\boldsymbol{S}$ and the canonical statistic, corresponding canonical parameter and expected value are respectively given by:

$$\boldsymbol{t} = \left(\ vec(\frac{n}{N}\boldsymbol{S})\ \right),\ \boldsymbol{\theta} = \left(\ -\left(\frac{n}{2}\right)\left(\frac{N}{n}\right) vec(\boldsymbol{\Sigma}^{-1})\ \right),\ E(\boldsymbol{t}) = \boldsymbol{m} = \left(\ vec(\boldsymbol{\Sigma})\ \right). \quad (6)$$

The latter is chosen as canonical statistic since the mle of $\boldsymbol{m}$ without any constraints is $\widehat{\boldsymbol{m}} = \boldsymbol{t}$. The covariance matrix of the canonical statistic is given by

$$\boldsymbol{V} = Cov(\boldsymbol{t}) = Cov(vec(\boldsymbol{S})) = \left(I_{p^2} + \boldsymbol{K}\right)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})/N \quad (7)$$

where the matrix $\boldsymbol{K}$ is given by $\boldsymbol{K} = \sum_{i,j=1}^{p}(\boldsymbol{H}_{ij} \otimes \boldsymbol{H}'_{ij})$ and $\boldsymbol{H}_{ij} : p \times p$ with $h_{ij} = 1$ and all other elements are equal to zero (Muirhead (1982, page 90)).

A patterned covariance matrix or a covariance matrix with zeros implies specific constraints on the elements of $\boldsymbol{\Sigma}$. Consider the general constraint $\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{g}[vec(\boldsymbol{\Sigma})]$ with $\boldsymbol{G}_m = \frac{\partial}{\partial vec(\boldsymbol{\Sigma})}\boldsymbol{g}[vec(\boldsymbol{\Sigma})]$, $\boldsymbol{G}_t = \frac{\partial}{\partial vec(\boldsymbol{\Sigma})}\boldsymbol{g}[vec(\boldsymbol{\Sigma})]|_{\boldsymbol{\Sigma}=\boldsymbol{S}}$. The mle of $\boldsymbol{\Sigma}$ under the constraint $\boldsymbol{g}(vec(\boldsymbol{\Sigma}))$ is obtained iteratively from (2):

$$\widehat{\boldsymbol{m}} = vec(\boldsymbol{S}) - (\boldsymbol{G}_m\boldsymbol{V})'\left(\boldsymbol{G}_t\boldsymbol{V}\boldsymbol{G}'_m\right)^{*}\boldsymbol{g}[vec(\boldsymbol{S})]$$

Constraints are imposed only on elements in the upper half or the lower half of the covariance matrix. Symmetry of the covariance matrix is preserved automatically.

Using the procedure (2) for estimating $\boldsymbol{\Sigma}$ under restrictions, the vector $\boldsymbol{t}$ of canonical statistics with corresponding covariance matrix $\boldsymbol{V}$ is used as point of departure for any model considered within this framework. The algorithm given in Section 2 is then used to obtain the mle's. In each section the model specific expressions for $\boldsymbol{g}(\boldsymbol{m})$ and $\boldsymbol{G}_m$ are given.

3.1 An introductory example

The following numerical example serves to illustrate the essence and simplicity of the estimation procedure presented in this paper. Rencher (1998) considered the heights (in inches) and weights (in pounds) for a random sample of 20 college-age males. The sample covariance matrix is equal to $\boldsymbol{S} = \begin{pmatrix} 14.576316 & 128.87895 \\ 128.87895 & 1441.2737 \end{pmatrix}$. We assume that this sample was drawn from a multivariate normal distribution with unknown mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$. In order to obtain the mle of the covariance matrix using the expression in (2), the canonical statistic $\boldsymbol{t}$ (cf. (6)) with corresponding covariance matrix $\boldsymbol{V}$ (cf. (7)) are required:

$$\boldsymbol{t} = \begin{pmatrix} (n-1)s_{11}/n \\ (n-1)s_{12}/n \\ (n-1)s_{21}/n \\ (n-1)s_{22}/n \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix}, \; E(\boldsymbol{t}) = \boldsymbol{m} = \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{21} \\ \sigma_{22} \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix}, \text{ say and}$$

$$Cov(\boldsymbol{t}) = \boldsymbol{V} = \left(I_{p^2} + \boldsymbol{K}\right)[\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}]/N = \frac{1}{N}\begin{pmatrix} 2\sigma_{11}^2 & 2\sigma_{11}\sigma_{12} & 2\sigma_{11}\sigma_{12} & 2\sigma_{12}^2 \\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & 2\sigma_{12}\sigma_{22} \\ 2\sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & \sigma_{11}\sigma_{22} + \sigma_{12}^2 & 2\sigma_{12}\sigma_{22} \\ 2\sigma_{12}^2 & 2\sigma_{12}\sigma_{22} & 2\sigma_{12}\sigma_{22} & 2\sigma_{22}^2 \end{pmatrix}.$$

The mle of the covariance matrix under a particular hypothesis is obtained by specifying constraints of the form $\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{g}[vec(\boldsymbol{\Sigma})] = \boldsymbol{0}$. Various constraints on the parameters may be considered.

Firstly, consider the hypothesis that the variances of height and weight are proportional with a known constant, e.g. $\sigma_{11} = 0.01\sigma_{22}$. In this case, $g(\boldsymbol{m}) = \sigma_{11} - 0.01\sigma_{22} = (1, 0, 0, -0.01)\boldsymbol{m} = \boldsymbol{C}\boldsymbol{m}$, say, with derivative $\boldsymbol{G}_m = \boldsymbol{C} = \boldsymbol{G}_t$. Consequently,

$$\boldsymbol{CV} = \left(2\sigma_{11}^2 - 0.02\sigma_{12}^2\;,\; 2\sigma_{11}\sigma_{12} - 0.02\sigma_{12}\sigma_{22}\;,\; 2\sigma_{11}\sigma_{12} - 0.02\sigma_{12}\sigma_{22}\;,\; 2\sigma_{12}^2 - 0.02\sigma_{22}^2\right)$$

and

$$\boldsymbol{CVC}' = 2\sigma_{11}^2 - 0.04\sigma_{12}^2 + 0.0002\sigma_{22}^2.$$

Table 1: Iteration process illustrated for a linear constraint.

