Maximum Likelihood Estimation of Some Patterned 2 x 2 Covariance Matrices

by

Bernhard Flury*
Purdue University and University of Berne

Technical Report #83-9

Department of Statistics
Purdue University

April 1983

*This work was partly supported by grant no. 82.008.0.82 of the Swiss National Science Foundation. I wish to thank Henri Carnal, David Moore, K.C.S. Pillai and Herman Rubin for some helpful discussions and remarks.

Abstract

Assume the bivariate random vectors $Y_{(i)}$ ($i = 1, \ldots, k$) are independently distributed as $N_2(0, \Sigma_{(i)})$, where $\Sigma_{(i)} = \begin{pmatrix} \sigma_{11}^{(i)} & \sigma_{12}^{(i)} \\ \sigma_{12}^{(i)} & \sigma_{22}^{(i)} \end{pmatrix}$. We wish to test the hypothesis

$$\frac{\sigma_{12}^{(i)}}{a\sigma_{11}^{(i)} + b\sigma_{22}^{(i)}} = \frac{\sigma_{12}^{(j)}}{a\sigma_{11}^{(j)} + b\sigma_{22}^{(j)}} \quad (i,j = 1, \ldots, k),$$

where a and b are fixed constants. The ML-estimates of $\Sigma_{(i)}$ under this hypothesis are derived, and their consistency is proved. The practical meaning of some special choices of a and b is explained. The case $a = 1$, $b = 0$ is discussed in detail for $k = 2$ populations. Generalizations to higher dimensions are outlined.

Keywords: bivariate normal distribution; log-likelihood ratio; characteristic vectors; equality of regression slopes

1. Introduction

It is often useful to model multivariate normal data with a priori restrictions on the covariance matrix. Such restrictions can arise from situations where some variables are independent, conditionally independent, equally correlated, etc. In a broad sense, most tests of one-sample-hypotheses about covariance matrices (Anderson 1958, p. 230ff, 247ff, Morrison 1976, p. 247, 250, 253) can be considered as tests for special patterns. ML-estimation of some special patterns is discussed by Anderson (1970, 1973), Szatrowski (1976, 1978, 1980) and by Rubin and Szatrowski (1981).

In contrast to this, the simultaneous estimation of $k \geq 2$ covariance matrices under restrictions of some parameters has not been treated. A frequent assumption is the one of equality of all k covariance matrices. When the hypothesis of equality is rejected, the usual reaction is to estimate every matrix individually. However, there might be some "similarity" between two or more matrices which could lead to a reduction of the number of parameters. This paper treats the special case of 2 dimensions and the restriction that in every population the ratio of the covariance to a weighted sum of the variances is the same, that is

$$H_0: \frac{\sigma_{12}^{(i)}}{a\sigma_{11}^{(i)} + b\sigma_{22}^{(i)}} = \frac{\sigma_{12}^{(j)}}{a\sigma_{11}^{(j)} + b\sigma_{22}^{(j)}} \quad \forall\, i,j \leq k \tag{1.1}$$

where a and b are fixed constants, and

$$\Sigma_{(i)} = \begin{pmatrix} \sigma_{11}^{(i)} & \sigma_{12}^{(i)} \\ \sigma_{12}^{(i)} & \sigma_{22}^{(i)} \end{pmatrix} \tag{1.2}$$

is the covariance matrix of the i-th population $(i = 1, \ldots, k)$.

Actually, the study of this problem was motivated by investigations on two properties of the characteristic vectors of the matrix $\Sigma_{(1)}^{-1}\Sigma_{(2)}$, which turned out to be special cases of the problem stated above, as discussed in section 4 of this paper. Since it was found that a common treatment could be given to both cases, the general solution is given here rather than the solutions of the two special cases.

2. Derivation of the ML-estimates

Assume $Y_{(i)}$ $i = 1, \ldots, k$ are independently distributed as $N_2(0,\Sigma_{(i)})$, $\Sigma_{(i)} > 0$. Samples $(Y_{(i)\ell})$, $\ell = 1, \ldots, n_i$ $(n_i \geq 2)$ are taken from the k populations. Let

$$S_{(i)} = \frac{1}{n_i} \sum_{\ell=1}^{n_i} Y_{(i)\ell} Y'_{(i)\ell} = \begin{pmatrix} s_{11}^{(i)} & s_{12}^{(i)} \\ s_{12}^{(i)} & s_{22}^{(i)} \end{pmatrix} \tag{2.1}$$

denote the sample covariance matrices. Then the common log-likelihood function of the k samples is

$$\ell(\Sigma_{(1)}, \ldots, \Sigma_{(k)}) = -\frac{1}{2} \sum_{i=1}^{k} \{n_i(\log|\Sigma_{(i)}| + tr(\Sigma_{(i)}^{-1} S_{(i)}))\}$$

$$- \log (2\pi) \sum_{i=1}^{k} n_i \tag{2.2}$$

While in the unrestricted case $S_{(i)}$ results as the ML-estimate for $\Sigma_{(i)}$ (Muirhead 1982, p. 84), estimation turns out to be more complicated under H_0.

Assume that $ab \neq 0$ and $a\sigma_{11}^{(i)} + b\sigma_{22}^{(i)} \neq 0$ $(i = 1, \ldots, k)$. Now we reparametrize as follows:

$$\left.\begin{aligned} \varphi_1^{(i)} &= a\sigma_{11}^{(i)} + b\sigma_{22}^{(i)} \\ \varphi_2^{(i)} &= a\sigma_{11}^{(i)} - b\sigma_{22}^{(i)} \\ c &= \sigma_{12}^{(i)}/\varphi_1^{(i)} \end{aligned}\right\} \quad i = 1, \ldots, k \qquad (2.3)$$

Thus the parameter space under H_0 is

$$p = \{\Sigma_{(1)}, \ldots, \Sigma_{(k)}: \Sigma_{(i)} = \begin{pmatrix} \frac{\varphi_1^{(i)} + \varphi_2^{(i)}}{2a} & c\,\varphi_1^{(i)} \\ c\,\varphi_1^{(i)} & \frac{\varphi_1^{(i)} - \varphi_2^{(i)}}{2b} \end{pmatrix}, \Sigma_{(i)} > 0, \forall i\} \qquad (2.4)$$

Since the reparametrization (2.3) is nonsingular, $\Sigma_{(i)}$ can be estimated by estimating $\varphi_1^{(i)}$, $\varphi_2^{(i)}$ and c (Anderson 1958, lemma 3.2.3., p. 47). Note that the reduced parameter space p contains only $2k + 1$ parameters, compared with $3k$ in the unrestricted case.

