MAXIMUM LIKELIHOOD ESTIMATES IN THE MULTIVARIATE NORMAL WITH PATTERNED MEAN AND COVARIANCE VIA THE EM ALGORITHM

by

Dalton F. Andrade

Department of Quantitative Methods
Embrapa, Brasilia, DF, Brasil

and

Ronald W. Helms

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1445

MAXIMUM LIKELIHOOD ESTIMATES IN THE MULTIVARIATE NORMAL WITH PATTERNED MEAN AND COVARIANCE VIA THE EM ALGORITHM

By Dalton F. Andrade

Department of Quantitative Methods,

Embrapa, Brasilia, DF, Brasil.

and Ronald W. Helms

Department of Biostatistics,

University of North Carolina,

Chapel Hill, North Carolina, U.S.A.

SUMMARY

The maximum likelihood equations for a multivariate normal model with structured mean and structured covariance matrix may not have an explicit solution. In some cases the model's error term may be decomposed as the sum of two independent error terms, each having a patterned covariance matrix, such that if one of the unobservable error terms is artificially treated as "missing data", the EM algorithm can be used to compute the maximum likelihood estimates for the original problem. Some decompositions produce likelihood equations which do not have an explicit solution at each iteration of the EM algorithm, but within-iteration explicit solutions are shown for two general classes of models including covariance component models used for analysis of longitudinal data.

Some Key words: Covariance components, EM algorithm, linear models, longitudinal data analysis, maximum likelihood, mixed models, random effects, repeated measures, two-stage models.

1. INTRODUCTION

There are many situations in which one is willing to assume that data represent a sample of N independent, identically distributed random vectors y_i, from a p-variate normal distribution with structured mean $\mu = X\beta$ and a covariance matrix Σ_0 with linear pattern $\Sigma_0 = \sum_{g=1}^{m} \tau_g G_g$ (Anderson, 1973), where the G_g are known, constant, p x p matrices which are linearly independent, i.e., $\Sigma_g a_g G_g = 0$ implies all $a_g = 0$. The parameters β and $\tau = (\tau_1, \ldots, \tau_m)^T$ are assumed to be unknown constants and X is a p x q known matrix of rank $q \leq p$. The vector y_i can be represented by the linear model

$$y_i = X\beta + \eta_i, \quad i = 1, 2, \ldots, N \tag{1}$$

where $V(\eta_i) = \Sigma_0$.

The maximum likelihood estimators of β and τ are the solutions of these likelihood equations:

$$\hat{\beta} = (X^T \hat{\Sigma}_0^{-1} X)^{-1} X^T \hat{\Sigma}_0^{-1} \bar{y} \tag{2}$$

$$\hat{\tau} = [(Tr\{\hat{\Sigma}_0^{-1} G_g \hat{\Sigma}_0^{-1} G_h\})_{gh}]^{-1} [(Tr\{\hat{\Sigma}_0^{-1} G_g \hat{\Sigma}_0^{-1} S\})_g] \tag{3}$$

$$g,h = 1, \ldots m$$

where:

$$\hat{\Sigma}_0 = \Sigma_g \hat{\tau}_g G_g \tag{4}$$

$$S = \Sigma_i (y_i - X\hat{\beta})(y_i - X\hat{\beta})^T / N \tag{5}$$

$\bar{y} = (\Sigma_i y_i)/N$, $[(\ )_{gh}]$ denotes a matrix with (g,h)--element displayed in the parentheses, and $[(\ )_g]$ is similar notation for a column vector.

In general the likelihood equations (2,3) do not have an explicit solution and must be solved by iterative techniques such as Fisher's method of scoring or the Newton-Raphson method (Anderson, 1970,1973), neither of which is guaranteed to converge. Szatrowski (1980) gives necessary and sufficient conditions on the structure of X and Σ_0 such that there exist explicit, noniterative solutions to the likelihood. We shall use the phrase "S-explicit representation" to represent Szatrowski's (1980) rigorous defintion of "an estimator has explicit representation" under the assumption that Σ_0 has "normal form". In the present context, $\hat{\beta}$ has S-explicit representation if and only if changing $\hat{\Sigma}_0$ to I in (2) does not change the value of the expression on the right hand side of (2). $\hat{\Sigma}_0$ (or, equivalently, $\hat{\tau}$) has S-explicit representation if and only if (a) $\hat{\beta}$ has S-explicit representation and (b) changing $\hat{\Sigma}_0$ to I in (3) does not change the value of the right hand side of (3). Clearly, if both $\hat{\beta}$ and $\hat{\Sigma}_0$ have S-explicit representation the likelihood equations have an explicit noniterative solution.

