The Annals of Mathematical Statistics
1969, Vol. 40, No. 3, 1100–1105

MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE COVARIANCE COMPONENTS FOR THE BALANCED ONE-WAY LAYOUT

By Jerome Klotz[1] and Joseph Putter[2]

University of Wisconsin

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 (ml) 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 ml 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 geneal P-variate case. In Section 4 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 ml estimators. Denote the P-variate observation row vectors by $\mathbf{x}_{jk}$. The variance component model corresponding to the balanced one-way layout is

$$\mathbf{x}_{jk} = \boldsymbol{\mu} + \mathbf{b}_j + \mathbf{w}_{jk} \qquad (j = 1, 2, \cdots, J; k = 1, 2, \cdots, K), \tag{2.1}$$

where $\boldsymbol{\mu}$ is a fixed mean vector, and the $J(K + 1)$ random multinormal vectors $\mathbf{b}_j : N(\mathbf{0}, \boldsymbol{\Sigma}_b)$ and $\mathbf{w}_{jk} : N(\mathbf{0}, \boldsymbol{\Sigma}_w)$ are independent. The within-groups covariance matrix $\boldsymbol{\Sigma}_w$ is assumed to be positive definite (pd), but the between-groups covariance matrix $\boldsymbol{\Sigma}_b$ may be positive semidefinite (psd). Denote $\mathbf{x}_{j\cdot} = \Sigma_k \mathbf{x}_{jk}/K$ and $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}_w + K\boldsymbol{\Sigma}_b$. Reduction of the sample space by sufficiency, using the factorization theorem, yields the complete sufficient statistic $(\mathbf{x}_{\cdot\cdot}, \mathbf{S}_b, \mathbf{S}_w)$ defined by

$$\mathbf{x}_{\cdot\cdot} = \textstyle\sum_j \sum_k \mathbf{x}_{jk}/(JK),$$
$$\mathbf{S}_b = K \textstyle\sum_j (\mathbf{x}_{j\cdot} - \mathbf{x}_{\cdot\cdot})'(\mathbf{x}_{j\cdot} - \mathbf{x}_{\cdot\cdot}) \qquad \text{and}$$
$$\mathbf{S}_w = \textstyle\sum_j \sum_k (\mathbf{x}_{jk} - \mathbf{x}_{j\cdot})'(\mathbf{x}_{jk} - \mathbf{x}_{j\cdot}).$$

Received 26 June 1968.

[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, Wisconsin, under Contract No. DA-31-124-ARO-D-462.

[2] On leave from the Volcani Institute of Agricultural Research, Bet-Dagan, Israel.

The three statistics $\mathbf{x}_{..}:N(\boldsymbol{\mu}, \boldsymbol{\Gamma}/(JK))$, $S_b:W(\boldsymbol{\Gamma}, J - 1)$, $\mathbf{S}_w:W(\boldsymbol{\Sigma}_w, J(K - 1))$ are independent where $W(\mathbf{M}, n)$ denotes a random matrix following the Wishart distribution with matrix parameter $\mathbf{M}$ and degrees of freedom n (see for example [1] p. 158). The likelihood is thus given by

$$\begin{aligned} &L(\boldsymbol{\mu}, \boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_w \mid \mathbf{x}_{..}, \mathbf{S}_b, \mathbf{S}_w) \\ (2.2) \qquad &= a\,|\boldsymbol{\Gamma}|^{-J/2}\,|\boldsymbol{\Sigma}_w|^{-J(K-1)/2} \exp\{-\tfrac{1}{2}[JK(\mathbf{x}_{..} - \boldsymbol{\mu})\boldsymbol{\Gamma}^{-1}(\mathbf{x}_{..} - \boldsymbol{\mu})' \\ &\qquad + \operatorname{tr}(\boldsymbol{\Gamma}^{-1}\mathbf{S}_b + \boldsymbol{\Sigma}_w^{-1}\mathbf{S}_w)]\} \end{aligned}$$

where a is a constant depending only on $\mathbf{S}_b$ and $\mathbf{S}_w$.

