Sankhyā : The Indian Journal of Statistics
1979, Volume 41, Series B, Pts. 3 and 4, pp. 138–153.

MINQE THEORY AND ITS RELATION TO ML AND MML ESTIMATION OF VARIANCE COMPONENTS

By C. RADHAKRISHNA RAO*
University of Pittsburgh
and
Indian Statistical Institute

SUMMARY. Starting with the general linear model $Y = X\beta + \epsilon$ where $E(\epsilon\epsilon') = \theta_1 V_1 + \ldots + \theta_p V_p$, the theory of minimum norm quadratic estimation (MINQE) of the parameter $\theta = (\theta_1, \ldots, \theta_p)'$ is developed. The method depends on the choice of a natural quadratic estimator of θ in terms of the unobservable variable ϵ and comparing it with a quadratic estimator $Y'AY$ in terms of the observable variable Y. The matrix A is determined by minimizing the difference between two quadratic forms. By placing restrictions on $Y'AY$ such as unbiasedness (U), invariance (I) under translation of Y by $X\beta$, different kinds of MINQE's such as MINQE (I), MINQE(U), MINQE(U, I), etc. are generated.

A class of iterated MINQE's (IMINQE's) is developed to obtain estimators free from apriori information used in the construction of MINQE's. This class is shown to include maximum likelihood (ML) and marginal ML (MML) estimators. Thus the MINQE principle provides a unified theory of estimation of variance components.

1. Introduction

MINQE (minimum norm quadratic estimation) of variance components in a general linear model was proposed by the author in a series of papers (Rao, 1970, 1971a, 1971b, 1972, 1973), the scope of which was extended to cover a wide variety of situations by Focke and Dewess (1972), Kleffe (1975, 1976, 1977a, b, 1978, 1979), Pukelsheim (1977) and Rao and Chaubey (1978). The purpose of this paper is to extended the MINQE theory a little further, introduce a class of estimators called IMINQE (iterated MINQE) and show its relationship to MLE (maximum likelihood estimator) of Hartley and J. N. K. Rao (1969) and MMLE (marginal maximum likelihood estimator) of Patterson and Thompson (1975). An extensive review of ML and MML estimation and the computational algorithms is contained in a recent paper by Harville (1977), where the MML was termed as the REML (restricted maximum likelihood).

* The work of this author is sponsored by the Air Force Office of Scientific Research, Air Force Systems Command under Contract F 49620-70-C-0161. Reproduction in whole or in part is permitted for any purpose of the United States Government.

2. MINQE PRINCIPLE

Mixed linear model. The usual mixed linear model discussed in the literature on variance components is

$$Y = X\beta + U_1\Phi_1 + \ldots + U_{p-1}\Phi_{p-1} + \Phi_p \quad \ldots \quad (2.1)$$

where $X(n \times m)$, U_i $(n \times n_i)$ are known matrices, β is a fixed unknown vector parameter and Φ_i are unobservable random variables such that

$$E(\Phi_i) = 0, \quad E(\Phi_i \Phi_j') = 0$$
$$E(\Phi_i \Phi_i') = \sigma_i^2 I_{n_i}, \quad i = 1, \ldots, p \quad \ldots \quad (2.2)$$

using the convention $n = n_p$. From (2.1) and (2.2),

$$E(Y) = X\beta, \; D(Y) = \sigma_1^2 V_1 + \ldots + \sigma_{p-1}^2 V_{p-1} + \sigma_p^2 I \quad \ldots \quad (2.3)$$

where $V_i = U_i U_i' \geqslant 0$ (i.e., non-negative definite). The unknown non-negative parameters $\sigma_1^2, \ldots, \sigma_p^2$ are called variance components.

General linear model. We consider a more general model

$$Y = X\beta + \epsilon \quad \ldots \quad (2.4)$$

where the stochastic structure of the unobservable random variable ϵ is not specified as in (2.1, 2.2) but it is known that

$$E(\epsilon) = 0, \quad E(\epsilon\epsilon') = \theta_1 V_1 + \ldots + \theta_p V_p \quad \ldots \quad (2.5)$$

where the parameters θ_i need not be non-negative and the matrices V_i are symmetric but need not be n.n.d. The model (2.4, 2.5) covers both variance and covariance component models, examples of which are given in Krishnaiah and Lee (1974), Kleffe (1978) and Rao and Kleffe (1979).

MINQE principle. Suppose that Φ_i in the model (2.1, 2.2) are observable. Then, natural estimators of the variance components are

$$\hat{\sigma}_i^2 = \Phi_i' \Phi_i \div n_i, \quad i = 1, \ldots, p \quad \ldots \quad (2.6)$$

and the estimator of a linear function $\psi = f_1\sigma_1^2 + \ldots + f_p\sigma_p^2$ is

$$\hat{\psi} = f_1\hat{\sigma}_1^2 + \ldots + f_p\hat{\sigma}_p^2 = \xi' N \xi \quad \ldots \quad (2.7)$$

where $\xi = (\alpha_1^{-1}\Phi_1', \ldots, \alpha_p^{-1}\Phi_p')'$, N is a suitably chosen matrix and $\alpha_1, \ldots, \alpha_p$ are apriori values of $\sigma_1^2, \ldots, \sigma_p^2$. (Note that $\xi' N \xi$ is independent of α_i).