		Iteration over $\boldsymbol{t}$	
	Iteration over $\boldsymbol{m}$	Starting Value	1
1	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 13.765613 \\ 122.39777 \\ 122.39777 \\ 1376.5613 \end{pmatrix}$
2	$\begin{pmatrix} 13.765613 \\ 122.39777 \\ 122.39777 \\ 1376.5613 \end{pmatrix}$	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 13.7698 \\ 122.435 \\ 122.435 \\ 1376.98 \end{pmatrix}$

In the iterative procedure the covariance matrix subject to the constraint $\sigma_{11} = 0.01\sigma_{22}$ is calculated using:

$$\widehat{\boldsymbol{m}} = \boldsymbol{t} - (\boldsymbol{CV})' \left(\boldsymbol{CVC}'\right)^{*} g(\boldsymbol{t}) = \boldsymbol{t} - \frac{t_1 - 0.01t_4}{2m_1^2 - 0.04m_2^2 + 0.0002m_4^2} \begin{pmatrix} 2m_1^2 - 0.02m_2^2 \\ 2m_1m_2 - 0.02m_2m_4 \\ 2m_1m_2 - 0.02m_2m_4 \\ 2m_2^2 - 0.02m_4^2 \end{pmatrix}.$$

The starting value for both $\boldsymbol{m}$ and $\boldsymbol{t}$ is $\boldsymbol{t} = \frac{n-1}{n}\boldsymbol{S} = \frac{19}{20} \begin{pmatrix} 14.576316 \\ 128.87895 \\ 128.87895 \\ 1441.2737 \end{pmatrix}$. The iteration process is summarized in Table 1. No iteration over $\boldsymbol{t}$ is required since $\boldsymbol{g}(\boldsymbol{m})$ is a linear function.

In order to illustrate the double iteration in the estimation procedure, consider now the hypothesis that the square of the variance of height is proportional to the variance of weight, i.e. $7\sigma_{11}^2 = \sigma_{22}$. Using traditional maximum likelihood methodology, this relation will have to be dealt with by reparameterization through substitution of σ_{22} by $7\sigma_{11}^2$ and derivation of a new set of likelihood equations. In this case, the implied constraint is used: $g(\boldsymbol{m}) = g[vec(\boldsymbol{\Sigma})] = 7\sigma_{11}^2 - \sigma_{22} = 7m_1^2 - m_4$ with corresponding matrices of derivatives $\boldsymbol{G}_m = 14m_1\frac{\partial m_1}{\partial \boldsymbol{m}} - \frac{\partial m_4}{\partial \boldsymbol{m}} = 14m_1(1,0,0,0) - (0,0,0,1)$ and $\boldsymbol{G}_t = \left.\frac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}}\right|_{\boldsymbol{m}=\boldsymbol{t}}$. The iteration process proceeds as indicated

Table 2: Double iteration process illustrated for a nonlinear constraint.

		Iteration over $\boldsymbol{t}$		
	Iteration over $\boldsymbol{m}$	Starting Value	1	2
1	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 14.0303 \\ 123.7619 \\ 123.7619 \\ 1377.7078 \end{pmatrix}$	$\begin{pmatrix} 14.0287 \\ 123.7506 \\ 123.7506 \\ 1377.6353 \end{pmatrix}$
2	$\begin{pmatrix} 14.0287 \\ 123.7506 \\ 123.7506 \\ 1377.6353 \end{pmatrix}$	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 14.0316 \\ 123.7799 \\ 123.7799 \\ 1377.9653 \end{pmatrix}$	$\begin{pmatrix} 14.0300 \\ 123.7682 \\ 123.7682 \\ 1377.8895 \end{pmatrix}$
3	$\begin{pmatrix} 14.0300 \\ 123.7682 \\ 123.7682 \\ 1377.8895 \end{pmatrix}$	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 14.0316 \\ 123.78 \\ 123.78 \\ 1377.9667 \end{pmatrix}$	$\begin{pmatrix} 14.0300 \\ 123.7684 \\ 123.7684 \\ 1377.8909 \end{pmatrix}$
4	$\begin{pmatrix} 14.0300 \\ 123.7684 \\ 123.7684 \\ 1377.8909 \end{pmatrix}$	$\begin{pmatrix} 13.8475 \\ 122.435 \\ 122.435 \\ 1369.21 \end{pmatrix}$	$\begin{pmatrix} 14.0316 \\ 123.7800 \\ 123.7800 \\ 1377.9667 \end{pmatrix}$	$\begin{pmatrix} 14.0300 \\ 123.7684 \\ 123.7684 \\ 1377.8909 \end{pmatrix}$

in Table 2. In each new iteration over $\boldsymbol{m}$, the observed $\boldsymbol{t}$ ($\boldsymbol{t}_0$), is used as initial value for $\boldsymbol{t}$.

4 Covariance matrices with given structure

4.1 Linearly patterned covariance matrices

Several methods for maximum likelihood estimation of patterned covariance matrices, which are used to model multivariate normal data, are given in the literature. Explicit forms for maximum likelihood estimates (mle's) exist in some cases. Szatrowski(1980) gives necessary and sufficient conditions on linear patterns such that there exists explicit solutions which are obtained in one iteration of the scoring equations from any positive definite starting point. However, patterns in applications often arise where explicit estimates do not exist and an iterative algorithm is required. Such algorithms include the well-known Newton-Raphson algorithm and the method of scoring (Anderson, 1970, 1973) which do not have guaranteed convergence for all patterns. The

EM algorithm (Rubin & Szatrowski, 1982) was proposed as an alternative method of estimation in cases where the patterned covariance matrices which do not have explicit mle's can be viewed as submatrices of larger patterned covariance matrices that do have explicit mle's. In this subsection, the exact maximum likelihood estimates of patterned covariance matrices are obtained directly by specifying specific structures in terms of constraints. Conditions for convergence are similar to those for Newton-Raphson.