The likelihood given by (2.2) is meaningfully defined only when $\boldsymbol{\Sigma}_w$ is pd and $\boldsymbol{\Gamma} - \boldsymbol{\Sigma}_w = K\boldsymbol{\Sigma}_b$ is psd. However, as in the univariate case, the supremum of the likelihood function may not be attained for pd values of $\boldsymbol{\Sigma}_w$. Also, it is not enough to define a ml estimator as a limit of a sequence of parameter values for which the likelihood tends to its supremum, because when $\mathbf{S}_w$ 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 ml estimator as follows. Denote a sample point by X, a parameter point by θ, and the corresponding likelihood function by $L(\theta \mid X)$. When $\sup_\theta L(\theta \mid X) < \infty$, then $\hat{\theta} = \hat{\theta}(X)$ is a ml esimator of θ if $\hat{\theta} = \lim_n \theta_n$ where $\lim_n L(\theta_n \mid X) = \sup_\theta L(\theta \mid X)$. When $\sup_\theta L(\theta \mid X) = \infty$, then $\hat{\theta}(X)$ is a ml estimator if $\hat{\theta}(X) = \lim_n \hat{\theta}(X_n)$ where $\sup_\theta L(\theta \mid X_n) < \infty$ and $\lim_n X_n = X$. (For the purpose of the definition the limits are considered in the pointwise sense although weaker types of convergence may give equivalent results in some cases.) This extended definition gives the solution of Herbach [3] in the univariate case.

We shall denote by $(\mathbf{H})_+$ the positive semidefinite part of the matrix $\mathbf{H}$, 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

$$(2.3) \qquad (\mathbf{H})_+ = \varphi(\mathbf{H})$$

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

$$(2.4) \qquad (\mathbf{M}\mathbf{H}\mathbf{M}^{-1})_+ = \mathbf{M}(\mathbf{H})_+\mathbf{M}^{-1}.$$

3. Derivation of ml **estimators.** For any fixed pd values of $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Gamma}$, the likelihood (2.2) is maximized when $\boldsymbol{\mu} = \mathbf{x}_{..}$, and hence the ml estimator of $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}} = \mathbf{x}_{..}$. To find $\hat{\boldsymbol{\Sigma}}_b = \hat{\boldsymbol{\Sigma}}_b(\mathbf{S}_b, \mathbf{S}_w)$ and $\hat{\boldsymbol{\Sigma}}_w = \hat{\boldsymbol{\Sigma}}_w(\mathbf{S}_b, \mathbf{S}_w)$ we have to maximize

$$\begin{aligned} (3.1) \quad L^*(\boldsymbol{\Gamma}, \boldsymbol{\Sigma}_w \mid \mathbf{S}_b, \mathbf{S}_w) = &-J \ln|\boldsymbol{\Gamma}| - J(K - 1)\ln|\boldsymbol{\Sigma}_w| \\ &- \operatorname{tr}(\boldsymbol{\Gamma}^{-1}\mathbf{S}_b + \boldsymbol{\Sigma}_w^{-1}\mathbf{S}_w), \end{aligned}$$

wrt $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_w$.

LEMMA 1. *Let* $\mathbf{S}_b$ *be* psd *and* $\mathbf{S}_w$ pd. *If* $\hat{\mathbf{\Sigma}}_w(\mathbf{S}_b, \mathbf{S}_w)$ *is* pd *then for any non-singular matrix* $\mathbf{C}$,

$$\hat{\mathbf{\Sigma}}_w(\mathbf{C}\mathbf{S}_b\mathbf{C}', \mathbf{C}\mathbf{S}_w\mathbf{C}') = \mathbf{C}\hat{\mathbf{\Sigma}}_w(\mathbf{S}_b, \mathbf{S}_w)\mathbf{C}'$$

and

$$\hat{\mathbf{\Sigma}}_b(\mathbf{C}\mathbf{S}_b\mathbf{C}', \mathbf{C}\mathbf{S}_w\mathbf{C}') = \mathbf{C}\hat{\mathbf{\Sigma}}_b(\mathbf{S}_b, \mathbf{S}_w)\mathbf{C}'.$$