The EM algorithm (Dempster, Laird, and Rubin, 1977), an iterative algorithm for maximum likelihood estimation for "incomplete data" situations, is usually not directly applicable to the "complete data" model (1). However, Dempster, Rubin, and Tsutakawa (1981), and Laird and Ware (1982), have used the idea of letting an unobservable random term in a mixed linear model be "missing data". The EM algorithm is then applied to this artificial incomplete data problem to compute maximum likelihood estimates of parameters in the original problem. Rubin and Szatrowski (1982), used a similar strategy on a model similar to (1) but with no structure in $\underset{\sim}{\mu}$. In the next section we show how this strategy may be applied to the model (1) by first inventing partitions of the error term and covariance matrix.

2. APPLICATION OF THE EM ALGORITHM.

Consider the situation in which Σ_0 can be written as the sum of two patterned matrices, $\Sigma_0 = \Sigma_1 + \Sigma_2$, where each Σ_j has linear pattern, $\Sigma_j = \Sigma_k \tau_{jk} G_{jk}$, where $\tau_j = (\tau_{j1}, \ldots, \tau_{jm_j})^T$ and for fixed j the G_{jk} are linearly independent. Let δ_i and ε_i be artificial random vectors such that $Var(\delta_i) = \Sigma_1$, $Var(\varepsilon_i) = \Sigma_2$, and the random term in (1) has the structure $\eta_i = \delta_i + \varepsilon_i$, $i = 1, 2, \ldots, N$. Although the vectors δ_i are unobservable, for the purpose of applying the EM algorithm let $\{y_i, \delta_i\}_1^N$ denote the "complete data" and $\{y_i\}_1^N$ denote the observed, "incomplete data". In this context the complete-data sufficient statistic is

$$(\Sigma_i y_i,\ \Sigma_i y_i y_i^T,\ \Sigma_i y_i \delta_i^T,\ \Sigma_i \delta_i,\ \Sigma_i \delta_i \delta_i^T).$$

The r-th iteration of the EM algorithm has two steps, the "E step" and the "M step". The details of these steps for this problem are given in the following paragraphs.

2.1 E Step

At the r-th iteration, the "E step" is the computation of the conditional expectation of the complete-data sufficient statistic given: (1) the observed data, $y_1, \ldots, y_N$, and (2) the estimated values of the parameters from the r-1 iteration. In this context the equations are:

$$E[\Sigma_i \underset{\sim}{y}_i \underset{\sim}{y}_i^T \mid \underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N; \underset{\sim}{\beta}^{(r-1)}, \underset{\sim 1}{\Sigma}^{(r-1)}, \underset{\sim 2}{\Sigma}^{(r-1)}] = \Sigma_i \underset{\sim}{y}_i \underset{\sim}{y}_i^T \quad (6)$$

$$E[\Sigma_i \underset{\sim}{y}_i \mid \#] = \Sigma_i \underset{\sim}{y}_i \quad (7)$$

$$E[\Sigma_i \underset{\sim}{y}_i \underset{\sim}{\delta}_i^T \mid \#] = \Sigma_i \underset{\sim}{y}_i \underset{\sim}{\delta}_i^{(r)^T} \quad (8)$$

$$E[\Sigma_i \underset{\sim}{\delta}_i \mid \#] = \Sigma_i \underset{\sim}{\delta}_i^{(r)} \quad (9)$$