Firstly, consider linearly patterned covariance matrices of the general form $\boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \sum_{g=0}^{m} \sigma_g \boldsymbol{C}_g$ (Szatrowski, 1980, Anderson, 1970,1973) where $\boldsymbol{\sigma}' = (\sigma_0, \sigma_1, \cdots, \sigma_m)$ are unknown coefficients and $\boldsymbol{C}_0, \boldsymbol{C}_1, \cdots, \boldsymbol{C}_m$ are known symmetric linearly independent matrices. The linear structure of the covariance matrix implies that

$$\boldsymbol{m} = vec(\boldsymbol{\Sigma}) = \sum_{g=0}^{m} \sigma_g vec(\boldsymbol{C}_g) = (vec\boldsymbol{C}_0, vec\boldsymbol{C}_1, \cdots, vec\boldsymbol{C}_m)\boldsymbol{\sigma} = \boldsymbol{C}\boldsymbol{\sigma}. \tag{8}$$

Consequently,

$$\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{Q}_C \boldsymbol{C}\boldsymbol{\sigma} = \boldsymbol{0} \text{ where } \boldsymbol{Q}_C = \boldsymbol{I} - \boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^* \boldsymbol{C}' \tag{9}$$

projects orthogonal to the columns of $\boldsymbol{C}$. The matrix of derivatives $\boldsymbol{G}_m = \dfrac{\partial \boldsymbol{g}(\boldsymbol{m})}{\partial \boldsymbol{m}} = \boldsymbol{Q}_C = \boldsymbol{G}_t$. Structures such as the complete symmetry pattern (with special cases the intraclass correlation pattern and sphericity test), block complete symmetry pattern, compound-, circular- and stationary symmetry pattern can be modelled within this framework.

Consider the 3×3 stationary covariance pattern (Rubin & Szatrowski, 1982):

$$\boldsymbol{\Sigma} = \begin{pmatrix} a & b & c \\ b & a & b \\ c & b & a \end{pmatrix}.$$

This may be expressed in linear form (8) with corresponding implied constraint (9) and matrix of derivatives $\boldsymbol{G}_m = \boldsymbol{C} = \boldsymbol{G}_t$. Alternatively, the following three simple constraints are required to obtain the mle's:

$$\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{33} \\ \sigma_{12} - \sigma_{23} \end{pmatrix} = \begin{pmatrix} vec(\boldsymbol{\Sigma})[1] - vec(\boldsymbol{\Sigma})[5] \\ vec(\boldsymbol{\Sigma})[1] - vec(\boldsymbol{\Sigma})[9] \\ vec(\boldsymbol{\Sigma})[2] - vec(\boldsymbol{\Sigma})[6] \end{pmatrix}$$

where $vec(\boldsymbol{\Sigma})[i]$ denotes the i-th element of $vec(\boldsymbol{\Sigma})$. The corresponding matrix of derivatives are given by:

$$\boldsymbol{G}_m = \boldsymbol{G}_t = \begin{pmatrix} \boldsymbol{I}_{p^2}[1,] - \boldsymbol{I}_{p^2}[5,] \\ \boldsymbol{I}_{p^2}[1,] - \boldsymbol{I}_{p^2}[9,] \\ \boldsymbol{I}_{p^2}[2,] - \boldsymbol{I}_{p^2}[6,] \end{pmatrix}$$

where $\boldsymbol{I}_{p^2}[i,]$ denotes the i-th row of the identity matrix $\boldsymbol{I} : p^2 \times p^2$.

Using the procedure (2) for maximum likelihood estimation under constraints, it is irrelevant whether an explicit mle exist for a specific symmetry structure since the exact mle's are obtained. The process of estimation follows the same straightforward method for any one of the patterns mentioned.

4.2 Covariance matrices with zeros

To find the mle of a covariance matrix under the constraint that certain covariances are zero, Chaudhuri, Drton & Richardson (2007) proposed an iterative conditional fitting algorithm (with guaranteed convergence properties) under the assumption of multivariate normality. Where the presence of zero covariances can be formulated as a linear hypotheses on the covariance matrix, e.g. covariance graph models (Chaudhuri, Drton & Richardson, 2007), procedures such as Newton-Raphson and the method of scoring (Anderson, 1970,1973) may be used for maximum likelihood estimation. However, the algorithm introduced by Chaudhuri et al. (2007) has clearer convergence properties than Anderson's algorithm. In this subsection, the exact maximum likelihood estimate of a covariance matrix with specific covariances equal to zero, is obtained by directly specifying these zeros as constraints on the elements of the covariance matrix. No likelihood equations are required. Convergence properties are similar to those of the Newton-Raphson method.

Estimation of the covariance matrix under the constraint that certain covariances are zero, represents another direct application of the procedure given in (2). Chaudhuri et al. (2007) consider a data example where they focus on $n = 134$ measurements of $p = 8$ genes related to galactose use in gene expression data from microarray experiments with yeast strands. The marginal correlations and standard deviations of these variables are given to two decimal places. The sample correlation matrix calculated from this is given below:

$$S = \begin{pmatrix} 0.1521 & 0.033696 & 0.014664 & -0.11934 & -0.0663 & -0.054756 & -0.050505 & -0.048048 \\ & 0.1296 & 0.038916 & -0.01836 & -0.0612 & 0.033696 & -0.05328 & -0.038808 \\ & & 0.2209 & 0.20774 & 0.22372 & 0.07332 & 0.182595 & 0.188188 \\ & & & 2.89 & 2.5143 & 0.58344 & 2.54745 & 2.27766 \\ & & & & 2.89 & 0.51714 & 2.7676 & 2.40856 \\ & & & & & 0.6084 & 0.7215 & 0.552552 \\ & & & & & & 3.4225 & 2.59259 \\ & & & & & & & 2.3716 \end{pmatrix}$$

Chaudhuri et al. (2007) give mle's of covariances for two different structures, here indicated by $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_d$. The variables are presented in the same order used by Chaudhuri et al. (2007).

