K.R. KOCH
Institut für Theoretische Geodäsie der Universität Bonn,
Nussallee 17, 5300 Bonn 1
Federal Republic of Germany

MAXIMUM LIKELIHOOD ESTIMATE OF VARIANCE COMPONENTS

Ideas by A.J. Pope

(In memory of Allen J. Pope, 11.10.1939–29.08.1985)

Summary

Using the orthogonal complement likehood function, an iterative procedure for the maximum likelihood estimates of the variance and covariance components is derived. It is shown that these estimates are identical with the reproducing estimates of the locally best invariant quadratic unbiased estimation of variance and covariance components. Successive approximations of the maximum likelihood estimates are given in addition.

1. Introduction

In September, 1983, when I worked as a Senior Visiting Scientist at the National Geodetic Survey in Rockville, Maryland, Allen Pope explained to me his ideas about the maximum likelihood estimates of variance and covariance components. I was interested in this topic because I saw an application for a problem I wanted to solve, and we agreed to cooperate on further investigations. But no progress was made on this subject during the last two years.

As usual, Allen Pope made notes when he explained his ideas. From a copy of his notes of September, 1983, the following derivation is taken. It is one example of his many brilliant ideas.

2. Model Space Likelihood and Orthogonal Complement Likelihood

We start from the model

$$\underline{X}\underline{\beta} = E(\underline{y}) = \underline{\mu}_y \quad \text{with} \quad D(\underline{y}) = \underline{\Sigma} = \underline{\Sigma}(\underline{\sigma}) \tag{2.1}$$

where $\underline{X}$ is the $n \times u$ matrix of known coefficients with rank $\underline{X} = u$, $\underline{\beta}$ the $u \times 1$ vector of unknown parameters, $\underline{y}$ the $n \times 1$ random vector of observations, $E(\underline{y}) = \underline{\mu}_y$ the $n \times 1$ vector of expected values of the observations and $D(\underline{y}) = \underline{\Sigma}$ the $n \times n$ positive definite covariance matrix of the observations. This matrix is assumed to be a function of the $k \times 1$ vector $\underline{\sigma}$ of unknown parameters, the so-called variance and

Bull. Géod. 60 (1986) pp. 329–338.

covariance components, hence $\underline{\Sigma} = \underline{\Sigma}(\underline{\sigma})$. Therefore, (2.1) is the Gauss-Markoff model with unknown variance and covariance components.

If the observations $\underline{y}$ are assumed to be normally distributed, the likelihood function $L(\underline{y};\underline{\beta},\underline{\sigma})$ of the observations $\underline{y}$ with the unknown parameters $\underline{\beta}$ and $\underline{\sigma}$ is given by

$$L(\underline{y};\underline{\beta},\underline{\sigma}) = \frac{1}{(2\pi)^{n/2}(\det \underline{\Sigma})^{1/2}} \, exp\left\{-\frac{1}{2}(\underline{y}-\underline{\mu}_y)'\,\underline{\Sigma}^{-1}(\underline{y}-\underline{\mu}_y)\right\} \tag{2.2}$$

The maximum likelihood estimates of $\underline{\beta}$ and $\underline{\sigma}$ are determined such that $L(\underline{y};\underline{\beta},\underline{\sigma})$ assumes a maximum; see, for instance, (Harville, 1977; Kubik, 1970).

The vector $\underline{y}$ of observations is now transformed by the $n \times n$ matrix $\underline{P}$ into the $n \times 1$ vector $\bar{\underline{y}}$ with

$$\bar{\underline{y}} = \underline{P}\underline{y}\,, \quad \underline{\mu}_{\bar{y}} = \underline{P}\underline{\mu}_y \quad \text{and} \quad \underline{D}(\bar{\underline{y}}) = \underline{P}\,\underline{\Sigma}\,\underline{P}' \tag{2.3}$$

so that we obtain instead of (2.2) the likelihood function of $\bar{\underline{y}}$

$$L(\bar{\underline{y}};\underline{\beta},\underline{\sigma}) = \frac{1}{(2\pi)^{n/2}(\det \underline{P}\,\underline{\Sigma}\,\underline{P}')^{1/2}} \, exp\left\{-\frac{1}{2}(\bar{\underline{y}}-\underline{\mu}_{\bar{y}})'(\underline{P}\,\underline{\Sigma}\,\underline{P}')^{-1}(\bar{\underline{y}}-\underline{\mu}_{\bar{y}})\right\} \tag{2.4}$$