To simplify notation, let $D_i = |\Sigma_{(i)}|$ and $Z_i = \sigma_{11}^{(i)}s_{22}^{(i)} + \sigma_{22}^{(i)}s_{11}^{(i)} - 2\sigma_{12}^{(i)}s_{12}^{(i)}$, then $\mathrm{tr}\,(\Sigma_{(i)}^{-1}S_{(i)}) = Z_i/D_i$. Note that in terms of the new parameters we have

$$D_i = \frac{1}{4ab}\left[(1-4abc^2)\ \varphi_1^{(i)^2} - \varphi_2^{(i)^2}\right], \qquad i = 1, \ldots, k \qquad (2.5)$$

and

$$Z_i = \frac{1}{2ab}\left[(w_1^{(i)} - 4abc\ s_{12}^{(i)})\ \varphi_1^{(i)} - w_2^{(i)}\ \varphi_2^{(i)}\right],\ i = 1, \ldots, k \qquad (2.6)$$

where we have set

$$\left.\begin{aligned} w_1^{(i)} &= as_{11}^{(i)} + bs_{22}^{(i)} \\ w_2^{(i)} &= as_{11}^{(i)} - bs_{22}^{(i)} \end{aligned}\right\}\quad i = 1, \ldots, k \qquad (2.7)$$

Now, equivalently to a maximization of (2.2), we can minimize the function

$$h((\varphi_1^{(i)}, \varphi_2^{(i)}),\ i = 1, \ldots, k,\ c) = \sum_{i=1}^{k} n_i\ (\log D_i + Z_i/D_i) \qquad (2.8)$$

Setting the first derivatives of (2.8) with respect to $\varphi_1^{(i)}$ and $\varphi_2^{(i)}$ equal to zero yields

$$(1-4abc^2)(Z_i - D_i)\ \varphi_1^{(i)} = (w_1^{(i)} - 4\ abcs_{12}^{(i)})\ D_i \qquad (2.9)$$

and

$$(Z_i - D_i)\ \varphi_2^{(i)} = w_2^{(i)} D_i \qquad (2.10)$$

Assume now that $4abc^2 \neq 1$, a condition which we will have to discuss later. Then solving (2.9) for $w_2^{(i)}(Z_i - D_i)$, putting this into (2.10) and solving for $w_2^{(i)}\varphi_1^{(i)}$ yields

$$w_2^{(i)} \varphi_1^{(i)} = \frac{w_1^{(i)} - 4abcs_{12}^{(i)}}{1-4abc^2} \varphi_2^{(i)} \qquad (2.11)$$

Putting (2.11) into (2.5) and (2.6) and some algebraic manipulation gives

$$w_2^{(i)^2} D_i = \frac{1}{4ab} \varphi_2^{(i)^2} g_i \qquad (2.12)$$

and

$$w_2^{(i)^2} Z_i = \frac{1}{2ab} w_2^{(i)} \varphi_2^{(i)} g_i \qquad (2.13)$$

and therefore

$$w_2^{(i)^2}(Z_i - D_i) = \frac{1}{2ab} \varphi_2^{(i)} g_i (w_2^{(i)} - \frac{1}{2} \varphi_2^{(i)}) \qquad (2.14)$$

where we have set

$$g_i = \frac{(w_1^{(i)} - 4abcs_{12}^{(i)})^2}{1 - 4abc^2} - w_2^{(i)^2} \qquad (2.15)$$

Putting (2.12) and (2.14) into (2.10) gives

$$\frac{1}{2ab} \varphi_2^{(i)^2} g_i (w_2^{(i)} - \frac{1}{2} \varphi_2^{(i)}) = w_2^{(i)} \frac{1}{4ab} g_i \varphi_2^{(i)^2} \qquad (2.16)$$

Assuming that $g_i \neq 0$, it follows that

$$w_2^{(i)} \; {\varphi_2^{(i)}}^2 = {\varphi_2^{(i)}}^3 \qquad (2.17)$$

The solution $\varphi_2^{(i)} = 0$ of (2.17) would imply either $w_2^{(i)} = 0$ or $D_i = 0$ by (2.10). The latter case would mean singularity of $\hat{\Sigma}_{(i)}$. Therefore we have the ML-estimate

$$\hat{\varphi}_2^{(i)} = w_2^{(i)} = as_{11}^{(i)} - bs_{22}^{(i)} \qquad (2.18)$$

Note that this solution holds also for $w_2^{(i)} = 0$, since we avoided division by $w_2^{(i)}$ in the above derivation. This result shows the rather surprising fact that the ML-estimate $\hat{\varphi}_2^{(i)}$ does not depend on the common parameter c. Actually, this is the justification for the reparametrization (2.3).

Before deriving the estimates for $\varphi_1^{(i)}$ $(i = 1, \ldots, k)$ and c, let us check the two assumptions made above.

Assumption $4abc^2 \neq 1$: Assume the opposite is true. Then we have $c^2 = 1/4ab$, and, by (2.9), $c = w_1^{(i)}/4abs_{12}^{(i)}$, since $D_i > 0$. Combining these two equations gives ${w_1^{(i)}}^2 = 4ab{s_{12}^{(i)}}^2$, or ${w_1^{(i)}}^2 - 4abs_{11}^{(i)}s_{22}^{(i)} = {w_2^{(i)}}^2 = 4ab(s_{12}^{(i)} - {s_{11}^{(i)}}^2 s_{22}^{(i)})$. Since $ab > 0$ in this case, the last equation can only hold if $w_2^{(i)} = \det S_{(i)} = 0$, an event which occurs with probability 0.