$$\mathbf{\Sigma}_s = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & \sigma_{14} & 0 & \sigma_{16} & 0 & 0 \\ & \sigma_{22} & \sigma_{23} & 0 & 0 & 0 & 0 & 0 \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} & \sigma_{37} & \sigma_{38} \\ & & & \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} & \sigma_{48} \\ & & & & \sigma_{55} & \sigma_{56} & \sigma_{57} & \sigma_{58} \\ & & & & & \sigma_{66} & \sigma_{67} & \sigma_{68} \\ & & & & & & \sigma_{77} & \sigma_{78} \\ & & & & & & & \sigma_{88} \end{pmatrix} \mathbf{\Sigma}_d = \begin{pmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ & \sigma_{22} & \sigma_{23} & 0 & 0 & 0 & 0 & 0 \\ & & \sigma_{33} & \sigma_{34} & \sigma_{35} & 0 & 0 & \sigma_{38} \\ & & & \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} & \sigma_{48} \\ & & & & \sigma_{55} & \sigma_{56} & \sigma_{57} & \sigma_{58} \\ & & & & & \sigma_{66} & \sigma_{67} & \sigma_{68} \\ & & & & & & \sigma_{77} & \sigma_{78} \\ & & & & & & & \sigma_{88} \end{pmatrix}$$

Using maximum likelihood estimation under constraints, the covariance matrices are estimated by setting specific covariances equal to zero by simply specifying the corresponding required constraints. Let $\boldsymbol{g}_s(\boldsymbol{m})$ and $\boldsymbol{g}_d(\boldsymbol{m})$ respectively denote the constraints to set the appropriate covariances equal to zero in the structures $\mathbf{\Sigma}_s$ and $\mathbf{\Sigma}_d$ given above. Then $\boldsymbol{g}_s(\boldsymbol{m}) = \boldsymbol{g}[vec(\mathbf{\Sigma}_s)] = \mathbf{0}$, $\boldsymbol{g}_d(\boldsymbol{m}) = \boldsymbol{g}[vec(\mathbf{\Sigma}_d)] = \mathbf{0}$ where:

$$\boldsymbol{g}_s(\boldsymbol{m}) = \begin{pmatrix} vec(\mathbf{\Sigma}_s)[3] \\ vec(\mathbf{\Sigma}_s)[5] \\ vec(\mathbf{\Sigma}_s)[7] \\ vec(\mathbf{\Sigma}_s)[8] \\ vec(\mathbf{\Sigma}_s)[12] \\ vec(\mathbf{\Sigma}_s)[13] \\ vec(\mathbf{\Sigma}_s)[14] \\ vec(\mathbf{\Sigma}_s)[15] \\ vec(\mathbf{\Sigma}_s)[16] \end{pmatrix}, \boldsymbol{G}_{sm} = \boldsymbol{G}_{st} = \begin{pmatrix} \boldsymbol{I}_{p^2}[3,] \\ \boldsymbol{I}_{p^2}[5,] \\ \boldsymbol{I}_{p^2}[7,] \\ \boldsymbol{I}_{p^2}[8,] \\ \boldsymbol{I}_{p^2}[12,] \\ \boldsymbol{I}_{p^2}[13,] \\ \boldsymbol{I}_{p^2}[14,] \\ \boldsymbol{I}_{p^2}[15,] \\ \boldsymbol{I}_{p^2}[16,] \end{pmatrix},$$

$$\boldsymbol{g}_d(\boldsymbol{m}) = \begin{pmatrix} vec(\boldsymbol{\Sigma}_d)[3] \\ vec(\boldsymbol{\Sigma}_d)[4] \\ vec(\boldsymbol{\Sigma}_d)[5] \\ vec(\boldsymbol{\Sigma}_d)[6] \\ vec(\boldsymbol{\Sigma}_d)[7] \\ vec(\boldsymbol{\Sigma}_d)[8] \\ vec(\boldsymbol{\Sigma}_d)[12] \\ vec(\boldsymbol{\Sigma}_d)[13] \\ vec(\boldsymbol{\Sigma}_d)[14] \\ vec(\boldsymbol{\Sigma}_d)[15] \\ vec(\boldsymbol{\Sigma}_d)[16] \\ vec(\boldsymbol{\Sigma}_d)[22] \\ vec(\boldsymbol{\Sigma}_d)[23] \end{pmatrix}, \boldsymbol{G}_{dm} = \boldsymbol{G}_{dt} = \begin{pmatrix} \boldsymbol{I}_{p^2}[3,] \\ \boldsymbol{I}_{p^2}[4,] \\ \boldsymbol{I}_{p^2}[5,] \\ \boldsymbol{I}_{p^2}[6,] \\ \boldsymbol{I}_{p^2}[7,] \\ \boldsymbol{I}_{p^2}[8,] \\ \boldsymbol{I}_{p^2}[12,] \\ \boldsymbol{I}_{p^2}[13,] \\ \boldsymbol{I}_{p^2}[14,] \\ \boldsymbol{I}_{p^2}[15,] \\ \boldsymbol{I}_{p^2}[16,] \\ \boldsymbol{I}_{p^2}[22,] \\ \boldsymbol{I}_{p^2}[23,] \end{pmatrix}.$$

The matrices $\boldsymbol{G}_{sm}$, $\boldsymbol{G}_{st}$, $\boldsymbol{G}_{dm}$, $\boldsymbol{G}_{dt}$ denote the corresponding matrices of derivatives.

The estimated covariances, which agree with those given by Chaudhuri et al. (2007), are given below:

$$\widehat{\boldsymbol{\Sigma}}_s = \begin{pmatrix} 0.155 & 0.038 & 0 & -0.075 & 0 & -0.063 & 0 & 0 \\ & 0.126 & 0.034 & 0 & 0 & 0 & 0 & 0 \\ & & 0.218 & 0.214 & 0.233 & 0.070 & 0.192 & 0.193 \\ & & & 2.808 & 2.452 & 0.558 & 2.492 & 2.225 \\ & & & & 2.847 & 0.489 & 2.726 & 2.373 \\ & & & & & 0.598 & 0.696 & 0.528 \\ & & & & & & 3.371 & 2.554 \\ & & & & & & & 2.336 \end{pmatrix}$$

$$\widehat{\boldsymbol{\Sigma}}_d = \begin{pmatrix} 0.150 & 0.030 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0.126 & 0.036 & 0 & 0 & 0 & 0 & 0 \\ & & 0.218 & 0.061 & 0.086 & 0 & 0 & 0.051 \\ & & & 2.763 & 2.374 & 0.557 & 2.453 & 2.167 \\ & & & & 2.723 & 0.485 & 2.646 & 2.277 \\ & & & & & 0.599 & 0.711 & 0.530 \\ & & & & & & 3.371 & 2.507 \\ & & & & & & & 2.266 \end{pmatrix}$$