The matrix $\underline{P}$ is defined by

$$\underline{P} = \begin{vmatrix} \underline{B} \\ \underline{X}'\,\underline{\Sigma}^{-1} \end{vmatrix} \tag{2.5}$$

with

$$\underline{B}\underline{X} = \underline{0} \tag{2.6}$$

where $\underline{B}$ is a $(n-u) \times n$ matrix with rank $\underline{B} = n-u$. The matrix $\underline{P}\,\underline{\Sigma}\,\underline{P}'$ is positive definite, since $\underline{P}$ is regular, which can be shown by

$$\underline{P}\,\underline{\Sigma}\,\underline{P}' = \begin{vmatrix} \underline{B}\,\underline{\Sigma}\,\underline{B}' & \underline{0} \\ \underline{0} & \underline{X}'\,\underline{\Sigma}^{-1}\underline{X} \end{vmatrix} \tag{2.7}$$

Taking the determinants of both sides leads to

$$(\det \underline{P})^2 = \det \underline{\Sigma}^{-1} \det \underline{B}\,\underline{\Sigma}\,\underline{B}' \det \underline{X}'\underline{\Sigma}^{-1}\underline{X}$$

The determinants on the right-hand side are positive, so that $\det \underline{P} \neq 0$.

Substituting (2.5) in (2.3) we get with (2.1) and (2.6)

$$\bar{\underline{y}} = \begin{vmatrix} \underline{B}\underline{y} \\ \underline{X}'\,\underline{\Sigma}^{-1}\,\underline{y} \end{vmatrix} \quad \text{and} \quad \underline{\mu}_{\bar{y}} = \begin{vmatrix} \underline{0} \\ \underline{X}'\,\underline{\Sigma}^{-1}\,\underline{X}\underline{\beta} \end{vmatrix} \tag{2.8}$$

and instead of (2.4)

$$L(\bar{y};\beta,\sigma) = \frac{1}{(2\pi)^{n/2} (\det B\Sigma B' \det X'\Sigma^{-1} X)^{1/2}}$$

$$exp\left\{-\frac{1}{2}\begin{vmatrix} By \\ X'\Sigma^{-1} y - X'\Sigma^{-1} X\beta \end{vmatrix}' \begin{vmatrix} (B\Sigma B')^{-1} & 0 \\ 0 & (X'\Sigma^{-1}X)^{-1} \end{vmatrix} \begin{vmatrix} By \\ X'\Sigma^{-1} y - X'\Sigma^{-1} X\beta \end{vmatrix}\right\}$$

or with $n = (n-u) + u$

$$L(\bar{y};\beta,\sigma) = L_1 (y;\sigma) L_2 (y;\beta,\sigma) \tag{2.9}$$

with

$$L_1 (y;\sigma) = \frac{1}{(2\pi)^{(n-u)/2} (\det B\Sigma B')^{1/2}} exp\left\{-\frac{1}{2}(By)'(B\Sigma B')^{-1}(By)\right\}$$

and

$$L_2 (y;\beta,\sigma) = \frac{1}{(2\pi)^{u/2} (\det X'\Sigma^{-1} X)^{1/2}} exp\left\{-\frac{1}{2}(X'\Sigma^{-1} y - X'\Sigma^{-1} X\beta)'(X'\Sigma^{-1} X)^{-} (X'\Sigma^{-1} y - X'\Sigma^{-1} X\beta)\right\}$$