PROOF. Since $L^*(\mathbf{\Gamma}, \mathbf{\Sigma}_w \mid \mathbf{C}\mathbf{S}_b\mathbf{C}', \mathbf{C}\mathbf{S}_w\mathbf{C}') = -JK \ln|\mathbf{C}\mathbf{C}'| + L^*(\mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}'^{-1}, \mathbf{C}^{-1}\mathbf{\Sigma}_w\mathbf{C}'^{-1} \mid \mathbf{S}_b, \mathbf{S}_w)$, we have $\mathbf{C}^{-1}\hat{\mathbf{\Sigma}}_w(\mathbf{C}\mathbf{S}_b\mathbf{C}', \mathbf{C}\mathbf{S}_w\mathbf{C}')\mathbf{C}'^{-1} = \hat{\mathbf{\Sigma}}_w(\mathbf{S}_b, \mathbf{S}_w)$, and similarly for $\mathbf{\Gamma}$ and therefore for $\hat{\mathbf{\Sigma}}_b$.

LEMMA 2. *Let* $\mathbf{S}_b$ *and* $\mathbf{S}_w$ *be symmetric* psd, $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w$, *and* $\mathbf{A} = u\mathbf{S}_b + v\mathbf{S}_w$. *Let* $\mathbf{S}_t^-$ *be any generalized inverse of* $\mathbf{S}_t$, *and* $\mathbf{H}$ *be any solution of* $\mathbf{S}_t\mathbf{H} = \mathbf{A}$. *Then* $\mathbf{S}_t(\mathbf{H})_+ = \mathbf{S}_t(\mathbf{S}_t^-\mathbf{A})_+$ *and the common value of these two products does not depend upon the choice of* $\mathbf{S}_t^-$ *and* $\mathbf{H}$.

PROOF. The general solution of $\mathbf{S}_t\mathbf{H} = \mathbf{A}$ is $\mathbf{H} = \mathbf{S}_t^-\mathbf{A} + (\mathbf{S}_t^-\mathbf{S}_t - \mathbf{I}_P)\mathbf{Z}$ where $\mathbf{I}_P$ is the $P \times P$ identity matrix and $\mathbf{Z}$ is any $P \times P$ matrix (see, for example, [6], p. 26). The lemma is trivial if $\mathbf{S}_t$ is nonsingular or $\mathbf{S}_t = \mathbf{0}$. If the rank of $\mathbf{S}_t$ is Q, $0 < Q < P$, then

$$\mathbf{S}_t = \mathbf{T}\begin{pmatrix}\mathbf{I}_Q & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{T}',$$

where $\mathbf{T}$ is non-singular. Writing $\mathbf{S}_b = \mathbf{T}\mathbf{U}\mathbf{T}'$ and $\mathbf{S}_w = \mathbf{T}\mathbf{V}\mathbf{T}'$, we have

$$\mathbf{U} + \mathbf{V} = \begin{pmatrix}\mathbf{I}_Q & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix},$$

and since $\mathbf{U}$ and $\mathbf{V}$ are symmetric psd, it follows easily that

$$\mathbf{U} = \begin{pmatrix}\mathbf{U}_{11} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix}\mathbf{V}_{11} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}, \qquad \text{and} \quad \mathbf{A} = \mathbf{T}\begin{pmatrix}\mathbf{A}_{11}^* & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{T}',$$

where $\mathbf{U}_{11}$, $\mathbf{V}_{11}$ are $Q \times Q$ matrices and $\mathbf{A}_{11}^* = u\mathbf{U}_{11} + v\mathbf{V}_{11}$. Writing $\mathbf{S}_t^- = \mathbf{T}'^{-1}\mathbf{W}\mathbf{T}^{-1}$ and using the property $\mathbf{S}_t\mathbf{S}_t^-\mathbf{S}_t = \mathbf{S}_t$ (see, for example, [6], p. 24), we obtain