The corresponding standard error of the estimates, $\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{m}})$, obtained using (3), are given below:

$$\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\Sigma}}_s) = \begin{pmatrix} 0.0178 & 0.0114 & 0 & 0.0259 & 0 & 0.0218 & 0 & 0 \\ & 0.0148 & 0.0135 & 0 & 0 & 0 & 0 & 0 \\ & & 0.0263 & 0.0677 & 0.0690 & 0.0304 & 0.0738 & 0.0621 \\ & & & 0.3388 & 0.3189 & 0.1187 & 0.3382 & 0.2899 \\ & & & & 0.3422 & 0.1171 & 0.3519 & 0.2989 \\ & & & & & 0.0712 & 0.1335 & 0.1092 \\ & & & & & & 0.4081 & 0.3247 \\ & & & & & & & 0.2828 \end{pmatrix}$$

$$\widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\Sigma}}_d) = \begin{pmatrix} 0.0165 & 0.0112 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0.0148 & 0.0132 & 0 & 0 & 0 & 0 & 0 \\ & & 0.0247 & 0.0383 & 0.0318 & 0 & 0 & 0.0252 \\ & & & 0.3192 & 0.2961 & 0.1098 & 0.3156 & 0.2705 \\ & & & & 0.3162 & 0.1083 & 0.3251 & 0.2757 \\ & & & & & 0.0665 & 0.1264 & 0.1020 \\ & & & & & & 0.3859 & 0.3028 \\ & & & & & & & 0.2641 \end{pmatrix}$$

The Wald statistic (4) for $\widehat{\boldsymbol{\Sigma}}_s$, W_s, is equal to 9.2767 ($p = 0.4121$) with $\nu = 9$ and for $\widehat{\boldsymbol{\Sigma}}_d$ is given by $W_d = 29.96$, ($p = 0.00477$) with $\nu = 13$.

5 Dependent multivariate normal samples

The problem of testing the homogeneity of the covariance matrices of dependent multivariate normal populations, is another example where no explicit analytical expression of the mle of the covariance matrix (under the null hypothesis) exist. Jiang et al. (1999) proposed a likelihood ratio test and modifications thereof, through an iterative scheme using PROC MIXED in SAS for finding the mle of the covariance matrix - not a straightforward process. An overview of preceding attempts and approaches to variations of this problem is given in the introduction of the paper by Jiang et al. (1999). In this section, it is illustrated how this problem is solved by obtaining the exact maximum likelihood estimates directly by specifying appropriate constraints on the covariance matrices under consideration. Methodology is simplified considerably.

Let $\boldsymbol{Y} : pk \times 1$ denote a random vector composed of k groups of p variables each and

$\boldsymbol{Y} \sim N_{pk}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unknown mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ respectively given by:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 : p \times 1 \\ \vdots \\ \boldsymbol{\mu}_k : p \times 1 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1k} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Sigma}_{k1} & \cdots & \boldsymbol{\Sigma}_{kk} \end{pmatrix} \text{ where } \boldsymbol{\Sigma}_{ij} : p \times p \text{ for } i, j = 1, \cdots k.$$

The hypothesis of interest is $H_0 : \boldsymbol{\Sigma}_{11} = \cdots = \boldsymbol{\Sigma}_{kk}$ against the alternative hypothesis that at least one of the equalities does not hold. No assumptions are made regarding the off-diagonal covariance matrices. Let $\boldsymbol{y}_{ij}$, $i = 1, \cdots, q$, $j = 1, \cdots, N_i$ denote N_i observations of the pk-variate random vector $\boldsymbol{Y}$ observed for population i. The sample covariance matrix is given by $\boldsymbol{S} = \frac{1}{n} \sum_{i=1}^{q} \sum_{j=1}^{N_i} (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}}_i)'$, $\overline{\boldsymbol{y}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \boldsymbol{y}_{ij}$, $n = N - q$, $N = \sum_{i=1}^{q} N_i$. The density function of $\boldsymbol{S}$ is given by (5) and expressed in terms of its canonical statistic and corresponding canonical parameter (6).

The example considered by Jiang & Sarkar (1998) and Jiang et al. (1999) on the bio-equivalence of two formulations of a drug product are used in the next two subsections to illustrate how a test for independence of groups of variables as well as a test for the homogeneity of covariance matrices (with or without given structure) can be performed. A standard 2×2 crossover experiment was conducted with 25 subjects to compare a new test formulation with a reference formulation. The two treatment periods were separated by a 7-day washout period. Two variables, namely the area under the plasma concentration-time curve (AUC) and the maximum plasma concentration (Cmax), were considered. The natural logarithm of AUC and Cmax are believed to be marginally normally distributed, i.e. the 2×2 variate random vector

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim N_4 \left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right]. \tag{10}$$

The two components of $\boldsymbol{y}$, given by $\boldsymbol{y}_{ijk}$, $i = 1, 2$, $j = 1, \cdots, N_i$, $k = 1$(test), 2(reference), represent the log(AUC) and log(Cmax) of the k-th formulation of the j-th subject in the i-th sequence. Assuming different effects of the period administering formulation k in sequence i, the sample covariance matrix with $N_1 = 12$, $N_2 = 13$ is given by (Jiang & Sarkar (1998)):

$$\boldsymbol{S} = \begin{pmatrix} 0.0657739 & 0.01602 & 0.0600567 & -0.002653, \\ 0.01602 & 0.0534908 & 0.0084126 & 0.0125034 \\ 0.0600567 & 0.0084126 & 0.0729088 & -0.000625 \\ -0.002653 & 0.0125034 & -0.000625 & 0.0459961 \end{pmatrix} \tag{11}$$

.

Three structures for $\boldsymbol{\Sigma}$, denoted by $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$ respectively, are considered in the next two subsections.