The likelihood function (2.9) of the transformed observations $\bar{y}$ is obtained as the product of the two likelihood functions L_1 and L_2. The first one, L_1, does not depend on the unknown parameters β, but only on the parameters σ because of $\Sigma = \Sigma(\sigma)$. Hence, this likelihood function will be used to estimate the unknown variance and covariance components σ. The second likelihood function, L_2, in (2.9) contains in addition to σ the unknown parameters β. If we assume for a moment σ as known so that the matrix Σ is given, the likelihood function L_2 assumes a maximum if

$$(X'\Sigma^{-1} y - X'\Sigma^{-1} X\beta)'(X'\Sigma^{-1} X)^{-1}(X'\Sigma^{-1} y - X'\Sigma^{-1} X\beta) = 0$$

which is fulfiled with

$$X'\Sigma^{-1} X\hat{\beta} = X'\Sigma^{-1} y \tag{2.10}$$

These are the well-known normal equations of the maximum likelihood estimate in case of normally distributed observations, of the method of least squares and of the best linear unbiased estimate, applied to the unknown parameters β of the Gauss-Markoff model.

As shown by (2.10), the second likelihood function L_2 in (2.9) leads to the estimate $\hat{\beta}$ in the parameter space for β, formed by the column space $R(X)$ of the model (2.1). This likelihood function will therefore be called the model space likelihood.

With (2.6) and rank $B = n - u$, it follows that the null space $N(\underline{X}')$ of $\underline{X}'$ is equal to the column space $R(\underline{B}')$ of $\underline{B}'$. Hence $N(\underline{X}') = R(\underline{B}')$, so that with $N(\underline{X}') = R(\underline{X})^{\perp}$ where $R(\underline{X})^{\perp}$ is the orthogonal complement of the column space $R(\underline{X})$, we get

$$R(\underline{X})^{\perp} = R(\underline{B}') \tag{2.11}$$

The first likelihood function L_1 in (2.9) is therefore called the orthogonal complement likelihood, and it will be used to estimate $\underline{\sigma}$ as already mentioned.

This approach is equivalent to the restricted maximum likelihood estimate, which takes care of the loss in degrees of freedom resulting from the estimates of the unknown parameters $\underline{\beta}$. The variance and covariance components $\underline{\sigma}$ are therefore determined by the so-called error contrasts, which fulfil the condition $E(\underline{a}'\underline{y}) = 0$, so that $\underline{a}'\underline{X} = \underline{0}$ with $\underline{a}$ being a given vector. Hence, the transformed observations $\underline{B}\underline{y}$ in (2.9) are error contrasts. A set of $n - u$ linearly independent error constrasts which have been chosen for the restricted maximum likelihood estimate are the ones obtained by the rank factorization of the projection matrix $\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$ (Harville, 1977, p. 325, Schaffrin, 1983, p. 85).

A similar representation of the likelihood function by two factors, one for estimating the unknown parameters $\underline{\beta}$, the other for estimating the variance and covariance components $\underline{\sigma}$ can be also found in (Kubik, 1970). However, the likelihood function (2.9) of the transformed observations leads to simpler formulas.

3. Estimate of the Variance and Covariance Components

In order to determine the extreme values of the orthogonal complement likelihood function L_1 in (2.9) we take its natural logarithm and obtain

$$\ln L_1(\underline{y};\underline{\sigma}) = -\frac{n-u}{2}\ln(2\pi) - \frac{1}{2}\ln\det \underline{B}\underline{\Sigma}\underline{B}' - \frac{1}{2}(\underline{B}\underline{y})'(\underline{B}\underline{\Sigma}\underline{B}')^{-1}(\underline{B}\underline{y}) \tag{3.1}$$

For the differentiation we apply the following two rules (see, for instance, Kubik, 1970). Let $\underline{A}$ be a regular symmetric $m \times m$ matrix, which is a function of $\underline{a}$ with $\underline{a} = (a_i)$, hence $\underline{A} = \underline{A}(\underline{a})$. Then with $\dot{\underline{A}}_i = \partial\underline{A}/\partial a_i$ we get

$$\partial \ln\det \underline{A}/\partial a_i = \mathrm{tr}\underline{A}^{-1}\dot{\underline{A}}_i \tag{3.2}$$

and

$$\partial \underline{A}^{-1}/\partial a_i = \underline{A}^{-1}\dot{\underline{A}}_i\underline{A}^{-1} \tag{3.3}$$