$$\mathbf{W} = \begin{pmatrix}\mathbf{I}_Q & \mathbf{W}_{12}\\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{pmatrix} \qquad \text{and} \quad \mathbf{S}_t^-\mathbf{A} = \mathbf{T}'^{-1}\begin{pmatrix}\mathbf{A}_{11}^* & \mathbf{0}\\ \mathbf{W}_{21}\mathbf{A}_{11}^* & \mathbf{0}\end{pmatrix}\mathbf{T}',$$

where $\mathbf{W}_{12}$, $\mathbf{W}_{21}$ and $\mathbf{W}_{22}$ are some matrices of the appropriate orders. Hence

$$\mathbf{H} = \mathbf{T}'^{-1}\begin{pmatrix}\mathbf{A}_{11}^* & \mathbf{0}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{pmatrix}\mathbf{T}' = \mathbf{T}'^{-1}\mathbf{G}\mathbf{T}', \qquad \text{say},$$

where $\mathbf{G}_{21}$ and $\mathbf{G}_{22}$ are also some matrices of appropriate orders. By (2.3) and (2.4) it follows that $(\mathbf{H})_+ = \mathbf{T}'^{-1}\varphi(\mathbf{G})\mathbf{T}'$, where φ is a polynomial satisfying

$\varphi(\lambda_i) = (\lambda_i)_+$ for all eigenvalues λ_i of $\mathbf{G}$ (i.e., of $\mathbf{A}_{11}^*$ and $\mathbf{G}_{22}$). Hence

$$\mathbf{S}_t(\mathbf{H})_+ = \mathbf{T}\begin{pmatrix} \mathbf{I}_Q & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \varphi(\mathbf{A}_{11}^*) & 0 \\ \mathbf{G}_{21}^* & \varphi(\mathbf{G}_{22}) \end{pmatrix}\mathbf{T}' = \mathbf{T}\begin{pmatrix} (\mathbf{A}_{11}^*)_+ & 0 \\ 0 & 0 \end{pmatrix}\mathbf{T}' = \mathbf{S}_t(\mathbf{S}_t^-\mathbf{A})_+,$$

because $\varphi(\mathbf{A}_{11}^*) = (\mathbf{A}_{11}^*)_+$, with the value of $\mathbf{G}_{21}^*$ immaterial.

THEOREM. *The maximum likelihood estimators of* $\mathbf{\mu}$, $\mathbf{\Sigma}_b$, *and* $\mathbf{\Sigma}_w$ *are given by*

$$\hat{\mathbf{\mu}} = \mathbf{x}_{..}$$

$$\hat{\mathbf{\Sigma}}_b = (JK)^{-1}\mathbf{S}_t(\mathbf{S}_t^-\mathbf{A})_+ \tag{3.2}$$

$$\hat{\mathbf{\Sigma}}_w = (JK)^{-1}\mathbf{S}_t[\mathbf{I}_p - (\mathbf{S}_t^-\mathbf{A})_+]$$

where $\mathbf{S}_t^-$ *is any generalized inverse of* $\mathbf{S}_t$, *and* $\mathbf{A} = \mathbf{S}_b - (K-1)^{-1}\mathbf{S}_w$.

PROOF. As shown above $\hat{\mathbf{\mu}} = \mathbf{x}_{..}$, and to obtain $\hat{\mathbf{\Sigma}}_b$ and $\hat{\mathbf{\Sigma}}_w$ we have to maximize (3.1). Consider first the case where $\mathbf{S}_w$ is pd. Then $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w$ is also pd and there exists a non-singular matrix $\mathbf{C}$ such that $\mathbf{S}_t = JK\mathbf{CC}'$ and $\mathbf{S}_b = JK\mathbf{C}\operatorname{diag}(d_1, d_2, \cdots, d_p)\mathbf{C}'$, where $d_m (m = 1, 2, \cdots, P)$ are the eigenvalues of $\mathbf{S}_t^{-1}\mathbf{S}_b$ and satisfy $0 \leqq d_m < 1$ since $\mathbf{S}_w$ is pd (see e.g. [1], p. 341). By Lemma 1, the problem is reduced to that of maximizing