5.1 Test for Independence

To test for independence of k sets of variates the off-diagonal covariance matrices are all set equal to zero:

$$\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{g}(vec(\boldsymbol{\Sigma})) = \begin{pmatrix} vec(\boldsymbol{\Sigma}_{12}) \\ \cdots \\ vec(\boldsymbol{\Sigma}_{k-1,k}) \end{pmatrix} \text{ with } \boldsymbol{G}_m = \begin{pmatrix} \frac{\partial vec(\boldsymbol{\Sigma}_{12})}{\partial vec(\boldsymbol{\Sigma})} \\ \cdots \\ \frac{\partial vec(\boldsymbol{\Sigma}_{k-1,k})}{\partial vec(\boldsymbol{\Sigma})} \end{pmatrix}.$$

For the example by Jiang et al. (1999), consider $H_0 : \boldsymbol{\Sigma}_{12} = 0$ where $\boldsymbol{\Sigma} : 4 \times 4$, given by (10). The constraints implied by the test for independence of the two sets of variates are:

$$\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} \sigma_{13} \\ \sigma_{14} \\ \sigma_{23} \\ \sigma_{24} \end{pmatrix} \begin{pmatrix} vec(\boldsymbol{\Sigma})[3] \\ vec(\boldsymbol{\Sigma})[4] \\ vec(\boldsymbol{\Sigma})[7] \\ vec(\boldsymbol{\Sigma})[8] \end{pmatrix} = \boldsymbol{0} \text{ with } \boldsymbol{G}_m = \begin{pmatrix} 0&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0 \end{pmatrix}$$

The maximum likelihood estimate of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1$ with corresponding standard errors under these constraints are:

$$\widehat{\boldsymbol{\Sigma}}_1 = \begin{pmatrix} 0.060512 & 0.014738 & 0 & 0 \\ 0.014738 & 0.049212 & 0 & 0 \\ 0 & 0 & 0.067076 & -0.000575 \\ 0 & 0 & -0.000575 & 0.042316 \end{pmatrix} \widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\Sigma}}_1) = \begin{pmatrix} 0.0063 & 0.0056 & 0 & 0 \\ 0.0056 & 0.0128 & 0 & 0 \\ 0 & 0 & 0.0069 & 0.0052 \\ 0 & 0 & 0.0052 & 0.0111 \end{pmatrix}$$

The Wald statistic, W_1, is equal to 12.523 ($p = 0.0139$) with $\nu = 4$.

5.2 Homogeneity of Covariance Matrices

In order to test for the homogeneity of covariance matrices, the $(k-1)\frac{1}{2}p(p+1)$ constraints implied by the model are simply:

$$\boldsymbol{g}(\boldsymbol{m}) = \boldsymbol{g}(vec(\boldsymbol{\Sigma})) = \begin{pmatrix} vec(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{22}) \\ \cdots \\ vec(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{kk}) \end{pmatrix} \text{ with } \boldsymbol{G}_m = \begin{pmatrix} \dfrac{\partial vec(\boldsymbol{\Sigma}_{11}) - vec(\boldsymbol{\Sigma}_{22})}{\partial vec(\boldsymbol{\Sigma})} \\ \cdots \\ \dfrac{\partial vec(\boldsymbol{\Sigma}_{11}) - vec(\boldsymbol{\Sigma}_{kk})}{\partial vec(\boldsymbol{\Sigma})} \end{pmatrix}. \quad (12)$$

To fit the model of homogeneity of covariance matrices of dependent multivariate normals to the data provided in the example by Jiang et al. (1999), the null hypothesis, $H_0 : \boldsymbol{\Sigma}_{11} = \boldsymbol{\Sigma}_{22}, \boldsymbol{\Sigma}_{12} \neq 0$, is considered. The appropriate constraints in terms of the typical elements σ_{ij} of $\boldsymbol{\Sigma}$ (cf. (10)) are given by:

$$\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} \sigma_{11} - \sigma_{33} \\ \sigma_{12} - \sigma_{34} \\ \sigma_{22} - \sigma_{44} \end{pmatrix} = \begin{pmatrix} vec(\boldsymbol{\Sigma})[1] - vec(\boldsymbol{\Sigma})[11] \\ vec(\boldsymbol{\Sigma})[2] - vec(\boldsymbol{\Sigma})[12] \\ vec(\boldsymbol{\Sigma})[6] - vec(\boldsymbol{\Sigma})[16] \end{pmatrix} \text{ with}$$

$$\boldsymbol{G}_m = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

These constraints yield the mle's of $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$ below. These values correspond with the results of Jiang et al. (1999) to five decimal places.

$$\widehat{\boldsymbol{\Sigma}}_2 = \begin{pmatrix} 0.064155 & 0.007345 & 0.055672 & 0.003795 \\ 0.007345 & 0.045719 & 0.002148 & 0.011277 \\ 0.0556724 & 0.002148 & 0.064155 & 0.007345 \\ 0.003800 & 0.011277 & 0.007345 & 0.045719 \end{pmatrix} \quad \widehat{\boldsymbol{\sigma}}(\widehat{\boldsymbol{\Sigma}}_2) = \begin{pmatrix} 0.0159 & 0.0082 & 0.0159 & 0.0086 \\ 0.0082 & 0.0092 & 0.0099 & 0.0094 \\ 0.0159 & 0.0099 & 0.0159 & 0.0082 \\ 0.0086 & 0.0094 & 0.0082 & 0.0092 \end{pmatrix}$$

This model provides improved fit: $W_2 = 1.915$, $(p = 0.5901)$ with $\nu = 3$.

Patterns can be imposed additionally on submatrices $\boldsymbol{\Sigma}_{ij} : p \times p$ for $i \leq j = 1, \cdots k$ by specifying the corresponding constraints. To illustrate, consider the sample covariance matrix in (11). It suggests that the dependence relation between the two sets of variates is contained within the sequence. Thus the relation between the two formulations are independent of the sequence in which they were administered. Consequently, a model for homogeneity of covariance matrices of such structurally dependent multivariate normals was fitted, i.e. $H_0 : \mathbf{\Sigma}_{11} = \mathbf{\Sigma}_{22},\ \mathbf{\Sigma}^*_{12} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ where a the covariance between AUC for the test and reference formulations and b the covariance between Cmax for the test and reference formulations. The constraints required are:

$$\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} vec(\mathbf{\Sigma})[1] - vec(\mathbf{\Sigma})[11] \\ vec(\mathbf{\Sigma})[2] - vec(\mathbf{\Sigma})[12] \\ vec(\mathbf{\Sigma})[6] - vec(\mathbf{\Sigma})[16] \\ vec(\mathbf{\Sigma})[4] \\ vec(\mathbf{\Sigma})[7] \end{pmatrix}$$

