Tatra Mt. Math. Publ. **17** (1999), 111–119

TATRA MOUNTAINS
Mathematical Publications

ON MAXIMUM LIKELIHOOD ESTIMATORS IN CERTAIN MULTIVARIATE NORMAL MIXED MODELS WITH TWO VARIANCE COMPONENTS

Stanisław Gnot — Dietmar Stemann —
Götz Trenkler — Agnieszka Urbańska-Motyka

ABSTRACT. We consider the problem of maximum likelihood (ML) estimation of variances in certain mixed linear models with two variance components. For those models a procedure to obtain ML estimators is described. The solutions produced by this procedure are described. It is shown that under some conditions the solutions of the ML equations cannot give ML estimators, i.e., estimators which maximize the loglikelihood function. For those models necessary and sufficient conditions are given, for which the solutions of the ML equations lead to ML estimators. Two way layouts model and equicorrelated model are considered in details.

1. Introduction

Consider a linear model $N\{y, X\beta, \Sigma(s) = \sigma_1^2 V + \sigma_2^2 I_n\}$, in which y is an $n \times 1$ normally distributed random vector with

$$E(y) = X\beta\,, \quad \operatorname{Cov}(y) = \Sigma(s) = \sigma_1^2 V + \sigma_2^2 I_n\,.$$

Here X is a known $n \times p$ matrix of rank p $(p < n)$, β is an unknown $p \times 1$ vector of fixed parameters, V is a known nonnegative definite matrix, I_n is the $n \times n$ identity matrix, while $s = (\sigma_1^2, \sigma_2^2)'$ is a vector of unknown variance components, $s \in \mathcal{S} = \{s = (\sigma_1^2, \sigma_2^2)' \colon \sigma_1^2 \geq 0,\ \sigma_2^2 > 0\}$.

The main problem considered in this paper is to find maximum likelihood (ML) estimators for β and s, i.e., to find $\hat\beta$ and $\hat s$ such that

$$(\hat\beta\,, \hat s) = \arg\max l(\beta, s, y)\,,$$

where $l(\beta, s, y)$ is the loglikelihood function of y given by

$$(\beta, s, y) = -\frac{1}{2}\,n\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}Q(y, \beta, \Sigma)\,, \tag{1}$$

2000 Mathematics Subject Classification: 62J1O.
Keywords: mixed linear model, variance components, ML-estimators.

with $Q(y,\beta,\Sigma) = (y - X\beta)'\Sigma^{-1}(y - X\beta)$. A general principle for maximizing $l(\beta, s, y)$, or equivalently

$$l_0(\beta, s, y) = -\log|\Sigma| - Q(y, \beta, \Sigma),$$

with respect to β and s is to equate $\frac{dl_0}{d\beta}$ and $\frac{dl_0}{ds}$ to zero and to solve the resulting equations for β and s. Since for

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

and for

$$Q_0 = (y - X\hat{\beta}_\Sigma)'\Sigma^{-1}(y - X\hat{\beta}_\Sigma)$$

the following equality holds,

$$Q(y, X\beta, \Sigma) = Q_0 + (\hat{\beta}_\Sigma - \beta)'X'\Sigma^{-1}X(\hat{\beta}_\Sigma - \beta), \tag{2}$$

the maximum of $l_0(\beta, s, y)$ with respect to β for a given s and y is attained at $\hat{\beta}_\Sigma$. This maximal value of $l_0(\beta, s, y)$ is

$$l_1(s, y) = -\log|\Sigma| - y'(M\Sigma M)^+ y \tag{3}$$

(see Rao, Kleffe 1988, p. 230). Here $M = I - P$, while $P = X(X'X)^{-1}X'$ is the orthogonal projection on $R(X)$. Thus the maximization of (3) determines $\hat{s}$. The value $\hat{s}$ that maximizes l_1 is a solution of $\frac{\partial l_1(s,y)}{\partial s} = 0$. But these solutions are not necessarily the ML estimators of s. In this paper we look for the solutions resulting from the equation system $\frac{\partial l_1}{\partial s} = 0$, and check which of them are ML-estimators, i.e., which of them maximize l_1, not necessarily in S.