$$L^*(\mathbf{\Sigma}_b, \mathbf{\Sigma}_w \mid JK\operatorname{diag}(d_1, \cdots, d_p), JK\operatorname{diag}(1-d_1, \cdots, 1-d_p)) \tag{3.3}$$
$$= J\ln|\mathbf{\Gamma}^{-1}| + J(K-1)\ln|\mathbf{\Sigma}_w^{-1}| - JK\textstyle\sum_{m=1}^{P}[\gamma^{mm}d_m + \sigma_w^{mm}(1-d_m)],$$

where $(\gamma^{mn}) = \mathbf{\Gamma}^{-1}$ and $(\sigma_w^{mn}) = \mathbf{\Sigma}_w^{-1}$. By Hadamard's inequality (see e.g. [6]' p. 45) we have $|\mathbf{\Gamma}^{-1}| \leqq \prod_{m=1}^{P}\gamma^{mm}$ and $|\mathbf{\Sigma}_w^{-1}| \leqq \prod_{m=1}^{P}\sigma_w^{mm}$, with either inequality strict unless the corresponding matrix is diagonal. Hence the matrices $\mathbf{\Sigma}_b$ and $\mathbf{\Sigma}_w$ which maximize (3.3) must be diagonal, and it remains to find the values γ^{mm} and σ_w^{mm} which maximize

$$\textstyle\sum_{m=1}^{P}[J\ln\gamma^{mm} + J(K-1)\ln\sigma_w^{mm} - JK(d_m\gamma^{mm} + (1-d_m)\sigma_w^{mm})]$$

subject to the restrictions $1/\gamma^{mm} = \gamma_{mm} \geqq \sigma_{wmm} = 1/\sigma_w^{mm} > 0$. But this is equivalent to solving P separate univariate problems, and the application of the univariate solution of Herbach [3] yields the value $\hat{\sigma}_{bmm} = [d_m - (K-1)^{-1}(1-d_m)]_+$ for the mth diagonal element of $\hat{\mathbf{\Sigma}}_b$ and the value $\hat{\sigma}_{wmm} = 1 - \hat{\sigma}_{bmm}$ for the mth diagonal element of $\hat{\mathbf{\Sigma}}_w$. Thus, putting $\mathbf{D} = \operatorname{diag}(d_1, d_2, \cdots, d_p)$,

$$\hat{\mathbf{\Sigma}}_b(JK\mathbf{D}, JK(\mathbf{I}_P - \mathbf{D})) = [\mathbf{D} - (K-1)^{-1}(\mathbf{I}_P - \mathbf{D})]_+, \tag{3.4}$$
$$\hat{\mathbf{\Sigma}}_w(JK\mathbf{D}, JK(\mathbf{I}_P - \mathbf{D})) = \mathbf{I}_P - \hat{\mathbf{\Sigma}}_b(JK\mathbf{D}, JK(\mathbf{I}_P - \mathbf{D})),$$

and the value of $\hat{\mathbf{\Sigma}}_w$ in (3.4) is pd because $\hat{\sigma}_{wmm} \geqq K(1-d_m)/(K-1) > 0$. Hence, by Lemma 1 and by (2.4), we have

$$\begin{aligned}\hat{\mathbf{\Sigma}}_b(\mathbf{S}_b, \mathbf{S}_w) &= \mathbf{C}[\mathbf{D} - (K-1)^{-1}(\mathbf{I}_P - \mathbf{D})]_+\mathbf{C}' \\ &= \mathbf{CC}'(\mathbf{C}'^{-1}[\mathbf{D} - (K-1)^{-1}(\mathbf{I}_P - \mathbf{D})]\mathbf{C}')_+ \\ &= (JK)^{-1}\mathbf{S}_t(\mathbf{S}_t^{-1}\mathbf{A})_+,\end{aligned}$$

and similarly for $\hat{\mathbf{\Sigma}}_w$, which proves (3.2) for $\mathbf{S}_w$ pd.