To prove (3.2) we use $\underline{A} = \underline{Y}\underline{\Lambda}\underline{Y}'$ with $\underline{Y}'\underline{Y} = \underline{I}$, where $\underline{\Lambda}$ is the diagonal matrix of the singular values λ_i of $\underline{A}$ and $\underline{Y}$ the matrix of singular vectors. We obtain

$$\ln\det\underline{A} = \ln\det\underline{\Lambda} = \ln\prod_{j=1}^{m}\lambda_j = \sum_{j=1}^{m}\ln\lambda_j$$

and

$$\partial \ln \det \underline{A} / \partial a_i = \sum_{j=1}^{m} \lambda_j^{-1} \dot{\lambda}_j = \text{tr}\,\underline{\Lambda}^{-1} \dot{\underline{\Lambda}}$$

Furthermore we get from $\underline{A} = \underline{Y}\underline{\Lambda}\underline{Y}'$

$$\dot{\underline{A}} = \dot{\underline{Y}}\underline{\Lambda}\underline{Y}' + \underline{Y}\dot{\underline{\Lambda}}\underline{Y}' + \underline{Y}\underline{\Lambda}\dot{\underline{Y}}'$$

and with $\underline{A}^{-1} = \underline{Y}\underline{\Lambda}^{-1}\underline{Y}'$

$$\text{tr}\,\underline{A}^{-1}\dot{\underline{A}} = \text{tr}\,\underline{Y}'\dot{\underline{Y}} + \text{tr}\,\underline{\Lambda}^{-1}\dot{\underline{\Lambda}} + \text{tr}\,\underline{Y}\dot{\underline{Y}}'$$

However $\dot{\underline{Y}}'\underline{Y} + \underline{Y}'\dot{\underline{Y}} = \underline{0}$ because of $\underline{Y}'\underline{Y} = \underline{I}$ so that

$$\text{tr}\,\underline{A}^{-1}\dot{\underline{A}} = \text{tr}\,\underline{\Lambda}^{-1}\dot{\underline{\Lambda}}$$

which gives (3.2). From $\underline{A}\underline{A}^{-1} = \underline{I}$ we obtain

$$\dot{\underline{A}}\underline{A}^{-1} + \underline{A}\dot{\underline{A}}^{-1} = \underline{0} \quad \text{and} \quad \dot{\underline{A}}^{-1} = -\underline{A}^{-1}\dot{\underline{A}}\underline{A}^{-1}$$

which gives (3.3).

Applying (3.2) and (3.3) and setting the derivative equal to zero we find with $\underline{\sigma} = (\sigma_i)$ and $\partial\underline{\Sigma}/\partial\sigma_i = \dot{\underline{\Sigma}}_i$ from (3.1)

$$\partial \ln L_1(\underline{y};\underline{\sigma})/\partial\sigma_i = -\frac{1}{2}\text{tr}\,(\underline{B}\underline{\Sigma}\underline{B}')^{-1}\underline{B}\dot{\underline{\Sigma}}_i\underline{B}'$$

$$+\frac{1}{2}(\underline{B}\underline{y})'(\underline{B}\underline{\Sigma}\underline{B}')^{-1}\underline{B}\dot{\underline{\Sigma}}_i\underline{B}'(\underline{B}\underline{\Sigma}\underline{B}')^{-1}(\underline{B}\underline{y}) = 0 \qquad (3.4)$$

