TECHNICAL REPORT NO. 166
June, 1968
MAXIMUM LIKELIHOOD ESTIMATION OF
MULTIVARIATE COVARIANCE COMPONENTS
FOR THE BALANCED ONE-WAY LAYOUT
by
Jerome Klotz\textsuperscript{1} and Joseph Putter\textsuperscript{2}
University of Wisconsin

\textsuperscript{1} Research partially supported by the Air Force under AFOSR Grant No. AF AFOSR-1158-66, and by the Math. Research Center, U.S. Army, Madison, under contract No. DA-31-124-AR-D-462.

\textsuperscript{2} On leave from the Volcani Institute of Agricultural Research, Bet-Dagan, Israel.
MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE
COVARIANCE COMPONENTS FOR THE BALANCED ONE-WAY LAYOUT

by

Jerome Klotz¹ and Joseph Putter²

University of Wisconsin at Madison

1. Introduction. Unbiased estimators of variance and covariance components for the balanced one-way layout have been extensively investigated in the literature. Unfortunately, they possess the unpleasant property of taking on inadmissible values such as negative variances and, more generally, non-positive-semidefinite covariance matrices. This in turn can lead to correlation coefficients that are imaginary or greater than one.

In the univariate case, the maximum likelihood (m.l.) estimators, which are free from these drawbacks, have been derived by Herbach [3] and shown in [5] to have uniformly, and in many cases considerably, smaller mean square errors than the unbiased estimators. Hence it is of interest to consider m.l. estimation in the multivariate case. Searle [7] computed the information matrix for the bivariate case, but did not derive explicit expressions for the estimators.

In this paper, we define (in Section 2) and derive (in Section 3) the maximum likelihood estimators for the general P-variate case. In Section 4

¹ Research partially supported by the Air Force under AFOSR Grant No. AF AFOSR-1158-66, and by the Math. Research Center, U.S. Army, Madison, under contract No. DA-31-124-AR-D-462.

² On leave from the Volcani Institute of Agricultural Research, Bet-Dagan, Israel.
the methods of computation are described, and in Section 5 explicit formulae are given for the bivariate case.

2. Model notation, and extended definition of m.l. estimators. Denote the P-variate observation row vectors by $x_{jk}$. The variance component model corresponding to the balanced one-way layout is

$$ x_{jk} = \mu + \beta_j + \epsilon_{jk} \quad (j = 1, 2, \ldots, J; \quad k = 1, 2, \ldots, K), $$

where $\mu$ is a fixed mean vector, and the $J(K+1)$ random multinormal vectors $\beta_j \sim N(0, \Sigma_b)$ and $\epsilon_{jk} \sim N(0, \Sigma_w)$ are independent. The within-groups covariance matrix $\Sigma_w$ is assumed to be positive definite (p.d.), but the between-groups covariance matrix $\Sigma_b$ may be positive semidefinite (p.s.d.). Denote $x_{..} = \Sigma_k x_{jk}/K$ and $\Gamma = \Sigma_w + K \Sigma_b$. Reduction of the sample space by sufficiency, using the factorization theorem, yields the complete sufficient statistic $(x_{..}, \Sigma_b, \Sigma_w)$ defined by

$$ x_{..} = \Sigma_j \Sigma_k x_{jk}/(JK), \quad \Sigma_b = K \Sigma_j (x_{..} - x_{..})'(x_{..} - x_{..}) $$

and

$$ \Sigma_w = \Sigma_j \Sigma_k (x_{jk} - x_{..})'(x_{jk} - x_{..}). $$

The three statistics $x_{..} \sim N(\mu, \Gamma/(JK))$, $\Sigma_b \sim W(\Gamma, J-1)$, $\Sigma_w \sim W(\Sigma_w, J(K-1))$ are independent. The likelihood is thus given by

$$ L(\mu, \Sigma_b, \Sigma_w | x_{..}, \Sigma_b, \Sigma_w) $$

$$ = a |\Gamma|^{-J/2} |\Sigma_w|^{-J(K-1)/2} \exp \left\{ -\frac{1}{2} [JK(x_{..} - \mu, -\mu) \Gamma^{-1}(x_{..} - \mu)'.