Remark 1.1. Throughout the paper we assume that $MVM \neq 0$, i.e., that $(M\Sigma M)^+ \neq M$. This assumption is necessary to avoid trivialities and contradictions.

Generally there are no analytical expressions for the ML estimators, but some iterative procedures to obtain a solution of the ML equation system are proposed in the literature (cf. Searle *et al.*, [4], Section 8). As it has been proved by Szatrowski [5] and Szatrowski and Miller [6], for a special case of the model, the likelihood equation system becomes linear, and ML estimators of β and s can be presented in an explicit form. The conditions are as follows

$$PV = VP, \tag{4}$$

$$\mathcal{V} = \mathrm{sp}\{V, I\} \quad \text{is a Jordan algebra}, \tag{5}$$

i.e., $V \in \mathcal{V}$ implies that $V^2 \in \mathcal{V}$. It is very easy to check that if V is singular, then (5) is equivalent to $V^2 = \alpha V$ for some positive α.

Lemma 1.1. *If V is such that $V^2 = \alpha V$ for some positive scalar α, then the condition $PV = VP$ holds iff $(MVM)^2 = \alpha MVM$.*

Proof. It is clear that the conditions $V^2 = \alpha V$ and $PV = VP$ imply $(MVM)^2 = \alpha MVM$. Now let $(MVM)^2 = \alpha MVM$. Then we have

$$\begin{aligned} MV^2M &= MV(P+M)VM \\ &= MVPVM + MVMVM = \alpha MVM . \end{aligned}$$

On the other hand, $MVMVM = (MVM)^2 = \alpha MVM$. It follows that $MVPVM = 0$. In consequence

$$MVP = PVM = 0 ,$$

and since $V = PVP + MVP + PVM + MVM$ we have $V = PVP + MVM$, and finally $VP = PV$.

Two examples given below show that the above conditions assuring the existence of the ML estimators in an explicit form are very restrictive. □

2. Examples

2.1 Two-way layout model.

Let us consider a two-way layout in which n experimental units are arranged in v rows and b columns according to the $v \times b$ incidence matrix N with entries $n_{ij} \geq 0$. The vector of row sums and the vector of column sums of the layout are $r = N1_b$ and $k = N'1_v$, respectively, where 1_z denotes the $z \times 1$ vector of ones. It is assumed that a linear model related to a given layout is the mixed model with random row effects, with fixed column effects, and without interactions. We can represent this model as

$$N\{y, X\beta, \Sigma(s) = \sigma_1^2 V + \sigma_2^2 I_n\} , \tag{6}$$

where $V = DD'$, while D and X are $n \times v$ and $n \times b$ known $(0,1)$-matrices of full column ranks v and b. It is clear that $D'X = N$, $X1_b = D1_v = 1_n$, while $X'X = \operatorname{diag}\{k_i\}$ and $D'D = \operatorname{diag}\{r_j\}$ are diagonal matrices with the diagonal elements $r_i = \sum_j n_{ij}$ and $k_j = \sum_i n_{ij}$, respectively ($r = (r_1, \dots, r_v)'$, $k = (k_1, \dots, k_b)'$). To characterize layouts for which (4) and (5) are fulfilled we assume that the layout under consideration is connected, i.e., $R(X) \cap R(D) = R(1_n)$. For general results see Kageyama and Zmyślony [2].

Proposition 2.1. *In the connected two-way layout under the assumption that $n > v + b - 1$ (V is singular), (4) and (5) are fulfilled iff $N = \frac{1}{v} 1_v k'$.*

Proof. Note first that (4) is equivalent to $R(DN) \subset R(1_n)$. Since $R(N) \subset R(D')$, it follows that rank $(N) = 1$, i.e., that $N = \frac{1}{n} rk'$. From (5) we find that

$V = DD'$ or, equivalently, $D'D = \operatorname{diag}\{r_i\}$ has only one positive eigenvalue. It means that all r_i are the same, and equal to α.

As we can see, the conditions (4) and (5) are fulfilled iff $N = \frac{1}{v}1_v k'$, i.e., iff the layout under consideration is orthogonal and equireplicated (see Caliński, 1993).