By introducing the identity

$$\underline{P}^{-1} = \underline{\Sigma}\underline{P}'(\underline{P}\underline{\Sigma}\underline{P}')^{-1}$$

we find with (2.5) and (2.7)

$$\underline{P}^{-1}\underline{P} = |\,\underline{\Sigma}\underline{B}'(\underline{B}\underline{\Sigma}\underline{B}')^{-1},\ \underline{X}(\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\,| \begin{vmatrix} \underline{B} \\ \underline{X}'\underline{\Sigma}^{-1} \end{vmatrix} = \underline{I}$$

and therefore

$$\underline{B}'(\underline{B}\underline{\Sigma}\underline{B}')^{-1}\underline{B} + \underline{\Sigma}^{-1}\underline{X}(\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{X}'\underline{\Sigma}^{-1} = \underline{\Sigma}^{-1}$$

With

$$\underline{W} = \underline{\Sigma}^{-1} - \underline{\Sigma}^{-1}\underline{X}(\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{X}'\underline{\Sigma}^{-1} \qquad (3.5)$$

we finally obtain

$$\underline{B}'(\underline{B}\underline{\Sigma}\underline{B}')^{-1}\underline{B} = \underline{W} \qquad (3.6)$$

As known, $\underline{y}'\underline{W}\underline{y}$ gives the weighted sum of the squares of the residuals $\hat{\underline{e}}$ in the Gauss-Markoff model which are obtained by

$$\hat{\underline{e}} = -(\underline{I}-\underline{X}(\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{X}'\underline{\Sigma}^{-1})\underline{y} = -\underline{\Sigma}\underline{W}\underline{y} \tag{3.7}$$

Substituting (3.6) in (3.4) we find the equations for the maximum likelihood estimates of the variance and covariance components $\underline{\sigma}$

$$\operatorname{tr}\underline{W}\dot{\underline{\Sigma}}_i - \underline{y}'\underline{W}\dot{\underline{\Sigma}}_i\underline{W}\underline{y} = 0 \quad \text{for } i\epsilon\{1,\ldots,k\} \tag{3.8}$$

or

$$\operatorname{tr}\underline{W}\dot{\underline{\Sigma}}_i - \hat{\underline{e}}'\underline{\Sigma}^{-1}\dot{\underline{\Sigma}}_i\underline{\Sigma}^{-1}\hat{\underline{e}} = 0 \quad \text{for } i\epsilon\{1,\ldots,k\} \tag{3.9}$$

(3.8) and (3.9) are nonlinear equations for the unknown parameters $\underline{\sigma}$. By introducing the $k \times 1$ vectors $\underline{q} = (q_i)$ and $\underline{v} = (v_i)$ we get instead of (3.8)

$$\underline{v} - \underline{q} = \underline{0} \tag{3.10}$$

with

$$q_i = \underline{y}'\underline{W}\dot{\underline{\Sigma}}_i\underline{W}\underline{y} \quad \text{and} \quad v_i = \operatorname{tr}\underline{W}\dot{\underline{\Sigma}}_i$$

4. Newton–Raphson Iteration

To solve the set of nonlinear equations (3.10) or its equivalent form (3.9) we will apply iterations. (3.10) is a function $\underline{f}(\underline{\sigma})$ of $\underline{\sigma}$

$$\underline{f}(\underline{\sigma}) = \underline{v} - \underline{q} = \underline{0} \tag{4.1}$$

where $\underline{f}(\underline{\sigma}) = (f_i(\underline{\sigma}))$ is a $k \times 1$ vector. By introducing the $k \times 1$ vector $\underline{\sigma}_0$ of approximate values of $\underline{\sigma}$, by using the Taylor series and restricting it to the linear terms, we get

$$\underline{f}(\underline{\sigma}) = f(\underline{\sigma}_0) + \left.\frac{\partial \underline{f}(\underline{\sigma})}{\partial \underline{\sigma}}\right|_{\underline{\sigma}=\underline{\sigma}_0} (\underline{\sigma}-\underline{\sigma}_0) = \underline{0}$$

and

$$\underline{f}(\underline{\sigma}_0) + \underline{H}(\underline{\sigma}-\underline{\sigma}_0) = \underline{0} \tag{4.2}$$

with

$$\underline{H} = (h_{ij}) = (\partial f_i(\underline{\sigma})/\partial \sigma_j)\,|_{\underline{\sigma}=\underline{\sigma}_0} \tag{4.3}$$

leading to the Newton–Raphson iteration

$$\underline{\sigma} = \underline{\sigma}_0 - \underline{H}^{-1} f(\underline{\sigma}_0) \tag{4.4}$$

provided the matrix $\underline{H}$ of partial derivatives is regular.