Let us consider an estimator of ψ in terms of the observable Y of the form

$$(Y-X\beta_0)'A(Y-X\beta_0) \qquad \ldots \quad (2.8)$$

where β_0 is an apriori value of β. The expression (2.8) is equal to

$$\xi' U_a' A U_a \xi + \nu' K^{\frac{1}{2}} X' A X K^{\frac{1}{2}} \nu + 2\nu' K^{\frac{1}{2}} X' A U_a \xi \qquad \ldots \quad (2.9)$$

where $\nu = K^{-\frac{1}{2}}(\beta-\beta_0)$, $U_a' = (\alpha_1^{\frac{1}{2}} U_1 : \ldots : \alpha_p^{\frac{1}{2}} U_p)$, K is a positive definite matrix, which is in the nature of an apriori dispersion matrix of β and $K^{\frac{1}{2}}$ is a symmetric square root of K. The difference between (2.9) and the natural estimator (2.7) is

$$(\xi' : \nu') \begin{pmatrix} U_a' A U_a - N & U_a A X K^{\frac{1}{2}} \\ K^{\frac{1}{2}} X' A U_a & K^{\frac{1}{2}} X' A X K^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \xi \\ \nu \end{pmatrix}. \qquad \ldots \quad (2.10)$$

The MINQE (minimum norm quadratic estimator) of ψ is $(Y-X\beta_0)'A_*(Y-X\beta_0)$ where A_* is chosen to minimize a suitably chosen norm of the quadratic form (2.10)

$$\left\| \begin{matrix} U_a' A U_a - N & U_a' A X K^{\frac{1}{2}} \\ K^{\frac{1}{2}} X' A U_a & K^{\frac{1}{2}} X' A X K^{\frac{1}{2}} \end{matrix} \right\| \qquad \ldots \quad (2.11)$$

with A restricted to a chosen class of matrices depending on desired properties for the estimators, such as unbiasedness, non-negative definiteness, etc.

The choice of a natural estimator is obvious when the linear model is as in (2.1, 2.2) with the structure of the error component in Y specified in terms of structural variables. The method outlined covers more general situations where $E(\Phi_i \Phi_i') = \Lambda_i$ containing unknown variance and covariance components. (see for example the problems considered by Rao, 1972, Rao and Kleffe, 1979 and Sinha and Wieand, 1977.)

Suppose the model is given in the form (2.4, 2.5)

$$Y = X\beta + \epsilon, \quad D(\epsilon) = \theta_1 V_1 + \ldots + \theta_p V_p \qquad \ldots \quad (2.12)$$

without specifying the structure of ϵ. Choosing $\alpha_1, \ldots, \alpha_p$ as prior values of $\theta_1, \ldots, \theta_p$ and (β_0, K) as the prior mean and dispersion matrix of β, we can transform (2.12) to

$$(Y - X\beta_0) = X K^{\frac{1}{2}} \nu + V_a^{\frac{1}{2}} \eta, \quad V_a = \alpha_1 V_1 + \ldots + \alpha_p V_p \qquad \ldots \quad (2.13)$$

and define a natural estimator of $\psi = f'\theta$ as

$$\eta' N_* \eta = \eta'(\Sigma \lambda_i V_\alpha^{-1} V_i V_\alpha^{-1})\eta \qquad \ldots \quad (2.14)$$

where $\lambda = (\lambda_1, \ldots, \lambda_p)'$ is determined such that $E(\eta' N_* \eta) = f'\theta$ which implies that λ is a solution of $[H(\alpha)]\lambda = f$ where

$$H(\alpha) = (\text{tr } V_\alpha^{-1} V_i V_\alpha^{-1} V_j). \qquad \ldots \quad (2.15)$$

Note that $H(\alpha)$ is non-singular if θ is identifiable, i.e., $\theta_1 V_1 + \ldots + \theta_p V_p = \theta_1' V_1 + \ldots + \theta_p' V_p \Longleftrightarrow \theta_i = \theta_i'$ for all i. The choice of a natural estimator as in (2.14) may be justified in the following way. Suppose ε in (2.12) has the structure of the error term in (2.1), in which case $\eta' N_* \eta = \xi' M \xi$. Then it is easily shown that N_* is as in (2.14) if we minimize the Euclidean norm of $(M - N)$, where N is as in (2.7), with respect M subject to the condition that $\xi' M \xi$ is unbiased for $f'\theta$.

If the estimator of ψ in terms of Y is

$$\begin{aligned} &(Y - X\beta_0)' A (Y - X\beta_0) \\ &= \eta' V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} \eta + \nu' K^{\frac{1}{2}} X' A X K^{\frac{1}{2}} \nu + 2\eta' V_\alpha^{\frac{1}{2}} A X K^{\frac{1}{2}} \nu \end{aligned} \qquad \ldots \quad (2.16)$$

then the norm of the difference between (2.16) and (2.14) is

$$\left\| \begin{matrix} V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N_* & V_\alpha^{\frac{1}{2}} A X K^{\frac{1}{2}} \\ K^{\frac{1}{2}} X' A V_\alpha^{\frac{1}{2}} & K^{\frac{1}{2}} X' A X K^{\frac{1}{2}} \end{matrix} \right\| . \qquad \ldots \quad (2.17)$$

For the MINQE estimator of ψ from the general model (2.13) we choose A_* which minimizes (2.17). Since the matrices A minimizing (2.11) and (2.17) may be different, the MINQE estimator may be dependent on the form in which the error component is specified in the linear model.