Assumption $g_i \neq 0$: Otherwise we would have

$$(w_1^{(i)} - 4\,abcs_{12}^{(i)})^2 = {w_2^{(i)}}^2\,(1-4abc^2) \qquad (2.19)$$

from which we could get an estimate of c by solving for c. Written as a polynomial in c, (2.19) becomes

$$(4abs_{12}^{(i)^2} + w_2^{(i)^2})\, c^2 - 2w_1^{(i)} s_{12}^{(i)} c + s_{11}^{(i)} s_{22}^{(i)} = 0 \qquad (2.20)$$

The discriminant of this equation is $4w_2^{(i)^2} (s_{12}^{(i)^2} - s_{11}^{(i)} s_{22}^{(i)})$, and a real solution exists only if $w_2^{(i)} = 0$ and det $S_{(i)} = 0$, which has probability 0. Therefore $g_i \neq 0$ can be assumed without loss of generality.

From (2.18) and (2.10) it follows now easily that

$$Z_i = 2D_i \quad (i = 1, \ldots, k) \qquad (2.21)$$

and therefore tr $\hat{\Sigma}_{(i)}^{-1} S_{(i)} = 2$ must hold for the ML-estimates $\hat{\Sigma}_{(1)}, \ldots, \hat{\Sigma}_{(k)}$. Putting (2.18) into (2.11) gives the ML-estimate for $\varphi_1^{(i)}$:

$$\hat{\varphi}_1^{(i)} = \frac{w_1^{(i)} - 4abcs_{12}^{(i)}}{1-4abc^2} = \frac{as_{11}^{(i)} + bs_{22}^{(i)} - 4abcs_{12}^{(i)}}{1-4abc^2} \qquad (2.22)$$

which depends on the estimate of the common parameter c. For estimating c, we need the first derivative of h with respect to c, set this equal to zero and use (2.21) to obtain

$$\sum_{i=1}^{k} n_i\, \varphi_1^{(i)} (c\, \varphi_1^{(i)} - s_{12}^{(i)})/D_i = 0 \qquad (2.23)$$

or, by multiplication with $\prod_{i=1}^{k} D_i/(n_1 + \ldots + n_k)$

$$\sum_{i=1}^{k} f_i \; \varphi_1^{(i)} \; (c \; \varphi_1^{(i)} - s_{12}^{(i)}) \quad \prod_{\substack{j=1 \\ j \neq i}}^{k} D_i = 0 \qquad (2.24)$$

where we have set $f_i = n_i / \sum_{j=1}^{k} n_j$.

Putting the estimates $\hat{\varphi}_1^{(i)}$ (2.22) and $\hat{\varphi}_2^{(i)}$ (2.18) into (2.5) yields

$$D_i = \frac{(w_1^{(i)} - 4abcs_{12}^{(i)})^2 - (1 - 4abc^2)\, w_2^{(i)^2}}{4ab(1 - 4abc^2)} \qquad (2.25)$$

$$= \frac{s_{11}^{(i)} s_{22}^{(i)} - 2w_1^{(i)} s_{12}^{(i)} c + (w_2^{(i)^2} + 4abs_{12}^{(i)^2}) c^2}{1 - 4abc^2}$$

and similarly we get

$$\varphi_1^{(i)} (c \; \varphi_1^{(i)} - s_{12}^{(i)}) = \frac{(w_1^{(i)} - 4\; abcs_{12}^{(i)})(w_1^{(i)} c - s_{12}^{(i)})}{(1 - 4abc^2)^2} . \qquad (2.26)$$

Thus multiplication of (2.24) with $(1 - 4abc^2)^{k+1}$ yields the equation

$$P(c) = \sum_{i=1}^{k} f_i \; [-w_1^{(i)} s_{12}^{(i)} + (w_1^{(i)^2} + 4abs_{12}^{(i)^2}) \; c - 4abw_1^{(i)} s_{12}^{(i)} c^2] \times$$

$$\times \prod_{\substack{j=1 \\ j \neq i}}^{k} [s_{11}^{(j)} s_{22}^{(j)} - 2w_1^{(j)} s_{12}^{(j)} \; c + (w_2^{(j)^2} + 4abs_{12}^{(j)^2}) \; c^2] = 0 \qquad (2.27)$$

$P(c)$ is a polynomial of degree 2k. Since for $S_{(i)} > 0$, the likelihood-function (2.2) must have a maximum, (2.27) has at least one real root, and therefore at least two. The ML-estimate $\hat{c}$ must thus be found by computing $\hat{\varphi}_1^{(i)}$ (2.22) and the value of the log-likelihood-function (2.2), taking as $\hat{c}$ the root which yields the largest value of (2.2). Let $\hat{\sigma}_{11}^{(i)} = (\hat{\varphi}_1^{(i)} + \hat{\varphi}_2^{(i)})/2a$,

$\hat{\sigma}_{22}^{(i)} = (\hat{\varphi}_1^{(i)} - \hat{\varphi}_2^{(i)})/2b$, $\hat{\sigma}_{12}^{(i)} = \hat{c}\ \hat{\varphi}_1^{(i)}$ and $\hat{\Sigma}_{(i)} = \begin{pmatrix} \hat{\sigma}_{11}^{(i)} & \hat{\sigma}_{12}^{(i)} \\ \hat{\sigma}_{12}^{(i)} & \hat{\sigma}_{22}^{(i)} \end{pmatrix}$, then,

using (2.21), the maximum of the log-likelihood-function is simply

$$\ell(\hat{\Sigma}_{(1)}, \ldots, \hat{\Sigma}_{(k)}) = -\frac{1}{2} \sum_{i=1}^{k} n_i \log|\hat{\Sigma}_{(i)}| - (\log 2\pi + 1) \sum_{i=1}^{k} n_i \qquad (2.28)$$

and therefore the log-likelihood-ratio statistic (Silvey 1970, p. 113, Serfling 1980, p. 157) for testing H_0 is

$$-2(\ell(\hat{\Sigma}_{(1)}, \ldots, \hat{\Sigma}_{(k)}) - \ell(S_{(1)}, \ldots, S_{(k)}) = \sum_{i=1}^{k} n_i \log \frac{|\hat{\Sigma}_{(i)}|}{|S_{(i)}|} \qquad (2.29)$$

Numerical examples show that "normally" P(c) has exactly 2 real roots, and more than 2 real roots occur when two or more of the sample covariance matrices $S_{(i)}$ are nearly singular and "far" from the pattern defined as H_0. In this case, some of the real roots of P(c) may even lead to non-positive definite estimates of $\Sigma_{(i)}$. No exact conditions are available to determine the number of real roots except for special cases (see section 4).