with matrix of derivatives

$$\boldsymbol{G}_m = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The maximum likelihood estimate of $\mathbf{\Sigma} = \mathbf{\Sigma}_3$ is:

$$\widehat{\mathbf{\Sigma}}_3 = \begin{pmatrix} 0.0634099 & 0.0046713 & 0.0548077 & 0 \\ 0.0046713 & 0.0456856 & 0 & 0.0111829 \\ 0.0548077 & 0 & 0.0634099 & 0.0046713 \\ 0 & 0.0111829 & 0.0046713 & 0.0456856 \end{pmatrix} \quad \widehat{\boldsymbol{\sigma}}(\widehat{\mathbf{\Sigma}}_3) = \begin{pmatrix} 0.0156 & 0.0033 & 0.0156 & 0.0000 \\ 0.0033 & 0.0092 & 0.0000 & 0.0091 \\ 0.0156 & 0.0000 & 0.0156 & 0.0033 \\ 0.0000 & 0.0091 & 0.0033 & 0.0092 \end{pmatrix}$$

The Wald statistic, W_3, is equal to 2.045 ($p = 0.843$) with $\nu = 5$. Of the three models considered, this one fits best.

5.3 An equivalent approach

The likelihood function of independent multivariate normal populations can of course also be used in a straightforward way as point of departure for this problem of testing for homogeneity of covariance matrices of groups of dependent multivariate normal variables. Theory is given and illustrated for two independent populations - again using the example by Jiang et al. (1999). Exactly the same results as those given in the previous subsection (cf. $\widehat{\boldsymbol{\Sigma}}_2$), are obtained. Generalization of the process to q populations follows directly.

Suppose that $\boldsymbol{y}_{i1}, \boldsymbol{y}_{i2}, \cdots, \boldsymbol{y}_{in_i}$ represent n_i independent observations of two independent random samples from $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ distributions $(i = 1, 2)$. Let $\overline{\boldsymbol{y}}_i$ represent the sample mean vector and $\boldsymbol{S}_i$ the matrix of mean sums of squares and products of the i-th sample:

$$\overline{\boldsymbol{y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}, \quad \boldsymbol{S}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}\boldsymbol{y}'_{ij}.$$

The likelihood function for the two multivariate samples can be expressed in its canonical exponential form as follows:

$$\begin{aligned}
&L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \\
&= \Pi_{i=1}^{2} det(2\pi\boldsymbol{\Sigma}_i)^{-n_i/2} exp\left\{-\frac{1}{2}tr\boldsymbol{\Sigma}_i^{-1}\left[\sum_{j=1}^{n_i}(\boldsymbol{y}_{ij}-\boldsymbol{\mu}_i)(\boldsymbol{y}_{ij}-\boldsymbol{\mu}_i)'\right]\right\} \\
&= \Pi_{i=1}^{2} exp\left\{n_i\boldsymbol{\mu}'_i\boldsymbol{\Sigma}_i^{-1}(\frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}) - \frac{n_i}{2}tr\boldsymbol{\Sigma}_i^{-1}(\frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{y}_{ij}\boldsymbol{y}'_{ij}) - \frac{n_i}{2}\boldsymbol{\mu}'_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{n_i}{2}\ell n[det(2\pi\boldsymbol{\Sigma}_i)]\right\} \\
&= exp\left[\boldsymbol{\theta}'\boldsymbol{t} - \kappa(\boldsymbol{\theta})\right] \text{ where}
\end{aligned}$$

$$\boldsymbol{t} = \begin{pmatrix} \frac{1}{n_1}\sum_{j=1}^{n_1}\boldsymbol{y}_{1j} \\ vec(\frac{1}{n_1}\sum_{j=1}^{n_1}\boldsymbol{y}_{1j}\boldsymbol{y}'_{1j}) \\ \frac{1}{n_2}\sum_{j=1}^{n_2}\boldsymbol{y}_{2j} \\ vec(\frac{1}{n_2}\sum_{j=1}^{n_2}\boldsymbol{y}_{2j}\boldsymbol{y}'_{kj}) \end{pmatrix} = \begin{pmatrix} \overline{\boldsymbol{y}}_1 \\ vec(\boldsymbol{S}_1) \\ \overline{\boldsymbol{y}}_2 \\ vec(\boldsymbol{S}_2) \end{pmatrix}, \boldsymbol{\theta} = \begin{pmatrix} n_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 \\ -\frac{n_1}{2}vec(\boldsymbol{\Sigma}_1^{-1}) \\ n_2\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 \\ -\frac{n_2}{2}vec(\boldsymbol{\Sigma}_2^{-1}) \end{pmatrix} \tag{13}$$

$$\text{and } E(\boldsymbol{t}) = \boldsymbol{m} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ vec(\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}'_1) \\ \boldsymbol{\mu}_2 \\ vec(\boldsymbol{\Sigma}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}'_2) \end{pmatrix} = \begin{pmatrix} \boldsymbol{m}_{11} \\ \boldsymbol{m}_{12} \\ \boldsymbol{m}_{21} \\ \boldsymbol{m}_{22} \end{pmatrix}. \tag{14}$$

The covariance matrix of $\boldsymbol{t}$ is given by

$$\boldsymbol{V} = Cov(\boldsymbol{t}) = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{22} \end{pmatrix} \text{ with } \boldsymbol{V}_{ii} = \begin{pmatrix} \boldsymbol{V}_{11}^i & \boldsymbol{V}_{12}^i \\ \boldsymbol{V}_{21}^i & \boldsymbol{V}_{22}^i \end{pmatrix}$$

where for $i = 1, 2$:

$$\boldsymbol{V}_{11}^{i} = \frac{1}{n_i}\boldsymbol{\Sigma}_i$$

$$\boldsymbol{V}_{21}^{i} = \frac{1}{n_i}\left(\boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i + \boldsymbol{\mu}_i \otimes \boldsymbol{\Sigma}_i\right) \quad \text{(Strydom \& Crowther (2012))}$$

$$\boldsymbol{V}_{12}^{i} = \boldsymbol{V}_{21}^{i\,\prime}$$

$$\boldsymbol{V}_{22}^{i} = \frac{1}{n_i}\left(I_{p^2} + \boldsymbol{K}\right)\left[\boldsymbol{\Sigma}_i \otimes \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i \otimes \boldsymbol{\mu}_i\boldsymbol{\mu}_i' + \boldsymbol{\mu}_i\boldsymbol{\mu}_i' \otimes \boldsymbol{\Sigma}_i\right] \quad \text{(Muirhead (1982, page 518)).}$$