The MINQE principle as stated requires for its application inputs such as prior value α of θ, prior mean and dispersion (β_0, K) of β, the choice of a natural estimator and a suitable norm of the difference between two quadratic forms (as in 2.11). If we have the necessary prior information, then there is no problem. Otherwise, we can use various devices. For instance, if we want the estimator $Y'AY$ to be translation invariant, i.e., $Y'AY = (Y - X\beta)'A(Y - X\beta) \forall \beta$, which may be a desirable property, then MINQE(I), i.e., subject to the condition of invariance, is independent of β_0 and K, so that any arbitrariness due to the choice of prior mean and dispersion of β is eliminated.

If β is known to be large compared to θ, then we may put $\beta_0 = 0$ and consider ∞-MINQE, the limit of MINQE as the diagonal elements of K tend to infinity as an appropriate estimator (see Focke and Dewess, 1972). It is shown that if we choose a simple Euclidean norm for (2.11) and impose the condition of unbiasedness, the MINQE (U), i.e., subject to the condition of unbiasedness, is independent of the choice of the natural estimator (whether 2.7 or 2.14 is chosen). Another possibility is to choose an initial value α for θ and compute the MINQE, say $\hat{\theta}_1$. Choosing $\hat{\theta}_1$ as the initial value we can recompute the MINQE, say $\hat{\theta}_2$. The limiting estimator, if it exists and is independent of α, is called IMINQE, (iterated MINQE). We shall examine some of these aspects in detail in the discussion of various MINQE's.

3. MINQE WITH UNBIASEDNESS

Consider the general model

$$Y = X\beta + \epsilon,\ D(\epsilon) = \theta_1 V_1 + \ldots + \theta_p V_p = V_\theta, \qquad \ldots \quad (3.1)$$

where $\theta \in \mathscr{F}$, an open set in R^p such that $V_\theta > 0$ (i.e., positive definite) for all $\theta \in \mathscr{F}$. (The condition $V_\theta > 0$ can be relaxed to $V_\theta \geqslant 0$, but the results are a little complicated.)

3.1 *MINQE* (U, I). Consider a quadratic estimator $Y'AY$ which satisfies the (U, I) condition, i.e., unbiased for $f'\theta$ and invariant for translation of Y by $X\beta$. It is seen that

$$Y'AY \text{ satisfies } (U, I) \Longleftrightarrow \text{tr}\, AV_i = f_i,\ i = 1, \ldots, p \text{ and } AX = 0. \quad \ldots \quad (3.2)$$

Under the condition (3.2), choosing the Euclidean norm for the difference of quadratic forms, the square of (2.17) becomes

$$\begin{aligned} &\text{tr}\,(V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N_*)(V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N_*) \\ &= \text{tr}\, V_\alpha A V_\alpha A - 2\,\text{tr}\, A V_\alpha^{\frac{1}{2}} N_* V_\alpha^{\frac{1}{2}} + \text{tr}\, N_* N_*. \end{aligned} \qquad \ldots \quad (3.3)$$

In view of (3.2) and the choice of N_* as in (2.14), the second and third terms in (3.3) are independent of A and the only quantity to be minimized is $\text{tr}\, AV_\alpha AV_\alpha$ subject to (3.2). Using the method outlined in Rao (1971), the optimum choice of A is

$$A_* = V_\alpha^{-1} R_\alpha (\Sigma \mu_i V_i) R_\alpha' V_\alpha^{-1} \qquad \ldots \quad (3.4)$$

where $R_\alpha = I - P_\alpha$, $P_\alpha = X(X'V_\alpha^{-1}X)^{-}X'V_\alpha^{-1}$ is the projection operator onto the space generated by the columns of X and $\mu = (\mu_1, \ldots, \mu_p)'$ is a solution of

$$[H_{UI}(\alpha)]\mu = f \qquad \ldots \quad (3.5)$$

where

$$H_{UI}(\alpha) = (\text{tr } V_\alpha^{-1}R_\alpha V_i V_\alpha^{-1}R_\alpha' V_j). \qquad \ldots \quad (3.6)$$

The MINQE (U, I) of $f'\theta$ is $Y'A_*Y$ which can be written in the form $f'\hat{\theta}$ where $\hat{\theta}$ is a solution of the consistent equation

$$[H_{UI}(\alpha)]\theta = h_I(Y, \alpha) \qquad \ldots \quad (3.7)$$

where $h_I(Y, \alpha)$ is a vector with the i-th component as

$$Y'V_\alpha^{-1}R_\alpha V_i R_\alpha' V_\alpha^{-1}Y. \qquad \ldots \quad (3.8)$$

3.2. *IMINQE* (U, I). The equation (3.7) is consistent, but the solution is not unique unless θ is identifiable on the basis of the distribution of $T'Y$ the maximal invariant of Y where $T = X^{\perp}$, i.e., unless the matrices $T'V_iT$, $i = 1, \ldots, p$ are linearly independent. In such a case $H_{UI}(\alpha)$ is non-singular and the MINQE (U, I) of θ exists and is given by

$$\hat{\theta}_1 = [H_{UI}(\alpha)]^{-1}h_I(Y, \alpha). \qquad \ldots \quad (3.9)$$

If we choose $\hat{\theta}_1$ as the initial value of θ and recompute the MINQE (U, I), we get

$$\hat{\theta}_2 = [H_{UI}(\hat{\theta}_1)]^{-1}\, h_I(Y, \hat{\theta}_1). \qquad \ldots \quad (3.10)$$

If the process is continued and the solution for θ converges, the limiting value satisfies the equation

$$[H_{UI}(\theta)]\theta = h_I(Y, \theta). \qquad \ldots \quad (3.11)$$

We define the solution of (3.11) as IMINQE (U, I), the iterated MINQE (U, I). The equation (3.11) can also be written in the form

$$\text{tr } V_\theta^{-1}R_\theta V_i = Y'R_\theta' V_\theta^{-1}V_i V_\theta^{-1}R_\theta Y, \; i = 1, \ldots, p \qquad \ldots \quad (3.12)$$

which are the MML (marginal maximum likelihood) equations of Patterson and Thompson (1975). IMINQE (U, I) may not be unbiased.