In deriving the above results we made use of the assumption that $\varphi_1^{(i)} \neq 0\ \forall i$. Actually, this assumption was already used in the form in which H_0: $\sigma_{12}^{(i)}/\varphi_1^{(i)} = \sigma_{12}^{(j)}/\varphi_1^{(j)}\ \forall\ i,j$ was stated. However if we write H_0 as $\sigma_{12}^{(i)}\ \varphi_1^{(j)} = \sigma_{12}^{(j)}\ \varphi_1^{(i)}\ \forall\ i,j$, there is no reason why $\varphi_1^{(i)} \neq 0$ should be assumed, and the reason to prefer the former version was only its solvability. Since in practical applications the case $\varphi_1^{(i)} = 0$ might occur, we give it a special treatment.

Suppose, for simplicity, that $\varphi_1^{(1)} = 0$. This implies that either a) $\sigma_{12}^{(1)} = 0$ or b) $\varphi_1^{(2)} = \ldots = \varphi_1^{(k)} = 0$.

Case a) In this case, $\Sigma_{(1)} = \begin{pmatrix} \varphi_2^{(1)}/2a & 0 \\ 0 & -\varphi_2^{(1)}/2b \end{pmatrix}$, which does not depend on the common parameter c, and therefore the likelihood function can be factorized in a part involving only $\Sigma_{(1)}$ and a part involving $\Sigma_{(2)}$ to $\Sigma_{(k)}$. ML-estimation of $\Sigma_{(1)}$ does therefore not depend on the second to k-th sample, and it can easily be shown that $\hat{\varphi}_2^{(1)} = w_2^{(1)} = as_{11}^{(1)} - bs_{22}^{(1)}$. (Actually, in this case the variances of two independent normal variates are estimated given a fixed constant of proportionality).

Case b) In this case, $\Sigma_{(i)} = \begin{pmatrix} \varphi_2^{(i)}/2a & \sigma_{12}^{(i)} \\ \sigma_{12}^{(i)} & \varphi_2^{(i)}/2b \end{pmatrix}$, and the likelihood function can be written as a product of k independent factors. It can be shown that $\hat{\varphi}_2^{(i)} = w_{(2)}^{(i)} = as_{11}^{(i)} - bs_{22}^{(i)}$ and $\hat{\sigma}_{12}^{(i)} = s_{12}^{(i)}$ $(i = 1, \ldots, k)$.

In finite samples from non-singular bivariate normal distributions, $\hat{\varphi}_1^{(i)} = 0$ will occur only with probability zero. However, values of $\hat{\varphi}_1^{(i)}$ close to zero indicate that a further reduction of the parameter space to the cases a) and b) above might be possible.

3. Consistency of the ML-estimates

If $\Sigma_{(i)} > 0$ $(i=1, \ldots, k)$, the asymptotic theory of likelihood ratio tests (Rao 1973) can be applied. Therefore, the log-likelihood-ratio statistic (2.29) is asymptotically ($\min(n_1, \ldots, n_k) \to \infty$) distributed as chi square with k-1 degrees of freedom, if H_0 is true.

However, we can learn more about the nature of the ML-estimates $\hat{\Sigma}_{(i)}$ by proving their consistency under H_0 directly. To do this, let us assume that $\varphi_1^{(i)}$, $\varphi_2^{(i)}$ and c are the true parameters of the i-th population, and write the polynomial (2.27) as a function of x rather than of c. If $\min_{1 \leq i \leq k} n_i \to \infty$, we have the following convergences in probability:

$$\left.\begin{aligned} \hat{\varphi}_2^{(i)} &= w_2^{(i)} = as_{11}^{(i)} - bs_{22}^{(i)} \to \varphi_2^{(i)} \\ w_1^{(i)} &= as_{11}^{(i)} + bs_{22}^{(i)} \to \varphi_1^{(i)} \\ s_{12}^{(i)} &\to \sigma_{12}^{(i)} = c\, \varphi_1^{(i)} \end{aligned}\right\} \quad i = 1, \ldots, k \qquad (3.1)$$

The polynomial (2.27) converges therefore to

$$P^*(x) = \sum_{i=1}^{k} f_i^* \left[-c\, \varphi_1^{(i)2} + \left(\varphi_1^{(i)2} + 4abc^2\, \varphi_1^{(i)2}\right)x - 4ab\,\varphi_1^{(i)2} c\, x^2\right] R_i^*(x) \qquad (3.2)$$

where

$$f_i^* = \lim_{\min n_j \to \infty} n_i/(n_1 + \ldots + n_k) \qquad (3.3)$$

$$R_i^*(x) = (1 - 4abx^2)^{k-1} \prod_{\substack{j=1 \\ j \neq i}}^{k} D_j^*(x) \tag{3.4}$$

and

$$D_j^*(x) = \frac{\sigma_{11}^{(j)}\sigma_{22}^{(j)} - 2c\varphi_1^{(j)^2} x + (\varphi_2^{(j)^2} + 4abc^2 \varphi_1^{(j)^2})x^2}{1 - 4abx^2} \tag{3.5}$$

Since $D_j^*(c) = |\Sigma_{(j)}| > 0$ by the assumption of non-singularity, and since $1 - 4abc^2 > 0$ (see below), it follows that $R_i^*(x) > 0$ for x in a sufficiently small neighborhood of c.

Since the term in rectangular brackets of (3.2), is

$$Q^*_i(x) = \varphi_1^{(i)^2}(c-x)(4abcx-1) \tag{3.6}$$

we have

$$P^*(x) = (c-x)(4abcx-1) \sum_{i=1}^{k} f_i^* \varphi_1^{(i)^2} R_i^*(x) \tag{3.7}$$

and thus clearly $P^*(c) = 0$.