By appling (3.3) to (3.6) we get

$$\partial \underline{W} / \partial \sigma_j = -\underline{B}' (\underline{B} \underline{\Sigma} \underline{B}')^{-1} \underline{B} \dot{\underline{\Sigma}}_j \underline{B}' (\underline{B} \underline{\Sigma} \underline{B}')^{-1} \underline{B} = -\underline{W} \dot{\underline{\Sigma}}_j \underline{W}$$

and with $\partial \dot{\underline{\Sigma}}_i / \partial \sigma_j = \ddot{\underline{\Sigma}}_{ij}$ the elements h_{ij} of the matrix $\underline{H}$ in (4.3)

$$h_{ij} = -\operatorname{tr} \underline{W} \dot{\underline{\Sigma}}_j \underline{W} \dot{\underline{\Sigma}}_i + \operatorname{tr} \underline{W} \ddot{\underline{\Sigma}}_{ij} + 2 \underline{y}' \underline{W} \dot{\underline{\Sigma}}_j \underline{W} \dot{\underline{\Sigma}}_i \underline{W} \underline{y} - \underline{y}' \underline{W} \ddot{\underline{\Sigma}}_{ij} \underline{W} \underline{y} \tag{4.5}$$

We recognize, that the coefficient matrix $\underline{H}$ in (4.2) for the equations of the variance and covariance components depends on the observations y. In order to facilitate the solution of this system of equations we make two compromises. First, we replace the matrix $\underline{H}$ by the matrix of expected values, thus eliminating the dependence on $\underline{y}$; and second, we neglect the second partial derivatives, so that instead of the matrix $\underline{H}$ the matrix $\underline{S}$ with $\underline{S} = (s_{ij})$ is obtained.

With

$$E(\underline{y}\underline{y}') = \underline{\Sigma} + \underline{\mu}_y \underline{\mu}_y'$$

where $\underline{\mu}_y = \underline{X}\underline{\beta}$ and from (3.5) with

$$\underline{W} = \underline{W} \underline{\Sigma} \underline{W} \quad \text{and} \quad \underline{W}\underline{X} = \underline{0} \tag{4.6}$$

we get

$$E(\underline{y}' \underline{W} \dot{\underline{\Sigma}}_j \underline{W} \dot{\underline{\Sigma}}_i \underline{W} \underline{y}) = \operatorname{tr}(\underline{W} \dot{\underline{\Sigma}}_j \underline{W} \dot{\underline{\Sigma}}_i \underline{W} E(\underline{y}\underline{y}')) = \operatorname{tr} \underline{W} \dot{\underline{\Sigma}}_i \underline{W} \dot{\underline{\Sigma}}_j$$

and therefore instead of (4.5)

$$s_{ij} = \operatorname{tr} \underline{W} \dot{\underline{\Sigma}}_i \underline{W} \dot{\underline{\Sigma}}_j \tag{4.7}$$

Substituting this result in (4.2) gives

$$\underline{S}(\underline{\sigma} - \underline{\sigma}_o) = \underline{q} - \underline{v} \tag{4.8}$$

It should be emphasized that according to (4.2) the matrix $\underline{S}$ and the vectors $\underline{q}$ and $\underline{v}$ in (4.8) have to be evaluated at the approximate values $\underline{\sigma}_o$.