3.3. *MINQE* (*U*). Consider the linear model with four observations

$$Y_1 = \beta_1+\epsilon_1, \quad Y_3 = \beta_2+\epsilon_3$$

$$Y_2 = \beta_1+\epsilon_2, \quad Y_4 = \beta_2+\epsilon_4$$

$$E(\epsilon_i\epsilon_j) = 0,\ E(\epsilon_1^2) = E(\epsilon_3^2) = \sigma_1^2,\ E(\epsilon_2^2) = E(\epsilon_4^2) = \sigma_2^2 \qquad \dots \quad (3.13)$$

which is a special case of the problem considered by Focke and Dewess (1972). It is easily seen that there does not exist a matrix A satisfying the condition (3.2) so that invariant unbiased estimators of σ_1^2 and σ_2^2 do not exist. In such cases we may impose only the condition of unbiasedness, i.e.,

$$E(Y'AY) = f'\theta \Longleftrightarrow X'AX = 0,\ \text{tr}\, AV_i = f_i,\ i = 1, \dots, p. \qquad \dots \quad (3.14)$$

Then the square of the norm (2.17) becomes

$$\text{tr}\,(V_\alpha^{\frac{1}{2}}AV_\alpha^{\frac{1}{2}}-N_\bullet)(V_\alpha^{\frac{1}{2}}AV_\alpha^{\frac{1}{2}}-N_\bullet)+2\,\text{tr}\, V_\alpha^{\frac{1}{2}}AXKX'AV_\alpha^{\frac{1}{2}}. \qquad \dots \quad (3.15)$$

The terms in (3.15) which depend on A can be written as

$$\text{tr}\,(V_\alpha+XKX')A(V_\alpha+XKX')A. \qquad \dots \quad (3.16)$$

Proceeding in the manner indicated in equations (3.4)-(3.8), MINQE (U) of $f'\theta$ is $f'\hat{\theta}$ where $\hat{\theta}$ is a solution of

$$[H_U(\alpha)]\,\theta = h_U(Y, \alpha) \qquad \dots \quad (3.17)$$

where

$$H_U(\alpha) = (\text{tr}\,(V_i+XKX')^{-1}(V_i-P_\alpha V_i P_\alpha^*)(V_\alpha+XKX')^{-1}V_j)$$

and the i-th component of $h_U(Y, \alpha)$ is

$$(Y-X\beta_0)'(V_\alpha+XKX')^{-1}(V_i-P_\alpha V_i P_\alpha')(V_\alpha+XKX')^{-1}(Y-X\beta_0) \qquad \dots \quad (3.18)$$

where $$P_\alpha = X(X'(V_\alpha+XKX')^{-1}X)^{-1}X'(V_\alpha+XKX')^{-1}.$$

If θ is identifiable, the IMINQE (U) of θ is the solution of

$$[H_U(\theta)]\theta = h_U(Y, \theta). \qquad \dots \quad (3.19)$$

Note 1. The equations (3.7) and (3.11) for MINQE (U, I) and IMINQE (U, I) and equations (3.17) and (3.19) for MINQE(U) and IMINQE(U)

remain the same even if we start with the model (2.1, 2.2) and minimize (2.11). Thus with unbiasedness, the MINQE is independent of the choice of the natural estimator as (2.7) or (2.14).

Note 2. MINQE (U, I) does not depend on β_0, K, but MINQE (U) does.

4. MINQE WITHOUT UNBIASEDNESS

We shall obtain the MINQE without unbiasedness but satisfying the condition of invariance for the model (2.1, 2.2) by minimizing (2.11) and for the general model (2.4, 2.5) by minimizing (2.17). We denote such an estimator by MINQE (I). Particular cases of MINQE (I) for the model (2.1, 2.2) have been studied by Rao and Chaubey (1978).

Alternative 1. For the model (2.1, 2.2), the matrix N of the natural estimator (2.7) of $f'\theta$ is a block diagonal matrix with the i-th block equal to $(\alpha_i f_i/n_i)I_{n_i}$. The expression to be minimized is

$$\operatorname{tr}(U_a'AU_a-N)(U_a'AU_a-N). \qquad \dots \quad (4.1)$$

The minimum of (4.1) is attained at A_* iff

$$\operatorname{tr}(U_a'A_*U_a-N)U_a'DU_a = 0 \qquad \dots \quad (4.2)$$

for all D such that $DX = 0 \Longrightarrow D = R_a'ER_a$ for arbitrary E. Substituting $D = R_a'ER_a$, the condition (4.2) becomes

$$\operatorname{tr}(R_aV_aA_*V_aR_a'-R_aU_aNU_a'R_a')E = 0$$

$$\Longrightarrow V_aR_a'A_*R_aV_a = R_a(c_1V_1+\dots+c_pV_p)R_a' \qquad \dots \quad (4.3)$$