$$E[\Sigma_i \underset{\sim}{\delta}_i \underset{\sim}{\delta}_i^T \mid \#] = N\, \underset{\sim}{V}^{(r)} + \Sigma_i \underset{\sim}{\delta}_i^{(r)} \underset{\sim}{\delta}_i^{(r)^T} \quad (10)$$

where the superscript (r) denotes the value of a parameter estimator computed in the r-th iteration, # is a typographical abbreviation for

$$"\underset{\sim}{y}_1, \ldots, \underset{\sim}{y}_N; \underset{\sim}{\beta}^{(r-1)}, \underset{\sim 1}{\Sigma}^{(r-1)}, \underset{\sim 2}{\Sigma}^{(r-1)}",$$

$$\underset{\sim}{\delta}_i^{(r)} = \underset{\sim 1}{\Sigma}^{(r-1)}[\underset{\sim 1}{\Sigma}^{(r-1)} + \underset{\sim 2}{\Sigma}^{(r-1)}]^{-1}[\underset{\sim}{y}_i - \underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}], \quad (11)$$

and

$$\underset{\sim}{V}^{(r)} = \underset{\sim 1}{\Sigma}^{(r-1)} - \underset{\sim 1}{\Sigma}^{(r-1)}[\underset{\sim 1}{\Sigma}^{(r-1)} + \underset{\sim 2}{\Sigma}^{(r-1)}]^{-1}\underset{\sim 1}{\Sigma}^{(r-1)}. \quad (12)$$

2.2 M Step

At the r-th iteration the "M Step" is the maximization of the "complete-data likelihood function" in which the complete-data sufficient statistic has been replaced by its conditional expectation obtained in the E Step. Setting derivatives of the complete-data likelihood function to zero and manipulating the results produces the likelihood equations:

$$\underset{\sim}{\beta}^{(r)} = (\underset{\sim}{X}^T \underset{\sim 2}{\Sigma}^{(r)^{-1}} \underset{\sim}{X})^{-1} \underset{\sim}{X}^T \underset{\sim 2}{\Sigma}^{(r)^{-1}} (\bar{\underset{\sim}{y}} - \bar{\underset{\sim}{\delta}}^{(r)}) \quad (13)$$

$$\underset{\sim 1}{\tau}^{(r)} = [(\mathrm{Tr}\{\underset{\sim 1}{\Sigma}^{(r)^{-1}} \underset{\sim}{G}_{1k} \underset{\sim 1}{\Sigma}^{(r)^{-1}} \underset{\sim}{G}_{1k'}\})_{kk'}]^{-1} \times [(\mathrm{Tr}\{\underset{\sim 1}{\Sigma}^{(r)^{-1}} \underset{\sim}{G}_{1k} \underset{\sim 1}{\Sigma}^{(r)^{-1}} \underset{\sim 1}{C}^{(r)}\})_k] \quad (14)$$

$k, k' = 1, 2, \ldots m_1$

$$\underset{\sim 2}{\tau} = [(\mathrm{Tr}\{\underset{\sim 2}{\Sigma}^{(r)^{-1}} \underset{\sim}{G}_{2k} \underset{\sim 2}{\Sigma}^{(r)^{-1}} \underset{\sim}{G}_{2k'}\})_{kk'}]^{-1} \times [(\mathrm{Tr}\{\underset{\sim 2}{\Sigma}^{(r)^{-1}} \underset{\sim}{G}_{2k} \underset{\sim 2}{\Sigma}^{(r)^{-1}} (\underset{\sim}{S}^{(r)} + \underset{\sim 1}{C}^{(r)} - 2\underset{\sim 2}{C}^{(r)})\})_k] \quad (15)$$