At this point we have to specify the parametrisation of the covariance matrix $\underline{\Sigma}$ of the observations $\underline{y}$. In general, $\underline{\Sigma} = \underline{\Sigma}(\underline{\sigma})$ will be a nonlinear function of the variance and covariance components. As mentioned, $\underline{\Sigma}(\underline{\sigma}_o)$ is needed. If we use the Taylor series and restrict it to the linear terms, we get

$$\underline{\Sigma}(\underline{\sigma}_o) = \underline{\Sigma}(\underline{\sigma}_e) + \sum_{i=1}^{k} \left. \frac{\partial \underline{\Sigma}(\underline{\sigma})}{\partial \sigma_i} \right|_{\underline{\sigma} = \underline{\sigma}_e} (\sigma_{oi} - \sigma_{ei})$$

where $\underline{\sigma}_e = (\sigma_{ei})$ is the point of expansion. We will assume

$$\underline{\sigma}_e = \underline{0} \quad \text{and} \quad \underline{\Sigma}(\underline{\sigma}_e) = \underline{0}$$

so that

$$\underline{\Sigma}(\underline{\sigma}_o) = \sum_{i=1}^{k} \frac{\partial \underline{\Sigma}(\underline{\sigma})}{\partial \sigma_i}\Bigg|_{\underline{\sigma}=\underline{\sigma}_o} \sigma_{oi} \qquad (4.9)$$

The derivative $\dot{\underline{\Sigma}}_i$ in (4.8) is therefore given by

$$\underline{\Sigma}_i = \partial \underline{\Sigma}(\underline{\sigma}) / \partial \sigma_i \,|_{\underline{\sigma}=\underline{\sigma}_o} \qquad (4.10)$$

Using (4.9) we can simplify (4.8) because of

$$\underline{S}\underline{\sigma}_o = \underline{v} \qquad (4.11)$$

To show this we compute the element i of the vector $\underline{S}\underline{\sigma}_o$

$$\sum_{j=1}^{k} s_{ij}\sigma_{oj} = \sum_{j=1}^{k} (\mathrm{tr}\underline{W}\dot{\underline{\Sigma}}_i \underline{W}\dot{\underline{\Sigma}}_j)\sigma_{oj} = \mathrm{tr}\underline{W}\dot{\underline{\Sigma}}_i\underline{W}\sum_{j=1}^{k}\dot{\underline{\Sigma}}_j\sigma_{oj}$$

and obtain with (4.9) and (4.6)

$$\sum_{j=1}^{k} s_{ij}\sigma_{oj} = \mathrm{tr}\underline{W}\dot{\underline{\Sigma}}_i\underline{W}\underline{\Sigma} = \mathrm{tr}\underline{W}\dot{\underline{\Sigma}}_i = v_i$$

Substituting (4.11) in (4.8) we finally get the system of equations for the estimates $\hat{\underline{\sigma}}$ of the variance and covariance components $\underline{\sigma}$

$$\underline{S}\hat{\underline{\sigma}} = \underline{q} \qquad (4.12)$$

As already mentioned, the matrix $\underline{S}$ and the vector $\underline{q}$ have to be evaluated at the approximate values $\underline{\sigma}_o$. According to (4.4) the estimates (4.12) have to be applied in iterations. By using the estimates to compute new approximate values, we can iterate until at the point of convergence the estimates reproduce themselves, which means

$$\hat{\underline{\sigma}} = \underline{\sigma}_o \qquad (4.13)$$

With this result we get from (4.11) and (4.12)

$$\underline{S}\underline{\sigma}_o = \underline{v} = \underline{S}\hat{\underline{\sigma}} = \underline{q} \qquad (4.14)$$

This equation fulfils (3.10) so that at the point of convergence the maximum likelihood estimate is obtained.