+ \text{tr}(\Gamma^{-1} \Sigma_b + \Sigma_w^{-1} \Sigma_w)] \right\}.$$
where \( \alpha \) is a constant depending only on \( \Sigma_b \) and \( \Sigma_w \).

The likelihood given by (2.2) is meaningfully defined only when \( \Sigma_w \) is p.d. and \( \Sigma - \Sigma_w = K \Sigma_b \) is p.s.d. However, as in the univariate case, the supremum of the likelihood function may not be attained for p.d. values of \( \Sigma_w \). Also, it is not enough to define a m.l. estimator as a limit of a sequence of parameter values for which the likelihood tends to its supremum, because when \( \Sigma_w \) is singular, the supremum is infinite and may be attained by different sequences with different limits. A similar type of difficulty involving infinite suprema of likelihood functions is mentioned by Kiefer and Wolfowitz ([4], p. 905).

To avoid these difficulties, we define a m.l. estimator as follows. Denote a sample point by \( X \), a parameter point by \( \theta \), and the corresponding likelihood function by \( L(\theta | X) \). When \( \sup_{\theta} L(\theta | X) < \infty \), then \( \hat{\theta} = \hat{\theta}(X) \) is a m.l. estimator of \( \theta \) if \( \hat{\theta} = \lim_{n \to \infty} \theta_n \) where \( \lim_{n \to \infty} L(\theta_n | X) = \sup_{\theta} L(\theta | X) \). When \( \sup_{\theta} L(\theta | X) = \infty \), then \( \hat{\theta}(X) \) is a m.l. estimator if \( \hat{\theta}(X) = \lim_{n \to \infty} \hat{\theta}(X_n) \) where \( \sup_{\theta} L(\theta | X_n) < \infty \) and \( \lim_{n \to \infty} X_n = X \). This extended definition gives the solution of Herbach [3] in the univariate case.

We shall denote by \( (H)_+ \) the positive semidefinite part of the matrix \( H \), which is defined by extending the function \( (h)_+ = \max(0, h) \) to a matrix function in the standard way (see, for example, [2], p. 96). We have

\[
(2.3) \quad (H)_+ = \varphi(H)
\]

where \( \varphi \) is any polynomial which satisfies \( \varphi(e_i) = (e_i)_+ \) for all the eigenvalues \( e_i \) of \( H \). If \( M \) is nonsingular, then

\[
(2.4) \quad (M H M^{-1})_+ = \varphi(MH) M^{-1}.
\]
3. Derivation of m.l. estimators. For any fixed p.d. values of $\Sigma_w$ and $\Sigma_b$, the likelihood (2.2) is maximized when $\mu = \mu_b$, and hence the m.l. estimator of $\mu$ is $\hat{\mu} = \mu_b$.

To find $\hat{\Sigma}_b = \hat{\Sigma}_b(\Sigma_b, \Sigma_w)$ and $\hat{\Sigma}_w = \hat{\Sigma}_w(\Sigma_b, \Sigma_w)$, we have to maximize

$$L^*(\mu, \Sigma_w | \Sigma_b, \Sigma_w) = -J \ln |\Sigma| - J(K-1) \ln |\Sigma_w| - \text{tr}(\Sigma^{-1} \Sigma_b + \Sigma_w^{-1} \Sigma_w)$$

w.r.t. $\Sigma_b$ and $\Sigma_w$.

**Lemma 1.** Let $\Sigma_b$ be p.s.d. and $\Sigma_w$ p.d. If $\hat{\Sigma}_w(\Sigma_b, \Sigma_w)$ is p.d., then for any non-singular matrix $Q$,

$$\hat{\Sigma}_w(Q \Sigma_b Q', Q \Sigma_w Q') = Q \hat{\Sigma}_w(\Sigma_b, \Sigma_w) Q'$$

and

$$\hat{\Sigma}_b(Q \Sigma_b Q', Q \Sigma_w Q') = Q \hat{\Sigma}_b(\Sigma_b, \Sigma_w) Q'.$$

**Proof.** Since $L^*(\mu, \Sigma_w | \Sigma_b Q Q', \Sigma_w Q Q') = -JK \ln |QQ'| + L^*(Q^{-1} \Sigma \Sigma^{-1} Q^{-1}, Q^{-1} \Sigma_w Q^{-1} | \Sigma_b, \Sigma_w)$, we have $Q^{-1} \hat{\Sigma}_w(Q \Sigma_b Q', Q \Sigma_w Q') Q^{-1} = \hat{\Sigma}_w(\Sigma_b, \Sigma_w)$, and similarly for $\hat{\Sigma}_b$.