Since the positive eigenvalues of $MVM = MDD'M$ are the same as those of $C = D'MD = \alpha I_v - N \operatorname{diag}\{\frac{1}{k_i}\}N'$, for the orthogonal and equireplicated layout the matrix C has the following form

$$C = \alpha\Big(I_v - \frac{1}{v}1_v 1_v'\Big).$$

A larger class of equireplicated and connected layouts can be obtained under the assumption that

$$C = m\Big(I_v - \frac{1}{v}1_v 1_v'\Big), \tag{7}$$

with $m \leq \alpha$. The layouts for which C has the form (7) are called totally balanced (see Caliński, 1993). Unfortunately, following Lemma 1.1 for the totally balanced layouts with $m < \alpha$, explicit forms of ML estimators of σ_1^2 and σ_2^2 do not exist. $\square$

2.2 Equicorrelated model.

Consider a linear regression model of the form:

$$y = K\beta + u\,, \quad E(u) = 0\,, \quad \operatorname{Cov}(y) = \sigma^2 W, \tag{8}$$

where y is an $n \times 1$ vector of observations, K is an $n \times p$ nonstochastic matrix of observations of explanatory variables (regressor matrix), β is an unknown $p \times 1$ vector of parameters, while u is an unobservable vector of disturbances. Let $\operatorname{rank}(K) = p$, $0 < p < n$. For W we assume an equicorrelated structure, i.e.,

$$W = \rho 1_n 1_n' + (1-\rho)I_n\,, \tag{9}$$

where ρ is an unknown correlation coefficient and 1_n is the $n \times 1$ vector of units. The problem is to estimate ρ and σ^2 from the independent sequence of observed vectors $y^{(1)}$, $y^{(2)}, \ldots, y^{(J)}$, each of them having the normal distribution with the covariance matrix W given by (9).

Let $Y = \operatorname{vec}\big(y^{(1)}, \ldots, y^{(J)}\big)$. It can be easily found out that Y has a normal distribution with

$$E(Y) = X\beta = (1_J \otimes K)\beta\,, \tag{10}$$

$$\operatorname{Cov}(Y) = \Sigma = \sigma^2\rho V + \sigma^2(1-\rho)I_N\,, \tag{11}$$

where $V = I_J \otimes 1_n 1_n'$, $N = n \cdot J$, while for arbitrary matrices A, B the symbol $A \otimes B$ denotes the Kronecker product of A and B. In this model $V^2 = nV$,

$P = \frac{1}{J}(1_J 1'_J \otimes P_K)$, where $P_K = K(K'K)^{-1}K'$. Moreover $MVM = (I_J \otimes 1_n 1'_n) - \frac{1}{J}(1_J 1'_J \otimes 1_n 1'_n P_K) - \frac{1}{J}(1_J 1'_J \otimes P_K 1_n 1'_n) + \frac{1}{J}(l_J l'_J \otimes P_K 1_n 1'_n PK)$. It is easy to check that $PV = VP$ iff $P_K 1_n 1'_n = P_K 1_n 1'_n$, i.e., iff $1_n \in R(K)$ or $1_n \in N(K')$. Let $a = 1'_n P_K 1_n$. Then since

$$(MVM)^2 = (n-a)MVM - a\Big[\Big(\frac{1}{J}1_J 1'_J \otimes 1_n 1'_n\Big) - (I_J \otimes 1_n 1'_n)\Big],$$

it follows that if $a = 0$ $\big(1_n \in N(K')\big)$ or if $J = 1$, then $(MVM)^2 = mMVM$ with $m = n - a$. As it has been mentioned earlier, in the considered model the condition (5) is always satisfied since $V^2 = nV$. The above considerations lead us to the following proposition.

Proposition 2.2. *In equicorrelated model the condition of Lemma 1.1 is fulfilled iff $1_n \in R(K)$ or $1_n \in N(K')$.*

Remark 2.1. Note that if $J = 1$, then since $(MVM)^2 = (n-a)MVM$, a has to be less then n (see Remark 1.1). In such a case we have the same situation as in totally balanced two-way layout (see Section 2.1).