Very often it is assumed that the covariance matrix $\underline{\Sigma}$ is a linear function of the variance and covariance components $\underline{\sigma}$

$$\underline{\Sigma} = \sum_{i=1}^{k} \underline{V}_i \sigma_i \qquad (4.15)$$

with

$$\underline{V}_i = \underline{T}_i \, \sigma_{oi}$$

where $\underline{V}_i$ and $\underline{T}_i$ are symmetric matrices and σ_{oi} are approximate values of the products $\sigma_{oi}\,\sigma_i$, so that the values of the unknown variance and covariance components are assumed to be close to one. With (4.15) $\dot{\underline{\Sigma}}_i$ in (4.12) is given by

$$\dot{\underline{\Sigma}}_i = \underline{V}_i \tag{4.16}$$

and the matrix $\underline{\Sigma}$ evaluated at the approximate values $\underline{\sigma}_o$ is obtained by

$$\underline{\Sigma}(\underline{\sigma}_o) = \sum_{i=1}^{k} \underline{V}_i \tag{4.17}$$

By substituting (4.16) and (4.17) in (4.12) the locally best invariant quadratic unbiased estimate of the variance and covariance components is obtained, see for instance (Koch, 1980, p. 211), which is identical with the MINQUE-estimate (Rao, 1973, p. 304).

The name locally is chosen because the estimates $\hat{\underline{\sigma}}$ depend through $\underline{\Sigma}(\underline{\sigma}_o)$ on the approximate values $\underline{\sigma}_o$. By using the estimates to compute new approximate values, we can iterate until at the point of convergence the estimates reproduce themselves. At this point (3.10) is fulfiled so that the reproducing estimates of the locally best invariant quadratic unbiased estimation are also maximum likelihood estimates. It should be noted here, that the iterated estimates might not be unbiased anymore.

5. Successive Approximations

To obtain successive approximations for the reproducing estimates computed with (4.12), (4.14) can be used. We replace the k x k matrix $\underline{S}$ by the k x k diagonal matrix $\underline{D} = diag\,(d_i)$

$$\underline{D}\,\underline{\sigma}_o = \underline{v}$$

so that

$$d_i = v_i / \sigma_{oi} \tag{5.1}$$

This result is substituted in $\underline{D}\,\hat{\underline{\sigma}} = \underline{q}$, and we obtain

$$\hat{\sigma}_i = (q_i / v_i)\,\sigma_{oi} \tag{5.2}$$

or for the model (4.15)

$$\hat{\sigma}_i = (q_i / v_i) \tag{5.3}$$

By using the estimates to improve the approximate values, (5.2) or (5.3) are applied successively, until at the point of convergence the estimates reproduce themselves, so that (3.10) is fulfilled. Hence, at the point of convergence the maximum likelihood estimates and the reproducing estimates of the locally best invariant quadratic unbiased estimation are obtained. (5.3) has already been derived using different approaches by (Förstner, 1979), (Persson, 1980, p. 89) or by (Lucas, 1985) who compared for the model (4.15) the computational effort to obtain the reproducing estimates when applying (4.12) and (5.3). He found the latter method to be more efficient.

o

o o

REFERENCES

W. FÖRSTNER : Ein Verfahren zur Schätzung von Varianz- und Kovarianz komponenten. Allgemeine Vermessungs-Nachrichten, 96, pp. 446–453, 1979.

D.A. HARVILLE : Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association, 72, pp. 320–338, 1977.

K.R. KOCH : Parameterschätzung und Hypothesentests in linearen Modellen. Dümmler, Bonn, 1980.

K. KUBIK : The estimation of the weights of measured quantities within the method of least squares. Bulletin Géodésique, 95, pp. 21–40, 1970.

J.R. LUCAS : A variance component estimation method for sparse matrix applications. NOAA Technical Report NOS 111 NGS 33, National Geodetic Survey, Rockville, Md., 1985.

C.–G. PERSSON : MINQUE and related estimators for variance components in linear models. Royal Institute of Technology, Stockholm, 1980.

C.R. RAO : Linear Statistical Inference and Its Applications. John Wiley, New York, 1973.

B. SCHAFFRIN : Varianz–Kovarianz–Komponenten–Schätzung bei der Ausgleichung heterogener Wiederholungsmessungen. Deutsche Geodätische Kommission, C, 282, München, 1983.

Received : 08.11.1985
Accepted : 26.06.86