$k, k' = 1, 2, \ldots m_2$

where:

$$\underset{\sim}{S}^{(r)} = \Sigma_i (\underset{\sim}{y}_i - \underset{\sim}{X}\underset{\sim}{\beta}^{(r)}) (\underset{\sim}{y}_i - \underset{\sim}{X}\underset{\sim}{\beta}^{(r)})^T / N \quad (16)$$

$$\underset{\sim 1}{C}^{(r)} = \underset{\sim}{V}^{(r)} + (\Sigma_i \underset{\sim}{\delta}_i^{(r)} \underset{\sim}{\delta}_i^{(r)^T}) / N = E[\Sigma_i \underset{\sim}{\delta}_i \underset{\sim}{\delta}_i^T \mid \#] / N \quad (17)$$

$$\underset{\sim 2}{C}^{(r)} = \Sigma_i (\underset{\sim}{y}_i - \underset{\sim}{X}\underset{\sim}{\beta}^{(r)}) \underset{\sim}{\delta}_i^{(r)^T} / N \quad (18)$$

and $\bar{\underset{\sim}{\delta}}^{(r)}$ is the mean of the $\underset{\sim}{\delta}_i^{(r)}$.

In general these likelihood equations do not have an explicit solution. The necessity to use an iterative technique to solve the equations, at each iteration of the EM algorithm, would render the entire method unusable.

Notice in these equations that it is possible for $\underset{\sim}{\beta}^{(r)}$ and $\underset{\sim}{\tau}_2^{(r)}$ to have S-explicit representation when $\underset{\sim}{\tau}_1^{(r)}$ does not, i.e., regardless of the form of $\underset{\sim}{\Sigma}_1$. This point leads to some important special cases for which these likelihood equations can be solved directly.

3. A SPECIAL CASE: $\Sigma_2 = \omega I$

When the model (1) is used for longitudinal data the term ε_i is often used to represent measurement errors, which may be assumed to be homoscedastic and uncorrelated, i.e. $\mathrm{Var}(\varepsilon_i) = \Sigma_2 = \omega I$. For the EM algorithm the complete-data sufficient statistic is $(\Sigma_i d_i^T d_i,\ \Sigma_i X^T d_i,\ \Sigma_i \delta_i \delta_i^T)$ where $d_i = y_i - \delta_i$.

<u>E Step</u>. The conditional expectation of the sufficient statistic at the r-th iteration, given the observed data and the values of the parameters from iteration r-1 is:

$$E[\Sigma_i d_i^T d_i \mid y_1, \ldots, y_N;\ \beta^{(r-1)},\ \Sigma_1^{(r-1)},\ \omega^{(r-1)})$$

$$= \Sigma_i (y_i - \delta_i^{(r)})^T (y_i - \delta_i^{(r)}) + N\ \mathrm{Tr}\{V^{(r)}\} \tag{19}$$

$$E[\Sigma_i X^T d_i \mid \#] = \Sigma_i X^T (y_i - \delta_i^{(r)}) \tag{20}$$

$$E[\Sigma_i \delta_i \delta_i^T \mid \#] = N\ V^{(r)} + \Sigma_i \delta_i^{(r)} \delta_i^{(r)^T} \tag{21}$$

where $\delta_i^{(r)}$ and $V^{(r)}$ are given by (11) and (12) with $\Sigma_2^{(r-1)} = \omega^{(r-1)} I$ respectively.

<u>M Step</u>. At the r-th iteration the complete-data likelihood equations are equation (14) for $\tau_1^{(r)}$,:

$$\underset{\sim}{\beta}^{(r)} = (\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T(\underset{\sim}{\overline{y}} - \underset{\sim}{\overline{\delta}}^{(r)}), \tag{22}$$

and

$$\omega^{(r)} = p^{-1}\mathrm{Tr}(\underset{\sim}{S}^{(r)} + \underset{\sim}{C}_1^{(r)} - 2\underset{\sim}{C}_2^{(r)}) \tag{23}$$

In these equations both $\underset{\sim}{\beta}^{(r)}$ and $\omega^{(r)}$ have S-explicit representation regardless of the form of $\underset{\sim}{\Sigma}_1$. Whether $\underset{\sim}{\tau}_1^{(r)}$ has explicit representation depends solely upon the structure of $\underset{\sim}{\Sigma}_1$ and does not depend upon $\underset{\sim}{X}$.

It is interesting that $\omega^{(r)}$ is the unweighted average of the diagonal elements of $\underset{\sim}{S}^{(r)} + \underset{\sim}{C}_1^{(r)} - 2\underset{\sim}{C}_2^{(r)}$; in fact, the off-diagonal elements of $\underset{\sim}{S}^{(r)}$ and $\underset{\sim}{C}_2^{(r)}$ need not be computed.