**Lemma 2.** Let $\Sigma_b$ and $\Sigma_w$ be symmetric p.s.d., $\Sigma_t = \Sigma_b + \Sigma_w$, and $A = u \Sigma_b + v \Sigma_w$. Let $\hat{\Sigma}_t$ be any generalized inverse of $\Sigma_t$, and $H$ be any solution of $\Sigma_t H = A$. Then $\Sigma_t (H)_+ = \hat{\Sigma}_t (\Sigma_t^{-} A)_+$ and the common value of these two products does not depend upon the choice of $\Sigma_t^{-}$ and $H$.

**Proof.** The general solution of $\Sigma_t H = A$ is $H = \Sigma_t^{-} A + (\Sigma_t^{-} \Sigma_t^{-} - I_P) Z$, where $I_P$ is the P×P identity matrix and $Z$ is any P×P matrix (see, for example, [6], p.26). The lemma is trivial if $\Sigma_t$ is nonsingular or $\Sigma_t = 0$. If the rank of $\Sigma_t$ is $Q$, $0 < Q < P$, then
\[ \mathcal{L}_t = \mathcal{T} \begin{pmatrix} \mathcal{L}_Q & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix} \mathcal{T}^\prime, \]

where \( \mathcal{T} \) is non-singular. Writing \( \mathcal{L}_b = \mathcal{T} \mathcal{U} \mathcal{T}^\prime \) and \( \mathcal{L}_w = \mathcal{T} \mathcal{V} \mathcal{T}^\prime \), we have

\[ \mathcal{U} + \mathcal{V} = \begin{pmatrix} \mathcal{L}_Q & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix}, \]

and since \( \mathcal{U} \) and \( \mathcal{V} \) are symmetric p.s.d., it follows easily that

\[ \mathcal{U} = \begin{pmatrix} \mathcal{U}_{11} & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix}, \quad \mathcal{V} = \begin{pmatrix} \mathcal{V}_{11} & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix}, \quad \text{and} \quad \mathcal{A} = \mathcal{T} \begin{pmatrix} \mathcal{A}_{11}^* & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix} \mathcal{T}^\prime, \]

where \( \mathcal{U}_{11}, \mathcal{V}_{11} \) are \( Q \times Q \) matrices and \( \mathcal{A}_{11}^* = u \mathcal{U}_{11} + v \mathcal{V}_{11} \). Writing \( \mathcal{L}_t = \mathcal{T}^{-1} \mathcal{W} \mathcal{T}^{-1} \) and using the property \( \mathcal{L}_t \mathcal{S}_t \mathcal{S}_t^\prime = \mathcal{L}_t \) (see, for example, [6], p.24), we obtain

\[ \mathcal{W} = \begin{pmatrix} \mathcal{L}_Q & \mathcal{W}_{12} \\ \mathcal{W}_{21} & \mathcal{W}_{22} \end{pmatrix} \quad \text{and} \quad \mathcal{S}_t \mathcal{A} = \mathcal{T}^{-1} \begin{pmatrix} \mathcal{A}_{11}^* & \mathcal{L} \\ \mathcal{W}_{21} \mathcal{A}_{11}^* & \mathcal{L} \end{pmatrix} \mathcal{T}^\prime, \]

where \( \mathcal{W}_{12}, \mathcal{W}_{21} \) and \( \mathcal{W}_{22} \) are some matrices of the appropriate orders.