where $c_i = \alpha_i^2 f_i/n_i$. From (4.3),

$$\begin{aligned} A_* &= V_a^{-1}R_a(c_1V_1+\dots+c_pV_p)R_a'V_a^{-1} \\ &= \Sigma c_i R_a'V_a^{-1}V_iV_a^{-1}R_a \qquad \dots \quad (4.4) \end{aligned}$$

and the MINQE (I) of $f_i'\theta$ is $Y'A_*Y$. In particular, the MINQE (I) of θ is

$$\hat{\theta} = \begin{Bmatrix} \frac{\alpha_1^2}{n_1} & & \\ & \ddots & \\ & & \frac{\alpha_p^2}{n_p} \end{Bmatrix} h_I(Y, \alpha) \qquad \dots \quad (4.5)$$

and the IMINQE (I) of θ is a solution of the equation

$$\theta = \begin{bmatrix} \frac{\theta_1^2}{n_1} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \frac{\theta_p^2}{n_p} \end{bmatrix} h_I(Y, \theta). \qquad \dots \quad (4.6)$$

Alternative 2. Let us consider the general model (2.12) with the natural estimator (2.14) and determine A by minimizing (2.17). The matrix N_* of the natural estimator is

$$\Sigma \lambda_i V_\alpha^{-\frac{1}{2}} V_i V_\alpha^{-\frac{1}{2}} \qquad \dots \quad (4.7)$$

where $\lambda = (\lambda_1, \dots, \lambda_p)'$ is a solution of $[H(\alpha)]\lambda = f$ where $H(\alpha) = (\text{tr } V_\alpha^{-1} V_i V_\alpha^{-1} V_j)$. The expression to be minimized is

$$\text{tr } (V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N_*)(V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N_*). \qquad \dots \quad (4.8)$$

The minimum of (4.8) is attained at A_* iff

$$\text{tr } (V_\alpha^{\frac{1}{2}} A V_\alpha^{\frac{1}{2}} - N_*) V_\alpha^{\frac{1}{2}} D V_\alpha^{\frac{1}{2}} = 0$$

for all D of the form $R_\alpha' E R_\alpha$ for arbitrary E. Then proceeding as in steps (4.2)-(4.4) we have

$$A_* = V_\alpha^{-1} R_\alpha (\Sigma \lambda_i V_i) R_\alpha' V_\alpha^{-1}$$

and the MINQE (I) of $f'\theta$ is $Y' A_* Y$, which can be written as $f'\hat{\theta}$ where $\hat{\theta}$ is a solution of

$$[H(\alpha)]\theta = h_I(Y, \alpha) \qquad \dots \quad (4.9)$$

where

$$H(\alpha) = \text{tr } (V_\alpha^{-1} V_i V_\alpha^{-1} V_j)$$

$$\dots \quad (4.10)$$

$$h_I(Y, \alpha) = (Y' V_\alpha^{-1} R_\alpha V_1 R_\alpha' V_\alpha^{-1} Y, \dots, Y' V_\alpha^{-1} R_\alpha V_p R_\alpha' V_\alpha^{-1} Y)'.$$

If θ is identifiable, the MINQE (I) of θ is the solution of (4.9) and the IMINQE(I) is the solution of

$$[H(\theta)]\theta = h_I(Y, \theta). \qquad \dots \quad (4.11)$$

The equation (4.11) can be written in the form

$$\text{tr}\, V_0^{-1} V_i = Y' R_0' V_0^{-1} V_i V_\theta^{-1} R_\theta Y, \quad i = 1, \ldots, p \qquad \ldots \quad (4.12)$$

which are the maximum likelihood (ML) equations of Hartley and Rao (1967).

The estimators (4.5) and (4.9) and the iterated versions (4.6) and (4.11) are in general different and depend on the choice of the natural estimator. It may be seen that if $V_i \geqslant 0$, then the estimator (4.5) uses the information on the actual structure of the error component where as (4.9) does not. The ML method ignores any given structure of the error component.

5. Maximum likelihood (ML) estimation

5.1. *ML equations.* We consider the general model

$$Y = X\beta + \epsilon, \; D(\epsilon) = \theta_1 V_1 + \ldots + \theta_p V_p = V_\theta \qquad \ldots \quad (5.1)$$

and discuss the ML estimation of θ under the assumption

$$Y \sim N_n(X\beta, V_\theta), \; \beta \in \mathcal{R}^m, \; \theta \in \mathcal{F}. \qquad \ldots \quad (5.2)$$

We assume that $V_\theta > 0 \; \forall \; \theta \in \mathcal{F}$, but do not place any restriction on θ_i such as non-negativity or on V_i such as non-negative definiteness. [However, we may note that by a linear transformation of θ we can write

$$V_\theta = \tau_1 W_1 + \ldots + \tau_p W_p \qquad \ldots \quad (5.3)$$

where $W_i > 0$ but τ_i are not necessarily non-negative.]