As a single real root of a polynomial is a continuous function of the polynomial coefficients, it remains to show that $P^*(x)$ has no second root at $x = c$. Since $f_i^* \geq 0$ and $R_i^*(x) > 0$ for x in a neighborhood of c, we have to show that the root $x = 1/4abc$ cannot be equal to c. Otherwise we would have $c = 1/4abc$ or $c^2 = 1/4ab$, which is only possible if $ab>0$. The determinant of the i-th covariance matrix would be $((1-4abc^2)\varphi_1^{(i)^2} - \varphi_2^{(i)^2})/4ab = -\varphi_2^{(i)^2}/4ab \leq 0$. This contradicts the assumption of all covariance matrices being positive definite. Therefore, one root of the polynomial (2.27) provides a consistent estimate of c, and the asymptotic theory of maximum likelihood indicates that this is the root which maximises (2.2).

4. The meaning of three special choices of a and b, and their generalizations to higher dimensions

As already mentioned in the introduction, this research was motivated by a study of certain properties of the matrix $\Sigma_{(1)}^{-1}\Sigma_{(2)}$. From now on we shall therefore restrict our considerations to the case of k = 2 populations. However, it must be borne in mind that if the pairs $(\Sigma_{(1)},\Sigma_{(2)})$ and $(\Sigma_{(2)},\Sigma_{(3)})$ satisfy H_0 for fixed a and b, the same is not necessarily true for $(\Sigma_{(1)},\Sigma_{(3)})$; that is, the property defined by H_0 is not transitive. Counter-examples to transitivity can easily be found by setting $\varphi_1 = 0$ for one population.

We will denote $\Omega = \begin{pmatrix}\omega_{11} & \omega_{12}\\ \omega_{21} & \omega_{22}\end{pmatrix} = \Sigma_{(1)}^{-1}\Sigma_{(2)}$ throughout this section.

The case a = b: In this case, $\omega_{21} = -\omega_{12}$, and if $(x_1,x_2)'$ is a characteristic vector of Ω, then the other characteristic vector is $(x_2,x_1)'$. An obvious generalization to higher dimensions is the case where $\Sigma_{(1)}^{-1}\Sigma_{(2)}$ can be written as the sum of a diagonal and a skew-symmetric matrix. However, this case does not seem to have any practical importance.

The case a = -b: Ω is symmetric in this case, and the generalization is obvious. It can be shown (Flury 1982) that $\Sigma_{(1)}$ and $\Sigma_{(2)}$ have the same characteristic vectors, and these are identical with the characteristic vectors of $\Sigma_{(1)}^{-1}\Sigma_{(2)}$. Thus this case has a relation to the principal component analysis of two (or more) groups, and the ML-estimation of the common principal axes of several populations is presently under study.

The case a = 1, b = 0: In this case, Ω is upper triangular ($\omega_{21} = 0$), the diagonal elements of Ω are the characteristic roots, and the characteristic vector associated to ω_{11} is $(1,0)'$. Note that the above derivation of the ML-estimates does not hold since ab = 0. However, we will discuss this case in detail in the next section. A straightforward generalization to higher dimensions is the condition that $\Sigma_{(1)}^{-1}\Sigma_{(2)}$ is triangular. A less restrictive generalization is as follows: Partition Ω as $\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$, where Ω_{11} is q x q and Ω_{21} is (p-q) x q, and let $\beta = (\beta_1, \ldots, \beta_q)'$ denote a characteristic vector of Ω_{11}. Then the generalization is to request $\Omega_{21}\beta = 0$, in which case $(\beta_1, \ldots, \beta_q, 0, \ldots, 0)'$ is a characteristic vector of Ω. If $q = 1$, $\Omega_{11} = \omega_{11}$ is scalar, its only characteristic vector is the scalar 1, and therefore $\Omega_{21}\beta = 0$ implies $\Omega_{21} = 0$. Thus $\Omega = \begin{pmatrix} \omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{pmatrix}$, and it can easily be seen that $(1,0, \ldots, 0)'$ is a characteristic vector of Ω associated to the characteristic value ω_{11}, while the remaining p-1 characteristic values of Ω are identical with those of Ω_{22}.

Note that the hypothesis $\sigma_{12}^{(1)}/\sigma_{11}^{(1)} = \sigma_{12}^{(2)}/\sigma_{11}^{(2)}$ states that $(1,0)'$ is a characteristic vector of $\Sigma_{(1)}^{-1}\Sigma_{(2)}$, and nothing is implied about the associated characteristic root or the second characteristic vector. If we wish to test simultaneously whether the characteristic root not associated with the hypothetical characteristic vector $(1,0)'$ is equal to unity, a test dervied by Khatri and Pillai (1969, p. 236ff) can be used.

Another interpretation of the hypothesis is equality of regression slopes in two populations. However, our approach is more general in the sense that it avoids conditioning as well as the assumption of equality of the residual variance in both populations.

5. The case a = 1, b = 0

As stated above, the derivation of the ML-estimates given in section 2 does not hold in this case. However, it is very similar and can be done without reparametrization (see Flury 1982). The result (in the case of k = 2 samples) is:

$$\left.\begin{aligned} \hat{\sigma}_{11}^{(i)} &= s_{11}^{(i)} \\ \hat{\sigma}_{22}^{(i)} &= s_{22}^{(i)} - 2\,\hat{c}s_{12}^{(i)} + 2\,\hat{c}^2 s_{11}^{(i)} \end{aligned}\right\} \quad i = 1,2 \tag{5.1}$$

and $\hat{c}$ is a real root of the polynomial $a_3x^3 - a_2x^2 + a_1x - a_0$, where

$$a_3 = s_{11}^{(1)}s_{11}^{(2)}$$

$$a_2 = [n_1(2s_{11}^{(1)}s_{12}^{(2)} + s_{12}^{(1)}s_{11}^{(2)}) + n_2(2s_{12}^{(1)}s_{11}^{(2)} + s_{11}^{(1)}s_{12}^{(2)})]/(n_1 + n_2)$$

$$a_1 = [n_1(s_{11}^{(1)}s_{22}^{(2)} + 2s_{12}^{(1)}s_{12}^{(2)}) + n_2(s_{22}^{(1)}s_{11}^{(2)} + 2s_{12}^{(1)}s_{12}^{(2)})]/(n_1 + n_2)$$

$$a_0 = (n_1s_{12}^{(1)}s_{22}^{(2)} + n_2s_{22}^{(1)}s_{12}^{(2)})/(n_1 + n_2) \tag{5.2}$$

If we set $\hat{\sigma}_{12}^{(i)} = \hat{c}\hat{\sigma}_{11}^{(i)}$ and $\hat{\Sigma}_{(i)} = \begin{pmatrix} \hat{\sigma}_{11}^{(i)} & \hat{\sigma}_{12}^{(i)} \\ \hat{\sigma}_{12}^{(i)} & \hat{\sigma}_{22}^{(i)} \end{pmatrix}$, then the log-likelihood ratio statistic takes again the form (2.29).