Hence

\[ \mathcal{H} = \mathcal{T}^{-1} \begin{pmatrix} \mathcal{A}_{11}^* & \mathcal{L} \\ \mathcal{S}_{21} & \mathcal{S}_{22} \end{pmatrix} \mathcal{T}^\prime = \mathcal{T}^{-1} \mathcal{G} \mathcal{T}^\prime, \]

where \( \mathcal{G}_{21} \) and \( \mathcal{G}_{22} \) are also some matrices of appropriate orders. By (2.3) and (2.4) it follows that \( (\mathcal{H})_+ = \mathcal{T}^{-1} \varphi(\mathcal{G}) \mathcal{T}^\prime \), where \( \varphi \) is a polynomial satisfying \( \varphi(\lambda_i) = (\lambda_i)_+ \) for all eigenvalues \( \lambda_i \) of \( \mathcal{G} \) (i.e., of \( \mathcal{A}_{11}^* \) and \( \mathcal{G}_{22} \)).

Hence

\[ \mathcal{L}_t(\mathcal{H})_+ = \mathcal{T} \begin{pmatrix} \mathcal{L}_Q & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix} \begin{pmatrix} \varphi(\mathcal{A}_{11}^*) & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix} \mathcal{T}^\prime = \mathcal{T} \begin{pmatrix} (\mathcal{A}_{11}^*)_+ & \mathcal{L} \\ \mathcal{L} & \mathcal{L} \end{pmatrix} \mathcal{T}^\prime = \mathcal{L}_t(\mathcal{G}_t \mathcal{A}), \]
because \( \Phi(\hat{\Sigma}^{-1}) = (\Sigma^{-1})_+ \), with the value of \( \hat{\Sigma}^{-1} \) immaterial.

**Theorem.** The maximum likelihood estimators of \( \mu \), \( \Sigma_b \), and \( \Sigma_w \) are given by

\[
\hat{\mu} = \bar{x}.
\]

\[
\hat{\Sigma}_b = (JK)^{-1} \Sigma_t (\Sigma_t^{-1} A)_+
\]

\[
\hat{\Sigma}_w = (JK)^{-1} \Sigma_t [I_p - (\Sigma_t^{-1} A)_+] \]

where \( \Sigma_t^{-1} \) is any generalized inverse of \( \Sigma_t \), and \( A = \Sigma_b - (K-1)^{-1} \Sigma_w \).

**Proof.** As shown above \( \hat{\mu} = \mu \ldots \), and to obtain \( \hat{\Sigma}_b \) and \( \hat{\Sigma}_w \) we have to maximize (3.1). Consider first the case where \( \Sigma_w \) is p.d. Then \( \Sigma_t = \Sigma_b + \Sigma_w \) is also p.d. and there exists a non-singular matrix \( Q \) such that \( \Sigma_t = JKQQ' \) and \( \Sigma_b = JKQ \text{diag}(d_1, d_2, \ldots, d_p)Q' \), where \( d_m (m = 1, 2, \ldots, P) \) are the eigenvalues of \( \Sigma_t^{-1} \Sigma_b \) and satisfy \( 0 \leq d_m < 1 \) since \( \Sigma_w \) is p.d. (see e.g. [1], p. 341). By Lemma 1, the problem is reduced to that of maximizing

\[
L^*(\Sigma_b, \Sigma_w) = (JK \text{diag}(d_1, \ldots, d_p), JK \text{diag}(1-d_1, \ldots, 1-d_p))
\]

\[
= J \ln |\Sigma_b^{-1}| + J(K-1) \ln |\Sigma_w^{-1}| - JK \sum_{m=1}^{P} \gamma_{mm} d_m + \sigma_{w}^{mm} (1-d_m),
\]

where \( (\gamma_{mn}) = \Sigma_b^{-1} \) and \( (\sigma_{w}^{mn}) = \Sigma_w^{-1} \). By Hadamard's inequality (see e.g. [6], p. 45) we have \( |\Sigma_b^{-1}| \geq \prod_{m=1}^{P} \gamma_{mm} \) and \( |\Sigma_w^{-1}| \geq \prod_{m=1}^{P} \sigma_{w}^{mm} \), with either inequality strict unless the corresponding matrix is diagonal. Hence the matrices \( \Sigma_b \) and \( \Sigma_w \) which maximize (3.3) must be diagonal, and it remains to find the values \( \gamma_{mm} \) and \( \sigma_{w}^{mm} \) which maximize