To fit the model of homogeneity of two covariance matrices of four multivariate normal variables (cf. (10)), in terms of the canonical statistics (13), the appropriate constraints are now given by:

$$\boldsymbol{g}(\boldsymbol{m}) = \begin{pmatrix} \boldsymbol{g}_1(\boldsymbol{m}) \\ \boldsymbol{g}_2(\boldsymbol{m}) \\ \boldsymbol{g}_3(\boldsymbol{m}) \end{pmatrix}, \boldsymbol{g}_i(\boldsymbol{m}) = \begin{pmatrix} vec(\boldsymbol{\Sigma}_i)[1] - vec(\boldsymbol{\Sigma}_i)[11] \\ vec(\boldsymbol{\Sigma}_i)[2] - vec(\boldsymbol{\Sigma}_i)[12] \\ vec(\boldsymbol{\Sigma}_i)[6] - vec(\boldsymbol{\Sigma}_i)[16] \end{pmatrix}, i = 1, 2, \quad \boldsymbol{g}_3(\boldsymbol{m}) = vec(\boldsymbol{\Sigma}_1) - vec(\boldsymbol{\Sigma}_2).$$

The corresponding derivatives are given by

$$\boldsymbol{G}_m = \begin{pmatrix} \frac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}[1,] - \frac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}[11,] \\ \frac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}[2,] - \frac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}[12,] \\ \frac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}[6,] - \frac{\partial vec(\boldsymbol{\Sigma}_i)}{\partial \boldsymbol{m}}[16,] \end{pmatrix}, i = 1, 2, \quad \boldsymbol{G}_{3m} = \frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} - \frac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}}$$

where

$$\frac{\partial vec(\boldsymbol{\Sigma}_1)}{\partial \boldsymbol{m}} = \frac{\partial[\boldsymbol{m}_{12} - vec(\boldsymbol{m}_{11}\boldsymbol{m}_{11}')]}{\partial \boldsymbol{m}} = (-\boldsymbol{I}_p \otimes \boldsymbol{m}_{11} - \boldsymbol{m}_{11} \otimes \boldsymbol{I}_p, \ \ \boldsymbol{I}_{p^2}, \ \ \boldsymbol{0}_{p^2\times p}, \ \ \boldsymbol{0}_{p^2\times p^2})$$

$$\frac{\partial vec(\boldsymbol{\Sigma}_2)}{\partial \boldsymbol{m}} = \frac{\partial[\boldsymbol{m}_{22} - vec(\boldsymbol{m}_{21}\boldsymbol{m}_{21}')]}{\partial \boldsymbol{m}} = (\boldsymbol{0}_{p^2\times p}, \ \ \boldsymbol{0}_{p^2\times p^2}, \ \ -\boldsymbol{I}_p \otimes \boldsymbol{m}_{21} - \boldsymbol{m}_{21} \otimes \boldsymbol{I}_p, \ \ \boldsymbol{I}_{p^2}).$$

6 Conclusion

The maximum likelihood estimation method used in this paper is a very flexible procedure for estimation in the Wishart class of distributions. It greatly simplifies maximum likelihood methodology for the analysis of covariance structures. In addition to its flexibility for modelling a single patterned covariance matrix, it is appropriate for modelling in the case of independent as well as dependent multivariate distributions in the exponential class. The methodology also provides for covariance matrices with non-linear structure, e.g. matrices with AR(1) or Toeplitz structure.

The procedure for maximum likelihood estimation under constraints provides a unified approach to solving estimation problems in a wide variety of applications where generally, estimation and hypothesis testing can be problematic.

References

Anderson, T.W., 1970. Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In: Bose, R.C. & Roy, S.N. (Eds), Essays in Probability and Statists, 1-24. Chapel Hill: University of North Carolina Press.

Anderson, T.W., 1973. Asymptotically efficient estimation of covariance matrices with linear structure. Ann. Statist., 1, 135-141.

Barndorff-Nielsen, O., 1982. Exponential Families. In: Kotz,S., Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences, Vol.2, 587-596. John Wiley & Sons, New York.

Brown, L.D., 1986. Fundamentals of Exponential Families with Applications in Statistical Decision Theory. Institute of Mathematical Statistics, Lecture Notes-Monograph Series, Volume 9.

Calvin, J.A.& Dykstra, R.L., 1995. REML estimation of covariance matrices with restricted parameter spaces. J. Amer. Statist. Assoc. 90, 321-329.

Chaudhuri, S., Drton, M., Richardson, T.S., 2007. Estimation of a covariance matrix with zeros. Biometrika, 94, 199-216.

Jiang, G., Sarkar, S.K. & Hsuan, F., 1999. A likelihood ratio test and its modifications for the homogeneity of the covariance matrices of dependent multivariate normals. Journal of Statistical Planning and Inference, 81, 95-111.

Matthews, G.B. & Crowther, N.A.S. 1995. A maximum likelihood estimation procedure when modelling in terms of constraints. S.A. Statist. J., 29, 29-50.

Matthews, G.B. & Crowther, N.A.S., 1998. A maximum likelihood estimation procedure for the generalized linear model with restrictions. S.A. Statist. J., 32, 119-144.

Muirhead, R.J., 1982,. Aspects of Multivariate Statistical Theory, New York: John Wiley & Sons, Inc.

Rencher, A.C., 1998,. Multivariate Statistical Inference and Applications, New York: John Wiley & Sons, Inc.

Rubin, D.B. & Szatrowski, T.H., 1982. Finding maximum likelihood estimates of patterned covariance matrices by the EM algorithm. Biometrika, 69(3), 657-660.

Strydom, H.F. & Crowther, N.A.S., 2012. Maximum likelihood estimation for multivariate normal samples. S.A. Statist. J., 46, 115-153.

Szatrowski, T.H., 1980. Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. Ann. Statist., 8, 802-810.