Since the polynomial (5.2) has degree 3, it has either one or three real roots. We will show a little later that "normally" there is exactly one solution.

Any real root of (5.2) yields a positive definite estimate of $\hat{\Sigma}_{(i)}$ (i = 1,2). This can be seen as follows (where the group subscript i is omitted for simplicity): Since $\hat{\sigma}_{22}$ is a continuous function of $\hat{c}$, s_{11} and s_{22}, it suffices to show that $\hat{\sigma}_{22}$ can take positive values, but not zero. Putting $\hat{c} = 0$ in (5.1) shows that $\hat{\sigma}_{22}$ can take positive values. If $\hat{\sigma}_{22}$ were zero, then the ML-estimate of c would be a root of the equation

$$2\,x^2 s_{11} - 2xs_{12} + s_{22} = 0 \quad . \tag{5.3}$$

The discriminant of this equation is $s_{12}^2 - 2s_{11}s_{22} < 0$ (since $|S| > 0$), and therefore no real solution exists. This proves $\hat{\sigma}_{22} > 0$. Now since $0 \leq (\hat{c}s_{11} - s_{12})^2 = \hat{c}^2 s_{11}{}^2 - 2\hat{c}s_{11}s_{12} + s_{12}{}^2 = |\hat{\Sigma}| - |S|$, it follows that $|\hat{\Sigma}| \geq |S| > 0$. As also $\hat{\sigma}_{11} > 0$, $\hat{\Sigma}$ is positive definite.

If $H_{01}(\omega_{21} = 0)$ and $H_{02}(\omega_{12} = 0)$ hold simultaneously, Ω is diagonal, and the two characteristic vectors are (1,0)' and (0,1)' (or can be chosen in this way if $\omega_{11} = \omega_{22}$. Thus $H_{01} \cap H_{02}$ means that the two covariance matrices are either proportional, or the two variables are uncorrelated in both groups.

Let us now derive some results about the number of real roots of the polynomial (5.2). First we write (5.2) in the form

$$\frac{s_{11}^{(2)}x^2 - 2s_{12}^{(2)}x + s_{22}^{(2)}}{s_{11}^{(2)}x - s_{12}^{(2)}} \Bigg/ \frac{s_{11}^{(1)}x^2 - 2s_{12}^{(1)}x + s_{22}^{(1)}}{s_{11}^{(1)}x - s_{12}^{(1)}}$$

$$= \frac{(x-a_2) + h_2/(x-a_2)}{(x-a_1) + h_1/(x-a_1)} = -\frac{n_2}{n_1} = -\lambda < 0 \tag{5.4}$$

where

$$\left.\begin{aligned} a_i &= s_{12}^{(i)}/s_{11}^{(i)} \\ h_i &= (s_{11}^{(i)}s_{22}^{(i)} - s_{12}^{(i)^2})/s_{11}^{(i)^2} \end{aligned}\right\} \quad i = 1,2 \tag{5.5}$$

Letting $y = x - a_1$ and $d = a_2 - a_1$, (5.4) can be written as

$$f(y) = \frac{y[(y-d)^2 + h_2]}{(y-d)\ (y^2+h_1)} = -\lambda = -\frac{n_2}{n_1} \tag{5.6}$$

To solve (5.6), we note that $h_1 > 0$ and $h_2 > 0$ (since $|S_{(i)}| > 0$, $i = 1,2$). We assume in the sequel that $d > 0$, but a similar treatment can be given for the case $d < 0$. The $f(y) < 0 \Leftrightarrow 0 < y < d$, and the solution of (5.6) is unique iff $f(y)$ is monotonic in the interval $(0,d)$ that is, $f'(y) \leq 0 \ \forall\ y \in (0,d)$ or $f'(y) \geq 0 \ \forall y \in (0,d)$. It can be shown that $f'(y) = 0 \Leftrightarrow (y-d)^2(dy^2+2h_1y-dh_1) = h_2(2y^3-dy^2+ch_1)$. Thus we look for conditions under which the equation

$$r(y) = h_2 \cdot s(y)$$

where $$r(y) = (y-d)^2(dy^2+2h_1y-dh_1) = r_1(y)r_2(y) \tag{5.7}$$

and $$s(y) = 2y^3 - dy^2 + dh_1$$

has solutions.

For the function $s(y)$, $s(0) = dh_1 > 0$ and $s(d) = d(d^2+h_1) > 0$. s decreases in the interval $(0,d/3)$ and takes the local minimum $d(h_1-d^2/27)$ at $y = d/3$. In $(d/3,d)$ $s(y)$ is increasing. Thus, if $h_1 < d^2/27$, $s(y)$ takes positive and negative values, which shows that under certain conditions the polynomial (5.2) has three real roots.

For the function $r(y)$ we have $r(0) = -d^3h_1 < 0$, $r(\frac{d}{2}) = d^5/8 > 0$ and $r(d) = 0$. Therefore there is $y_0 \in (0,d/2)$ with $r(y_0) = 0$. Since y_0 must be a root of $r_2(y)$, it follows that

$$y_0 = h_1(\sqrt{1+d^2/h_1} - 1)/d \tag{5.8}$$

Therefore $r(y)$ takes positive values for all d and h_1, and (5.6) has solutions if h_2 is small enough. Asymptotically we have $\lim_{d\to\infty} y_0 = \sqrt{h_1}$, that is, $y_0 \approx \sqrt{h_1}$ if d is large. If $d^2/h_1 < 1$, we note that the function $\sqrt{1+x}$ can be written as the power series $1 + \frac{x}{2} - \frac{1}{8}x^2 + \frac{3}{48}x^3 - \ldots = 1 + \frac{x}{2} - \sigma(x)$, and therefore $\lim_{d\to 0} \frac{y_0}{d} = \frac{h_1(-1+1+\frac{d^2}{2h_1} - \sigma\left(\frac{d^2}{h_1}\right))}{d^2} = \frac{1}{2}$. Thus $y_0 \approx \frac{d}{2}$ for small d (but $y_0 < d/2$ holds for all d since $r(\frac{d}{2}) > 0$).