\[
\sum_{m=1}^{P} \left[ J \ln \gamma_{mm} + J(K-1) \ln \sigma_{w}^{mm} - JK \gamma_{mm} (d_m \gamma_{mm} + (1-d_m) \sigma_{w}^{mm}) \right]
\]
subject to the restrictions \( \frac{1}{\gamma_{mm}} = \sigma_{mm} \geq \sigma_{wmm} = \frac{1}{\sigma_{w}} > 0 \). But this is equivalent to solving \( P \) separate univariate problems, and the application of the univariate solution of Herbach [3] yields the value \( \hat{\sigma}_{bmm} = [d_m^{-1}(1-d_m^{-1})_+ \) for the \( m \)-th diagonal element of \( \hat{\sigma}_b \) and the value \( \hat{\sigma}_{wmm} = 1 - \hat{\sigma}_{bmm} \) for the \( m \)-th diagonal element of \( \hat{\sigma}_w \). Thus, putting \( \mathcal{R} = \text{diag} (d_1, d_2, \ldots, d_p) \),

\[
\hat{\mathcal{L}}_b(JK \mathcal{R}, JK(L_p - \mathcal{R})) = [\mathcal{R} - (K^{-1})^{-1}(L_p - \mathcal{R})]_+ ,
\]

(3.4)

\[
\hat{\mathcal{L}}_w(JK \mathcal{R}, JK(L_p - \mathcal{R})) = L_p - \hat{\mathcal{L}}_b(JK \mathcal{R}, JK(L_p - \mathcal{R})) ,
\]

and the value of \( \hat{\mathcal{L}}_w \) in (3.4) is p.d. because \( \hat{\sigma}_{wmm} \geq K(1-d_m)/(K-1) > 0 \).

Hence, by Lemma 1 and by (2.4), we have \( \hat{\mathcal{L}}_b(S_b, S_w) = [\mathcal{R} - (K^{-1})^{-1}(L_p - \mathcal{R})]_+ \mathcal{L}_t' = \mathcal{L}_t'([\mathcal{R} - (K^{-1})^{-1}(L_p - \mathcal{R})]_+ \mathcal{L}_t) = (JK)^{-1} \mathcal{L}_t \mathcal{L}_t^{-1} \), and similarly for \( \hat{\mathcal{L}}_w \), which proves (3.2) for \( \mathcal{L}_w \) p.d.

If \( \mathcal{L}_w \) is singular, then the supremum of (3.1) is infinite and we apply the extended definition of m.l. estimators given in Section 2. Consider a sequence of p.d. \( \mathcal{L}_{wn} \) converging to \( \mathcal{L}_w \) and a sequence of p.s.d. \( \mathcal{L}_{bn} \) converging to \( \mathcal{L}_b \), and put \( \mathcal{L}_{tn} = \mathcal{L}_{bn} + \mathcal{L}_{wn} \), \( \mathcal{A}_n = \mathcal{L}_{bn}^{-1}(K^{-1})^{-1} \mathcal{L}_{wn} \). Any limit point \( \mathcal{H}_t \) of \( \mathcal{L}_{tn}^{-1} \mathcal{A}_n \) must satisfy \( \mathcal{L}_t \mathcal{H}_t = \mathcal{A}_n \), and an application of Lemma 2 completes the proof of the theorem.

REMARKS. (i) When \( \mathcal{A} \) is positive semidefinite, then \( \hat{\mathcal{L}}_b = \mathcal{A}/(JK) \).