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

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

(ii) If, following Thompson [8], the restricted maximum likelihood estimators using the maximal invariant are desired, then for $\hat{\mathbf{\Sigma}}_b$, $\mathbf{A}$ is replaced in (3.2) by $\mathbf{A} = J(J-1)^{-1}\mathbf{S}_b - (k-1)^{-1}\mathbf{S}_w$ and $\hat{\mathbf{\Sigma}}_w = \mathbf{S}_t(JK-1)^{-1} - \hat{\mathbf{\Sigma}}_b$.

4. Computation of $\hat{\mathbf{\Sigma}}_b$. In computing $\hat{\mathbf{\Sigma}}_b$, the first step is to calculate $\mathbf{A}$, $\mathbf{S}_t^-$, and $\mathbf{H} = \mathbf{S}_t^-\mathbf{A}$ (for the construction of generalized inverses, see [6], p. 26). In most cases (theoretically with probability one), $\mathbf{S}_t$ will be non-singular and $\mathbf{S}_t^- = \mathbf{S}_t^{-1}$. Next, compute the eigenvalues $e_1, e_2, \cdots, e_P$ of the matrix $\mathbf{H}$ and denote by $f_1, f_2, \cdots, f_R$ the distinct values of $\{e_m\}$. Define φ to be the unique polynomial of degree $R-1$ that satisfies $\varphi(f_r) = (f_r)_+ = \max(0, f_r)$ for $r = 1, 2, \cdots, R$. Then calculate

$$\hat{\mathbf{\Sigma}}_b = \mathbf{S}_t\varphi(\mathbf{H})/(JK). \tag{4.1}$$

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

$$\hat{\mathbf{\Sigma}}_b = (JK)^{-1}\mathbf{S}_t\textstyle\sum_{r'}f_{r'}\prod_{s\neq r'}(\mathbf{H} - f_s\mathbf{I}_P)/(f_{r'} - f_s), \tag{4.2}$$

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

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 ρ_b and ρ_w corresponding to $\mathbf{\Sigma}_b$ and $\mathbf{\Sigma}_w$ 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 $|\mathbf{A}|$.

Case 1. $|\mathbf{A}| \geqq 0$. If none of the diagonal elements of $\mathbf{A} = (a_{mn})$ are negative, then $\hat{\mathbf{\Sigma}}_b = \mathbf{A}/(JK)$. If either a_{11} or a_{22} is negative, then $\hat{\mathbf{\Sigma}}_b = \mathbf{0}$.

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

$$\hat{\mathbf{\Sigma}}_b = (\mathbf{A} - e_1\mathbf{S}_t)e_2/(JKg).$$

With respect to estimating ρ_b, the situation depends on the diagonal elements of $\hat{\mathbf{\Sigma}}_b$. If both of these elements are positive, then the maximum-likelihood estimator of ρ_b is the correlation coefficient $\hat{\rho}_b$ corresponding to $\hat{\mathbf{\Sigma}}_b$; note that $|\hat{\rho}_b| < 1$

if $|\hat{\mathbf{\Sigma}}_b| > 0$ (which will be the case if and only if $|\mathbf{A}| > 0$ and both a_{11} and a_{22} are positive), and $|\hat{\rho}_b| = 1$ if $|\hat{\mathbf{\Sigma}}_b| = 0$. If one or both of the diagonal elements of $\hat{\mathbf{\Sigma}}_b$ vanish, then no meaningful maximum-likelihood estimator of ρ_b seems to exist. The estimation of ρ_w depends similarly on the diagonal elements of $\hat{\mathbf{\Sigma}}_w$.

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

REFERENCES

[1] ANDERSON, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] GANTMACHER, F. R. (1959). *Matrix Theory* Vol. I. Chelsea, New York.

[3] HERBACH, L. H. (1959). Properties of model II-type analysis of variance tests. *Ann. Math. Statist.* **30** 939–959.

[4] KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

[5] KLOTZ, J. H., MILTON, R. C., and ZACKS, S. (1968). Mean square efficient estimators of variance components. Technical Report #136, Department of Statistics, Univ. of Wisconsin.

[6] RAO, C. R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

[7] SEARLE, S. R. (1956). Matrix methods in components of variance and covariance analysis. *Ann. Math. Statist.* **27** 737–748.

[8] THOMPSON, W. A. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33** 273–289.