The log likelihood of the unknown parameters β, θ is

$$l(\beta, \theta, Y) = -\log |V_\theta| - (Y - X\beta)' V_\theta^{-1} (Y - X\beta). \qquad \ldots \quad (5.4)$$

The proper ML estimator of (β, θ) is a value $(\hat{\beta}, \hat{\theta})$ such that

$$l(\hat{\beta}, \hat{\theta}, Y) = \sup_{\beta \in \mathcal{R}^m, \, \theta \in \mathcal{F}} l(\beta, \theta, Y). \qquad \ldots \quad (5.5)$$

The ML equations obtained by equating the derivatives of (5.4) w.r.t. (β, θ) to zero are

$$X' V_\theta^{-1} X \beta = X' V_\theta^{-1} Y \qquad \ldots \quad (5.6)$$

$$\text{tr}\, V_\theta^{-1} V_i = (Y - X\beta)' V_\theta^{-1} V_i V_0^{-1} (Y - X\beta), \quad i = 1, \ldots, p. \qquad \ldots \quad (5.7)$$

Substituting for β in (5.7) from (5.6), the equations become

$$X\beta = P_\theta Y, \quad \ldots \quad (5.8)$$

$$[H(\theta)]\theta = h_I(Y, \theta) \quad \ldots \quad (5.9)$$

where $P_\theta = X(X'V_\theta^{-1}X)^- X'V_\theta^{-1}$ and $H(\theta)$ and $h_I(Y, \theta)$ are as defined in (4.9).

We make a few comments on the ML equation (5.9) which provide direct estimates of θ.

(i) The ML equation (5.9) is the same as that for the second alternative IMINQE (I) given in (4.11).

(ii) The equations (5.6) and (5.7) are unbiased while (5.9) is not in the sense

$$E[h_I(Y, \theta)] \neq [H_I(\theta)]\theta. \quad \ldots \quad (5.10)$$

The equation obtained by equating $h_I(Y, \theta)$ to its expected value is

$$[H_{UI}(\theta)]\theta = h_I(Y, \theta) \quad \ldots \quad (5.11)$$

which is the MML equation of Patterson and Thompson (1975).

(iii) There may be no solution to (5.9) in the admissible set $\mathcal{F}$ to which θ belongs. This may happen when the supremum is attained at a boundary point of $\mathcal{F}$.

(iv) As observed by Harville (1977), the MLE is invariant for translation of Y by $X\beta$ for any β, i.e., the MLE is a function of the maximal invariant $T'Y$ of Y where $T = X^\perp$.

Suppose that θ in the model (5.1) is identifiable on the basis of the distribution of Y but not on the basis of the distribution of $T'Y$ as in the example of Focke and Dewess (1972). Such a situation arises when $T'V_iT$ are linearly dependent while V_i are not. In such cases the proper ML estimates (5.5) do not exist and ML equation estimates exhibit functional relationships which may not hold for true values of the parameters. Thus the invariance property of the MLE limits the scope of the application of the ML method. In such situations MINQE (U) or IMINQE (U) given by (3.17, 3.19) can be used.

(v) *Computational algorithms :* The equation (5.9) for the estimation of θ is, in general, very complicated and no closed form solution is possible.

One has to adopt iterative procedures. Harville (1977) has reviewed some of the existing methods in the special case of the mixed model. The following comments are valid in the general case :

(a) If V_i are linearly independent, then $H(\theta)$ is non-singular in which case the $(k+1)$-th approximation to the solution of (5.9) may be obtained using $\hat{\theta}_k$ the k-th approximation as

$$\hat{\theta}_{k+1} = [H(\hat{\theta}_k)]^{-1} h_l(Y, \hat{\theta}_k) \qquad \ldots \quad (5.12)$$

as in the case of IMINQE (I). Iterative procedure of the type (5.12) is mentioned by Anderson (1973), LaMotte (1973) and Rao (1972) in different contexts. However, it is not known whether the procedure converges and provides a solution at which the supremum of the likelihood is attained.

(b) Hartley and Rao (1967), Henderson (1977) and Harville (1977) proposed algorithms suitable for the special case when one of the V_i is an identity matrix (or at least non-singular). An extension of their method for the general case is to obtain the $(k+1)$-th approximation of the i-th component of θ as

$$\hat{\theta}_{i,k+1} = \hat{\theta}_{ik} \frac{Y'(I-P_{\hat{\theta}_k})' V_{\hat{\theta}_k}^{-1} V_i V_{\hat{\theta}_k}^{-1}(I-P_{\theta_k}) Y}{\operatorname{tr} V_{\theta_k}^{-1} V_i}, \quad i = 1, \ldots, p. \quad \ldots \quad (5.13)$$

In the special case when V_i are non-negative definite and the initial θ_i are chosen as non-negative, the successive approximations of θ_i using the algorithm (5.13) stay non-negative. This may be a "good property" of the algorithm, but it is not clear what happens when the likelihood equation (5.9) does not have a solution in the admissible region.

(c) Hemmerle and Hartley (1973) and Goodnight and Hemmerle (1978) developed the method of W transformation for solving the ML equations. Miller (1979) has given a different approach. Possibilities of using the variable-metric algorithms of Davidson-Fletcher-Powell described by Powell (1970) are mentioned by Harville (1977). The method of scoring described in (Rao, 1973, p. 366) can be used if the second order differentials of the likelihood can be easily computed. As it stands, further research is necessary for finding a satisfactory method of solving the equation (5.9) and ensuring that the solution provides a maximum of the likelihood.

5.2. *Marginal maximum likelihood (MML) equations.* As observed earlier the ML equation (5.9) is not unbiased, i.e.,

$$E[h_I(Y, \theta] \neq [H(\theta)]\theta. \quad \ldots \quad (5.14)$$

If we replace the equation (5.9) by

$$h_I(Y, \theta) = E[h_I(Y, \theta)] = [H_{UI}(\theta)]\theta \quad \ldots \quad (5.15)$$

we obtain the IMINQE (U, I) defined in (3.11).