3. ML estimators under conditions $V^2 = \alpha V$ and $(MVM)^2 = mMVM$

In this section we assume that in the considered model $N\{y, X\beta, \Sigma(s) = \sigma_1^2 V + \sigma_2^2 I_n\}$ for some positive α and m, the following conditions hold

$$V^2 = \alpha V \quad \text{and} \quad (MVM)^2 = mMVM. \tag{12}$$

Proposition 3.1. *If for some positive scalars α and m greater than one $V^2 = \alpha V$ and $(MVM)^2 = mMVM$, then $m \le \alpha$.*

P r o o f. Let us introduce the following parametrization:

$$\sigma^2 = \sigma_1^2 + \sigma_2^2, \qquad \rho = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

Then

$$\sigma_1^2 = \sigma^2 \rho, \qquad \sigma_2^2 = \sigma^2(1-\rho).$$

First note that $\Sigma = \sigma^2\big[\rho V + (1-\rho)I_n\big]$. Since $\Sigma > 0$ implies $M\Sigma M \ge 0$, it follows that under the assumptions $V^2 = \alpha V$ and $(MVM)^2 = mMVM$, the matrix $\Sigma > 0$ iff $\rho > \frac{-1}{\alpha - 1}$, while $M\Sigma M \ge 0$ iff $\rho \ge \frac{-1}{m-1}$. From the above we find that $-\frac{1}{\alpha-1} \ge -\frac{1}{m-1}$, and in consequence $m \le \alpha$.

In the rest of the paper we assume that $m < \alpha$. Under (12), $\mathrm{sp}\{MVM, M\}$ is a Jordan algebra, and

$$(M\Sigma M)^{+} = \lambda_1 MVM + \lambda_2 M\,, \tag{13}$$

where $\lambda_1 = -\frac{\sigma_1^2}{\sigma_2^2(m\sigma_1^2+\sigma_2^2)}$, while $\lambda_2 = \frac{1}{\sigma_2^2}$. Using (13) we find that $\det(\Sigma) = (\alpha\sigma_1^2 + \sigma_2^2)^{\nu}\sigma_2^{2(n-\nu)}$, where $\nu = \mathrm{rank}(V)$, while n is the total number of observations. It follows that

$$l_1 = -\nu\log(\alpha\sigma_1^2 + \sigma_2^2) - (n-\nu)\log(\sigma_2^2) - \lambda_1 y'MVMy - \lambda_2 y'My \tag{14}$$

and

$$\frac{\partial l_1}{\partial \sigma_1^2} = -\frac{\nu\alpha}{\alpha\sigma_1^2 + \sigma_2^2} + \frac{1}{(m\sigma_1^2 + \sigma_2^2)^2}\, y'MVMy\,, \tag{15}$$

$$\frac{\partial l_1}{\partial \sigma_2^2} = -\frac{\nu}{\alpha\sigma_1^2 + \sigma_2^2} - \frac{n-\nu}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma_2^4(m\sigma_1^2 + \sigma_2^2)}\, y'MVMy + \frac{1}{\sigma_2^4}\, y'My\,. \tag{16}$$

Equating (15) to zero we find that

$$y'MVMy = \frac{\nu\alpha(m\sigma_1^2 + \sigma_2^2)^2}{\alpha\sigma_1^2 + \sigma_2^2}\,. \tag{17}$$

Putting (17) to (16) and equating (16) to zero we find that

$$y'My = \frac{\nu\alpha\sigma_1^2(m\sigma_1^2 + \sigma_2^2)}{\alpha\sigma_1^2 + \sigma_2^2} + n\sigma_2^2\,. \tag{18}$$

The equation system (17) and (18) can be expressed in terms of sufficient and independent statistics $T_1 = \frac{1}{m\nu_1} y'MVMy$ and $T_2 = \frac{1}{\nu_2}(y'My - \nu_1 T_1)$ in the following way

$$\nu_1 T_1 = \frac{\nu\alpha(m\sigma_1^2 + \sigma_2^2)^2}{m(\alpha\sigma_1^2 + \sigma_2^2)}\,, \tag{19}$$

$$\nu_2 T_2 = \frac{\nu\alpha\sigma_1^2(m\sigma_1^2 + \sigma_2^2)}{\alpha\sigma_1^2 + \sigma_2^2} - \frac{\nu\alpha(m\sigma_1^2 + \sigma_2^2)^2}{m(\alpha\sigma_1^2 + \sigma_2^2)} + n\sigma_2^2\,, \tag{20}$$

where $\nu_1 = \mathrm{rank}(MVM)$, while $\nu_2 = n - p - \nu_1$.