Since $r(y_0) = r(d) = 0$, r must take a local maximum at a point $y_1 \in (y_0,d)$. By differentiating r we get

$$y_1 = \sqrt{\left(\frac{3h_1-d^2}{4d}\right)^2 + h_1} - \frac{3h_1-d^2}{4d} \tag{5.9}$$

Asymptotically, we have $\lim_{d\to\infty} \frac{y_1}{d} = \frac{1}{2}$ and therefore $y_1 \approx \frac{d}{2}$ for large d.

For $d^2 < 3h_1$, we put $z = \frac{3h_1-d^2}{4d}$ and have therefore $y_1 = \sqrt{z^2+h_1} - z$. Since

$\lim_{d\to 0} z = \infty$, we look at the behavior of $\sqrt{z^2+h_1} - z$ for $z \to \infty$. Since

$(\sqrt{z^2+h_1} - z)(\sqrt{z^2+h_1} + z) = h_1$ and $\lim_{z\to\infty} \frac{\sqrt{z^2+h_1} + z}{2z} = 1$, it follows that

$\lim_{z\to\infty} 2z\,(\sqrt{z^2+h_1} - z) = h_1$. Therefore we have $h_1 = \lim_{d\to 0} \left(y_1 - \frac{3h_1 - d^2}{2d}\right) =$

$\lim_{d\to 0} \left[y_1\left(\frac{3h_1}{2d} - \frac{d}{2}\right)\right]$. Since $\lim_{d\to 0} y_1 d = 0$, $\lim_{d\to 0} \left(y_1 \cdot \frac{3h_1}{2d}\right) = h_1$ holds, and therefore $y_1 \approx \frac{2}{3} d$ for d small.

Since $r'(\frac{1}{2}d) = \frac{1}{2}d^2 h_1 > 0$ and $r'(\frac{2}{3}d) = -\frac{4}{27} a^4 < 0$, it follows that $\frac{1}{2}d < y_1 < \frac{2}{3}d$, and therefore we try to give conditions for the solvability of (5.7) based on the values of the functions $r(y)$ and $s(y)$ at $y = \frac{1}{2}d$ and $y = \frac{2}{3}d$.

<u>Bound for $y = \frac{2}{3}d$:</u> It can easily be checked that $r(\frac{2}{3}d) = d^3(4d^2+3h_1)/81$ and $s(\frac{2}{3}d) = d(4d^2 + 27h_1)/27$. For $k_1 = r(\frac{2}{3}d)/s(\frac{2}{3}d) = \frac{d^2(4d^2+3h_1)}{3(4d^2+27h_1)}$ we have $r(\frac{2}{3}d) = k_1 \cdot s(\frac{2}{3}d)$. As $\max_{0\le y\le d} r(y) \ge r(\frac{2}{3}d)$, (5.6) has at least one solution for $h_2 < k_1$. The critical values of n_1 and n_2 can be computed by putting $y = \frac{2}{3}d$ in (5.6).

<u>Bound for $y = \frac{1}{2}d$:</u> Thanks to $r(\frac{1}{2}d) = d\cdot(\frac{d}{2})^4$ and $s(\frac{d}{2}) = dh_1$, (5.7) has at least one solution for $h_2 \le k_2 = (\frac{d}{2})^4/h_1$. The critical values of n_1 and n_2 follow from (5.6) as $\frac{n_2}{n_1} = \frac{(d/2)^2}{h_1}$.

<u>Comparison of the two bounds</u>: For $h_2 \le k = \max(k_1, k_2)$, (5.7) has always at least one solution. By solving the equation $k_1 = k_2$ for d, it can easily

be seen that

$$k = \begin{cases} k_1 & \text{if} \quad d \leq 1.1889 \sqrt{h_1} \\ k_2 & \text{if} \quad d > 1.889 \sqrt{h_1} \end{cases} \tag{5.10}$$

Therefore k_1 gives the better bound for small d.

Thus we have found a condition under which the polynomial (5.2) has three real roots for certain sample sizes n_1 and n_2. However, it might also be interesting to know conditions under which the solution of (5.2) is unique for all n_1 and n_2, that is, equation (5.6) has no solution. This is certainly the case if $\max\limits_{0 \leq y \leq d} r(y) \leq h_2 \cdot \min\limits_{0 \leq y \leq d} s(y)$. While $\min\limits_{0 \leq y \leq d} s(y) = d(h_1 - d^2/27)$ was obtained above, the exact maximum of r(y) is a rather complicated expression, and we will give therefore an upper bound for it. Since at $y = y_1$ both $r_1(y)$ and $r_2(y)$ are positive, an upper bound for $r(y_1)$ can be obtained as the product of two upper bounds for $r_1(y_1)$ and $r_2(y_1)$. Since $\frac{1}{2}d < y_1 < \frac{2}{3}d$, $|y_1 - d| < \frac{d}{2}$ and $(y_1-d)^2 = r_1(y_1) < \frac{d^2}{4}$ hold. By the same argument, $r_2(y_1) = dy_1^2 + 2h_1y_1 - dh_1 \leq d \cdot (\frac{2}{3}d)^2 + 2h_1(\frac{2}{3}d) - dh_1 = \frac{1}{3}d(\frac{4}{3}d^2+h_1)$, and therefore

$$\max_{0 \leq y \leq d} r(y) = r(y_1) < \frac{d^3}{12}\left(\frac{4}{3}d^2+h_1\right) \tag{5.11}$$

Thus we get the bound

$$k^* = \frac{d^2(h_1+4d^2/3)}{12(h_1-d^2/27)} \tag{5.12}$$

and for $h_2 \geq k^*$ the polynomial (5.2) has exactly one real root for all sample sizes n_1, n_2.