The equation (5.15) is obtained by Patterson and Thompson (1975) by maximizing the marginal likelihood of θ based on $T'Y$, where T is any choice of $X^{\perp}$, which is the maximal invariant of Y. Now

$$l(\theta, T'Y) = -\log |T'V_\theta T| - Y'T(T'V_\theta T)^{-1} T'Y. \quad \ldots \quad (5.16)$$

Differentiating (5.16) w.r.t. θ_i we obtain the MML (marginal ML) equation

$$\text{tr}\,(T(T'V_\theta T)^{-1}T'V_i) = Y'T(T'V_\theta T)^{-1}T'V_iT(T'V_\theta T)^{-1}T'Y, \; i = 1, \ldots, p. \quad \ldots \quad (5.17)$$

Using the identity (Rao, 1973, p. 77)

$$T(T'V_\theta T)^{-1} T'$$

$$= V_\theta^{-1} - V_\theta^{-1}X(X'V_\theta^{-1}X)^{-1}X'V_\theta^{-1} = V_\theta^{-1}(I - P_\theta) \quad \ldots \quad (5.18)$$

the equation (5.17) becomes

$$\text{tr}\,(V_\theta^{-1}(I-P_\theta)V_i) = Y'V_\theta^{-1}(I-P_\theta)V_i(I-P_\theta')V_\theta^{-1}Y, \quad i = 1, \ldots, p \quad \ldots \quad (5.19)$$

which is independent of the choice of $T = X^{\perp}$ used in the construction of the maximal invariant of Y. It is easy to see that (5.19) can be written as

$$[H_{UI}(\theta)]\theta = h_I(Y, \theta) \quad \ldots \quad (5.20)$$

which is the equation (5.15).

(i) Both ML and MML estimates depend on the maximal invariant $T'Y$ of Y. Both the methods are not applicable when θ is not identifiable on the basis of $T'Y$.

(ii) The bias in MMLE may not be as heavy as in MLE and may be more useful as a point estimator.

(iii) The solution of (5.20) may not lie in the admissible set of θ as in the case of the ML equation.

(iv) If $\hat{\theta}_k$ is the k-th approximation, then the $(k+1)$-th approximation can be obtained as

$$\hat{\theta}_{k+1} = [H_{UI}(\hat{\theta}_k)]^{-1} h_I(Y, \hat{\theta}_k). \quad \dots \quad (5.21)$$

It is not known whether the proceess converges and yields a solution which maximizes the marginal likelihood.

(v) Another algorithm for MMLE similar to (5.13) is to compute the $(k+1)$-th approximation to the i-th component of θ as

$$\hat{\theta}_{i,\, k+1} = \hat{\theta}_{i,\, k} \frac{Y'(I-P'_{\hat{\theta}_k})V^{-1}_{\hat{\theta}_k} V_i V^{-}_{\hat{\theta}_k}(I-P_{\hat{\theta}_k})Y}{\text{tr } V^{-1}_{\hat{\theta}_k}(I-P_{\hat{\theta}_k})V_i}. \quad \dots \quad (5.22)$$

It is seen that both ML and MML estimators can be obtained as iterated MINQE's, MLE being IMINQE (I) defined in (4.11) and MMLE being IMINQE (U, I) defined in (3.11). There are other iterated MINQE's which can be used in cases where ML and MML methods are not applicable.

Note : The computation of MINQE's, MLE and MMLE as they are defined by various equations involve the inversion of $V_\theta = \theta_1 V_1 + \dots + \theta_p V_p$, an n-th order matrix, and the evaluation of matrices like $V^{-1}(I-P_\theta)$ where $P_\theta = X(X'V_{\theta}^{-1}X)^{-}X'V_{\theta}^{-1}$. These cannot be avoided in a general situation when the linear model is specified in the form (2.4, 2.5). However, the computations can be considerably simplified in special cases. An elegant method for computing V_θ^{-1} and terms like tr $V_\theta^{-1}V_iV_\theta^{-1}V_j$ which occur in the computation of MINQE's is given by Giesbrecht and Burrows (1978) in the case of hierarchical classification linear model. Thompson (1979) gives similar expressions for computing the MMLE from linear models arising out of orthogonal block designs.

In the case of a mixed model, V_θ is of the form

$$V_\theta = \theta_1 U_1 U_1' + \dots + \theta_{p-1} U_{p-1} U_{p-1}' + \theta_p I \quad \dots \quad (5.23)$$

in which case we can use the well-known formulae

$$\theta_p V_\theta^{-1} = I - U(U'U + G^{-1})^{-1}U' \quad \dots \quad (5.24)$$

$$\theta_\theta V_\theta^{-1}(I-P_\theta) = M - MU(U'MU + G^{-1})^{-1}U'M \quad \dots \quad (5.25)$$

where $M = I - X(X'X)^{-}X'$, $U = (U_1 : \dots : U_{p-1})$ and G is a block diagonal matrix with the i-th block equal to $\nu_i I_{n_i}$, where $\nu_i = \theta_i/\theta_p$ and n_i is the number of columns of U_i. (see Harville, 1977, Corbeil and Searle, 1976 and Rao, 1973).