For a given scalar c, let us denote $\phi_c(\rho) = (c-1)\rho + 1$. Note that $c\sigma_1^2 + \sigma_2^2 = \sigma^2\phi_c(\rho)$. Finally, (19) and (20) reduce to

$$\frac{\nu_1 m}{\nu\alpha}\, T_1 = \frac{\phi_m^2(\rho)}{\phi_\alpha(\rho)}\,\sigma^2\,, \tag{21}$$

$$m\nu_2 T_2 = \frac{\sigma^2\big(m\nu\alpha\rho\phi_m(\rho) - \nu\alpha\phi_m^2(\rho) + nm(1-\rho)\phi_\alpha(\rho)\big)}{\phi_\alpha(\rho)}\,. \tag{22}$$

Dividing (21) by (22) we get

$$\frac{\phi_m^2(\rho)}{(1-\rho)\big[nm\phi_\alpha(\rho)-\nu\alpha\phi_m(\rho)\big]}=Q\,, \tag{23}$$

where $Q=\frac{\nu_1 T_1}{\nu\nu_2\alpha T_2}$. Let $t=\frac{m\rho}{1-\rho}$. Then $\rho=\frac{t}{m+t}$. There is one to one correspondence between t and ρ, and if ρ runs through the interval $(-\frac{1}{\alpha-1},1)$ (if Σ is positive definite), then t runs through the interval $(-\frac{m}{\alpha},+\infty)$. Denote

$$f(t)=\frac{(1+t)^2}{(n-\nu)\alpha t+nm-\nu\alpha}\,. \tag{24}$$

It is easy to check that (23) is equivalent to

$$f(t)=Q\,, \tag{25}$$

while from (21) we have

$$\sigma^2=\frac{m\nu_1\phi_\alpha(\rho)}{\nu\alpha\phi_m^2(\rho)}\,T_1\,. \tag{26}$$

The problem arises, under which conditions equation (25) has at least one solution. The Proposition presented below answers this question. □

PROPOSITION 3.2. *The likelihood equation* (25) *has*

(i) *two solutions iff* $Q>c$,

(ii) *one solution iff* $Q=c$,

(iii) *no solution iff* $0<Q<c$,

where

$$c=\frac{4n(\alpha-m)}{\alpha^2(n-v)^2}\,.$$

Note that since $\alpha>m$ and $n>v$, the constant c is always positive.

Suppose that Q satisfies (i). Let $\hat t(1)$ and $\hat t(2)$ be two solutions of (25) ($\hat t(1)<\hat t(2)$) and let $\hat{\sigma}^2(1)$ and $\hat{\sigma}^2(2)$ be the estimators of σ^2, taken from (26) at $\hat t(1)$ and $\hat t(2)$, respectively. Then the following lemma can be proved.

LEMMA 3.1. *The loglikelihood function l_1 as a function of t and σ^2 has its saddlepoint at $\big(\hat t(1),\hat{\sigma}^2(1)\big)$, and achieves its global maximum at $\big(\hat t(2),\hat{\sigma}^2(2)\big)$.*

To prove the lemma it is sufficient to calculate the second derivatives of l_1 with respect to σ_1^2 and σ_2^2 to establish definiteness of the Hessian

$$H=\begin{pmatrix}\frac{\partial l_1^2}{\partial\sigma_1^4} & \frac{\partial l_1^2}{\partial\sigma_1^2\sigma_2^2}\\ \frac{\partial l_1^2}{\partial\sigma_1^2\sigma_2^2} & \frac{\partial l_1^2}{\partial\sigma_2^4}\end{pmatrix}$$

(for general considerations see Searle *et al.* [4], Section 6.3). After arduous though straightforward calculations we find that

$$\det(H) = \frac{(n-\nu)\nu^5\alpha^5 m^2(t+1)^7\left[\alpha(n-\nu)t - \nu\alpha + 2nm - n\alpha\right]}{\nu_1^4(\alpha t + m)^6 T_1^4}\,.$$