(ii) If, following Thompson [8], the restricted maximum likelihood estimators using the maximal invariant are desired, then for \( \hat{\mathcal{L}}_b \), \( \mathcal{A} \) is replaced in (3.2) by \( \mathcal{A} = (J(1 - K^{-1})^{-1} \mathcal{L}_b - (K^{-1})^{-1} \mathcal{L}_w \) and \( \hat{\mathcal{L}}_w = \mathcal{L}_t(JK(1 - K^{-1}))^{-1} \hat{\mathcal{L}}_b \).
4. Computation of $\hat{\Sigma}_b$. In computing $\Sigma_b$, the first step is to calculate $A$, $\xi_t^-$, and $H = \xi_t^+ A$ (for the construction of generalized inverses, see [6], p. 26). In most cases (theoretically with probability one), $\xi_t^-$ will be non-singular and $\xi_t^+ = \xi_t^{-1}$. Next, compute the eigenvalues $e_1, e_2, \ldots, e_p$ of the matrix $H$ and denote by $f_1, f_2, \ldots, f_R$ the distinct values of $\{e_m\}$. Define $\varphi$ to be the unique polynomial of degree $R-1$ that satisfies $\varphi(f_r) = (f_r)_r = \max(0, f_r)$ for $r = 1, 2, \ldots, R$. Then calculate

$$\hat{\Sigma}_b = \xi_t^+ \varphi(H)/(JK).$$  

Any of the well-known representation formulae for $\varphi$ can be used. For example, using the Lagrange interpolation formula we have

$$\hat{\Sigma}_b = (JK)^{-1} \xi_t^+ \sum_{r'} f_{r'} \Pi_{s \neq r'} (H - f_s \xi_p)/(f_{r'} - f_s),$$  

where the sum is taken over all $r'$ for which $f_{r'} > 0$ and each product is taken over all $s = 1, 2, \ldots, R$ different from $r'$. If $f_r \geq 0$ for most $r$, it is more convenient to determine $\hat{\Sigma}_b$ by taking the sum in (4.2) over the negative values of $f_r$ and subtracting it from $A/(JK)$.

5. The bivariate case. Many practical (e.g. genetic) applications of variance components involve the bivariate case, with special emphasis on estimating the "between" and "within" correlation coefficients $\rho_b$ and $\rho_w$ corresponding to $\Sigma_b$ and $\Sigma_w$ respectively. It is therefore of interest to consider this simple case in more detail, with some explicit formulae for the estimators. Two cases have to be distinguished, according to the sign of the determinant $|A|$.

Case I: $|A| \geq 0$. If none of the diagonal elements of $A = (a_{mn})$ are
negative, then \( \hat{\Sigma}_b = \mathcal{A}/(JK) \). If either \( a_{11} \) or \( a_{22} \) is negative, then \( \hat{\Sigma}_b = \mathcal{Q} \).

**Case II:** \( |\mathcal{A}| < 0 \). In this case compute the matrix \( \mathcal{Q}^{-1} \mathcal{A} = \mathcal{H} = (h_{mn}) \) and its eigenvalues \( e_1 = (h_{11} + h_{22} - g)/2 < 0 \) and \( e_2 = (h_{11} + h_{22} + g)/2 > 0 \) where \( g = [(h_{11} - h_{22})^2 + 4h_{12}h_{21}]^{1/2} \). By (4.2), using \( \mathcal{Q}^{1/2} = \mathcal{A} \), we have

\[
\hat{\Sigma}_b = (\mathcal{A} - e_1 \mathcal{Q}^{1/2})e_2/(JKg).
\]

With respect to estimation \( \rho_b \), the situation depends on the diagonal elements of \( \hat{\Sigma}_b \). If both of these elements are positive, then the maximum-likelihood estimator of \( \rho_b \) is the correlation coefficient \( \hat{\rho}_b \) corresponding to \( \hat{\Sigma}_b \); note that \( |\hat{\rho}_b| < 1 \) if \( |\hat{\Sigma}_b| > 0 \) (which will be the case if and only if \( |\mathcal{A}| > 0 \) and both \( a_{11} \) and \( a_{22} \) are positive), and \( |\hat{\rho}_b| = 1 \) if \( |\hat{\Sigma}_b| = 0 \). If one or both of the diagonal elements of \( \hat{\Sigma}_b \) vanish, then no meaningful maximum-likelihood estimator of \( \rho_b \) seems to exist. The estimation of \( \rho_w \) depends similarly on the diagonal elements of \( \hat{\Sigma}_w \).

**Acknowledgements** The authors wish to thank G. Shook for bringing this problem to their attention and B. Harris for a helpful conversation.
REFERENCES