Let us denote

$$(U'U + G^{-1})^{-1} = (T_{ij}), \quad \dots \quad (5.26)$$

$$(U'MU + G^{-1})^{-1} = (T^*_{ij}), \quad \dots \quad (5.27)$$

$$u_i = (T^*_{i1}U'_1M + \dots + T^*_{i\,p-1}U'_{p-1}M)Y.$$

Then the ML equations (5.7) can be written as

$$n\theta_p = Y'MY - Y'MU(T^*_{ij})U'MY \quad \dots \quad (5.28)$$

$$\theta_i\left(n_i - \frac{\text{tr } T_{ii}}{\nu_i}\right) = u'_i u_i,\ i = 1, \dots, p-1 \quad \dots \quad (5.29)$$

which are essentially the equations given by Henderson (1977) except for the method of computing u_i and the right-hand side expression of (5.28). The MML equations (5.10) are

$$(n-r)\,\theta_p = Y'MY - Y'MU(T^*_{ij})U'MY \quad \dots \quad (5.30)$$

$$\theta_i\left(n_i - \frac{\text{tr } T^*_{ii}}{\nu_i}\right) = u'_i u_i \quad \dots \quad (5.31)$$

where the right-hand side expressions are the same as those for ML equations as observed earlier.

Similarly, the MINQE equations for any apriori value α of θ can be set up using the expressions (5.24) and (5.25).

References

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

Corbeil, R. R. and Searle, S. R. (1976): Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, 18, 31-38.

Focke, J. and Dewess, G. (1972): Uber die Schatzmethode MINQUE von C. R. Rao and ihre Verallgemeinerung. *Math. Operationforsch. u. Statist.*, 3, 129-143.

Giesbrecht, Francis, G. and Burrows, Peter, M. (1978): Estimating variance components in hierarchical structures using MINQUE and restricted maximum likelihood. *Comm. Statist. Theor. Meth.*, A7(9), 891-904.

Goodnight, J. H. and Hemmerle, W. J. (1978): A simplified algorithm for the W-transformation in variance component estimation. SAS Tech. Rept. R-104, Raleigh, N.C.

Hartley, H. O. and Rao, J. N. K. (1957): Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54, 93-108.

HARVILLE, DAVID A. (1977): Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.*, **72**, 320-340.

HEMMERLE, W. J. and HARTLEY, H. O. (1973): Computing maximum likelihood estimates for the mixed AOV model using the *W*-transformation. *Technometrics*, **15**, 819-831.

HENDERSON, C. R. (1977): Prediction of future records. *Proc. Int. Conf. on Quantitative Genetics*, 616-638.

KLEFFE, J. (1975): Quadratische Bayes-Schatzungen fur Lineare Parameter Der-Kovarianzmatrix imm Gemishten Linearen Modellen. Dissertation, Humboldt Univ., Berlin.

———— (1976): Best quadratic unbiased estimators for variance components in mixed linear models. *Sankhyā*, B, **38**, 179-186.

———— (1977a): Invariant methods for estimating variance components in mixed linear models. *Math. Operationforsch. Statist.*, **8**, 233-250.

———— (1977b): A note on ∞-MINQUE in variance convariance components models. *Math. Operationforsch Statist.*, **8**, 337-343.

———— (1978): Simultaneous estimation of expectation and covariance matrix in linear models. *Math. Oper. Statist. Ser. Statistics*, **9**, 443-478.

———— (1979): C. R. Rao's MINQUE for replicated and multivariate observations. Tech. Rept. Zimm der AdW der DDR, Berlin.

KRISHNAIAH, P. R. and LEE, JACK C. (1974): On covariance structures. *Sankhyā*, **33A**, 357-371.

LAMOTTE, LYNN R. (1973): Quadratic estimation of variance components. *Biometrics*, **29**, 311-330.

MILLER, J. J. (1977): Asymptotic properties of maximum likelihood estimates in the mixed model of analysis of variance. *Ann. Statist.*, **5**, 746-762.

———— (1979): Maximum likelihood estimation of variance components—a Monte Carlo Study. *J. Stat. Comp. and Simulation*, **8**, 175-190.

PATTERSON, H. D. and THOMPSON, R. (1975): Maximum likelihood estimation of components of variance. *Proc. of 8-th International Biometric Conference*, 197-207.

POWELL, M. J. D. (1970): A survey of numerical methods for unconstrained optimization. *SIAM Review*, **12**, 79-97.

PUKELSHEIM, F. (1977): Linear models and convex programs: Unbiased non-negative estimation in variance components. *Tech. Report* 104, Stanford University.

RAO, P. S. R. S. and CHAUBEY, Y. P. (1978): Three modifications of the principle of the MINQUE. *Comm. Statist. Meth.*, **A7**, 767-778.

RAO, C. RADHAKRISHNA (1970): Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.*, **67**, 161-172.

———— (1971a): Estimation of variance and covariance components. *J. Multivar. Anal.*, **1**, 257-275.

———— (1971b): Minimum variance quadratic unbiased estimation of variance components. *J. Multivar. Anal.*, **1**, 445-456.

———— (1972): Estimation of variance and covariance components in linear models. *J. Amer. Stat. Assoc.*, **67**, 112-115.

———— (1973): *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley, New York.

RAO, C. RADHAKRISHNA and KLEFFE, J. (1979): Estimation of variance components. *Tech. Rept.* 79-1. University of Pittsburgh.

SINHA, B. K. and WIEAND, H. S. (1977): MINQUE'S of variance and covariance components of certain covariance structures. Indian Statistical Institute, *Tech. Rept.* 28/77.

THOMPSON, R. (1979): Maximum likelihood estimation of variance components. Reprint. Akademie Der Wissenshaften Der DDR., Berlin.

Paper received: October, 1979.