Besides

$$\frac{\partial l_1^2}{\partial \sigma_1^4} = \frac{\nu^3\alpha^3 m^2(t+1)^3(\alpha - \alpha t - 2m)}{\nu_1(\alpha t + m)^4 T_1^2}\,.$$

Let

$$t_0 = \frac{\nu\alpha - 2mn + n\alpha}{\alpha(n-\nu)}\,.$$

The following relations hold:

(i) the solutions $\hat{t}_1$ and $\hat{t}_2$ of the ML equation (25) satisfy the following inequality: $\hat{t}_1 < t_0 < \hat{t}_2$,

(ii) $\det(H) < 0$ for $t < t_0$ and $\det(H) > 0$ for $t > t_0$,

(iii) $\frac{\partial l_1^2}{\partial \sigma_1^4} < 0$ for $\frac{\alpha - m}{\alpha} < t_0$.

It follows that the Hessian H is negative definite at $\hat{t}_2$ and indefinite at $\hat{t}_1$. From

$$\sigma^2 = \sigma_1^2 + \sigma_2^2\,, \qquad \rho = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{t}{m+t} \qquad \left(t = \frac{m\sigma_1^2}{\sigma_2^2}\right)$$

we have

$$\sigma_1^2 = \sigma^2\rho = \frac{\sigma^2 t}{m+t}\,, \qquad \sigma_2^2 = \sigma^2(1-\rho) = \frac{\sigma^2 m}{m+t}\,.$$

Let

$$\hat{\sigma}_1^2(i) = \frac{\hat{\sigma}^2(i)\hat{t}(i)}{m + \hat{t}(i)}\,, \qquad i = 1,2\,, \tag{27}$$

$$\hat{\sigma}_2^2(i) = \frac{\hat{\sigma}^2(i)m}{m + \hat{t}(i)}\,, \qquad i = 1,2\,. \tag{28}$$

Since there is one to one correspondence between (σ_1^2, σ_2^2) and (σ^2, t), Lemma 3.1 leads to the following corollary.

COROLLARY 3.1. *The loglikelihood function l_1 as a function of σ_1^2 and σ_2^2 has its saddlepoint at $\left(\hat{\sigma}_1^2(1), \hat{\sigma}_2^2(1)\right)$, and achieves its global maximum at $\left(\hat{\sigma}_1^2(2), \hat{\sigma}_2^2(2)\right)$, where $\hat{\sigma}_1^2(i)$ and $\hat{\sigma}_2^2(i)$ are defined by (27) and (28).*

REFERENCES

[1] CALIŃSKI, T.: *Balance, efficiently and orthogonality concepts in block designs*, J. Statist. Plann. Inference **36** (1993), 283–300.

[2] KAGEYAMA, S.—ZMYŚLONY, R.: *Characterization and optimality of block designs with estimation of parameters under mixed model*, in: Statistical Sciences and Data Analysis, Proceedings of the Third Pacific Area Statistical Conference, 1993, pp. 559–570.

[3] RAO, C. R.—KLEFFE, J.: *Estimation of Variance Components and Applications*, Elsevier Science Publications, B. V., 1988.

[4] SEARLE, S. R.—CASELLA, G.—Mc CULLOCH, CH. E.: *Variance Components*, John Wiley & Sons, New York, 1992.

[5] SZATROWSKI, T. H.: *Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances*, Ann. Statist. **8** (1980), 803–810.

[6] SZATROWSKI, T. H.—MILLER, J. J.: *Explicit maximum likelihood estimates from balanced data in the mixed model of the analysis of variance*, Ann. Statist. **8** (1980), 811–819.

Received May 25, 1998

Stanisław Gnot,
Agnieszka Urbańska-Motyka
Institute of Mathematics
Pedagogical University
Plac Słowiański 9
PL–65–069 Zielona Góra
POLAND
E-mail: gnot@omega.im.wsp.zgora.pl

Dietmar Stemann
Department of Statistics
University of Hagen
Postfach 940, Roggenkamp 6
D–58009 Hagen
GERMANY
E-mail: Dietmar.Stemann@FernUni-Hagen.de

Götz Trenkler
Department of Statistics
University of Dortmund
D–44221 Dortmund
GERMANY