It may be noted that, lacking a better solution, the problem was treated somehow asymmetrically, treating h_1 and h_2 in a different way. However, by exchanging $S_{(1)}$ and $S_{(2)}$, more conditions can be obtained.

As a numerical example, let $S_{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $S_{(2)} = \begin{pmatrix} 1 & 1 \\ 1 & w \end{pmatrix}$, that is, $d = 1$, $h_1 = 1$ and $h_2 = w-1$. Using (5.12) we get $k^* = 21/104$, which means that if $w \geq 1.2$, the solution of (5.2) will be unique for all n_1, n_2. On the other hand, using the simple bound $k_2 = (\frac{d}{2})^4/h_1$ (though k_1 would be a little more precise) we get $k_2 = 1/16$ and $\lambda = 1/4$, which means that (5.2) will have 3 real roots for $S_{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $S_{(2)} = \begin{pmatrix} 1 & 1 \\ 1 & 1.0625 \end{pmatrix}$, and $n_1 = 4n_2$. These roots are .8, .5 and .5.

The bounds k and k^* given above indicate clearly that three real roots of (5.2) must be expected when h_1 or h_2 are small compared with d, that is, when $S_{(1)}$ or $S_{(2)}$ is close to singularity. Moreover, under H_0, we expect d close to zero. Therefore the solution of (5.2) is asymptotically unique under H_0.

6. Conclusions

As shown in section 2, the derivation of the ML-estimates is quite complicated. As the derivation isn't much simpler in each of the three special cases treated in section 4, we can assume that the generalizations to higher dimension will be of considerable difficulty. However, we could try to replace the ML-principle by some other (preferably easier) conditions and hope to get results which do not depend on normality assumptions. For instance in the case $a = -b$, we note that tr $\hat{\Sigma}_{(i)}$ = tr $S_{(i)}$ $(i = 1,2)$, that is, the total variance is unaffected in both groups. This condition can easily be generalized to higher dimensions; however, it doesn't determine the solution entirely.

In the case $a = 1$, $b = 0$ Pillai*) has suggested the following large sample estimate for the case of two groups: Since $\Sigma^{-1}_{(1)}\Sigma_{(2)}$ is triangular, its diagonal elements are identical with the characteristic values λ_i, that is, $\lambda_1 = (\sigma^{(1)}_{11}\sigma^{(2)}_{22} - c^2\sigma^{(1)}_{11}\sigma^{(2)}_{11})/\Delta$ and $\lambda_2 = (\sigma^{(1)}_{22}\sigma^{(2)}_{11} - c^2\sigma^{(1)}_{11}\sigma^{(2)}_{11})/\Delta$, where $\Delta = \sigma^{(1)}_{11}\ {}^{(1)}_{22} - c^2\sigma^{(1)2}_{11}$. By taking the ratio λ_1/λ_2 we get easily

$$c^2 = \frac{\dfrac{\sigma^{(1)}_{22}}{\sigma^{(1)}_{11}}\lambda_1 - \dfrac{\sigma^{(2)}_{22}}{\sigma^{(2)}_{11}}\lambda_2}{\lambda_1 - \lambda_2} \qquad (6.1)$$

Thus an estimate of $\Sigma_{(i)}$ $(i = 1,2)$ is $\hat{\sigma}^{(i)}_{11} = s^{(i)}_{11}$, $\hat{\sigma}^{(i)}_{22} = s^{(i)}_{22}$, and $\hat{\sigma}^{(i)}_{12} = \hat{c} \cdot \hat{\sigma}^{(i)}_{11}$, where $\hat{c}$ is obtained from (5.1) by using the above estimates and replacing λ_1 and λ_2 by the characteristic values of $S^{-1}_{(1)}S_{(2)}$. This estimate differs from the ML-estimate and has the advantage that $\hat{\sigma}_{22}$ is unbiased even if H_0 is false. Moreover, it does not depend on assumptions about the distribution of the two sample covariance matrices. However, it has two serious shortcomings: First, it is not obvious how to determine the sign of $\hat{c}$, when $s^{(1)}_{12}$ and $s^{(2)}_{12}$ have different signs. Second, since $S^{-1}_{(1)}S_{(2)}$ will in general not be exactly triangular, it is also not obvious, which characteristic root should be labeled as λ_1 and which as λ_2. Nevertheless, Pillai's suggestion gives an alternative idea about the estimation of special patterns avoiding assumptions on underlying distributions.

*) personal communication

7. References

Anderson, T.W. (1958): "An Introduction to Multivariate Statistical Analysis", New York: John Wiley & Sons.

Anderson, T.W. (1970): "Estimation of Covariance Matrices which are Linear Combinations or Whose Inverse are Linear Combinations of Given Matrices", in Essays in Probability and Statistics, ed. R.C. Bose et. al., University of North Carolina Press, 1-24.

Flury, B. (1982): "Analyse von Linear kombinationen mit extremen Varianzenquotienten", Dissertation Universität Bern

Khatri, C.G. and Pillai, K.C.S. (1969): "Distributions of Vectors Corresponding to the Largest Roots of Three Matrices", in Multivariate Analysis, vol. 2, ed. P.C. Krishnaiah, New York: Academic Press.

Morrison, D.F. (1976): "Multivariate Statistical Methods", New York: McGraw-Hill.

Muirhead, R. (1982): "Aspects of Multivariate Statistical Theory", New York: Wiley.

Rao, C.R. (1973): "Linear Statistical Inference and its Applications", second edition, New York: John Wiley.

Rubin, D.B. and Szatrowski, T.H. (1981): "Finding MLE of Patterned Covariance Matrices by the EM Algorithm", Technical Report No. 81-18, Educational Testing Service, Princeton, NJ 08541.

Silvey, S.D. (1975): "Statistical Inference", New York: Halsted Press.

Szatrowski, T.H. (1976): "Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure", Technical Report no. 107, Dept. of Statistics, Stanford University.

Szatrowski, T.H. (1978): "Explicit solutions, one iteration convergence and averaging in the multivariate normal estimation problem for patterned means and covariances", Ann. Stat. Math. 30, Part A, 81-88.

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.