4. EXAMPLES WITH R-TH ITERATE EXPLICIT SOLUTIONS

Olkin and Press (1969) studied the model $\underset{\sim}{y}_i=\underset{\sim}{\mu}+\underset{\sim}{\eta}_i$, $\underset{\sim}{\eta}_i \sim \text{NID}(\underset{\sim}{0}, \underset{\sim}{\Sigma}_0)$, where $\underset{\sim}{\Sigma}_0$ has circular symmetry pattern, i.e.,

$$\underset{\sim}{\Sigma}_0 = \begin{bmatrix} \sigma_1 & \sigma_2 & \sigma_3 & \cdots & \sigma_p \\ \sigma_p & \sigma_1 & \sigma_2 & \cdots & \sigma \\ & \cdot & \cdot & \cdot & \\ \sigma_2 & \sigma_3 & \sigma_4 & \cdots & \sigma_1 \end{bmatrix}, \tag{24}$$

where $\sigma_j = \sigma_{p-j+2}$.

They showed that when there are no restrictions on $\underset{\sim}{\mu}$ this model has S-explicit representation for both $\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{\Sigma}}_0$. They also showed some situations in which this model provides a good fit for actual data.

Consider the model above generalized so that $\underset{\sim}{\mu}$ has an assumed structure, $\underset{\sim}{\mu}=\underset{\sim}{X}\underset{\sim}{\beta}$, which implies that neither $\hat{\underset{\sim}{\beta}}$ nor $\hat{\underset{\sim}{\Sigma}}_0$ will have explicit representation except for very special combinations of $\underset{\sim}{\Sigma}_0$ and $\underset{\sim}{X}$. Thus, an iterative technique will be required to compute maximum likelihood estimators. We can use an artifice to put this model into the framework of the previous section: let

$$\underset{\sim}{\eta}_i = \underset{\sim}{\delta}_i + \underset{\sim}{\varepsilon}_i$$

$$\begin{pmatrix} \underset{\sim}{\delta}_i \\ \underset{\sim}{\varepsilon}_i \end{pmatrix} \sim \text{NID}\left[\underset{\sim}{0}, \begin{pmatrix} \underset{\sim}{\Sigma}_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Sigma}_2 \end{pmatrix} \right],$$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} \tau_1 & \tau_2 & \tau_3 & \cdots & \tau_p \\ \tau_p & \tau_1 & \tau_2 & \cdots & \tau_{p-1} \\ & \cdot & \cdot & \cdot & \\ \tau_2 & \tau_3 & \tau_4 & \cdots & \tau_1 \end{bmatrix} \qquad (25)$$

$$\underset{\sim}{\Sigma}_2 = \omega \underset{\sim}{I}_p$$

$$\sigma_1 = \tau_1 + \omega$$

$$\sigma_j = \tau_j, \; j = 2, 3, \ldots, p.$$

Note that $\tau_j = \tau_{p-j+2}$ so that $\underset{\sim}{\Sigma}_1$ has circular symmetry pattern and therefore the vector $\underset{\sim}{\tau}_1$ has only INT[(p + 2)/2] elements.

This model is a special case of the model of section 3 in which $\underset{\sim}{\Sigma}_1$ has the circular symmetry pattern, which is a special case of linear structure. We can use the EM algorithm as described in section 3 to obtain maximum likelihood estimates of $\underset{\sim}{\beta}$, $\underset{\sim}{\tau}_1$, and, ω therefore, of $\underset{\sim}{\beta}$ and $\underset{\sim}{\sigma}=(\sigma_1, \sigma_2, \ldots, \sigma_p)^T$. At the r-th iteration, by the results of section 3, $\underset{\sim}{\beta}$, and ω have

S-explicit representation regardless of the structure of $\underset{\sim}{\Sigma}_1$. Since $\underset{\sim}{\Sigma}_1$ (eg.25) has circular symmetry pattern, $\hat{\underset{\sim}{\tau}}_1$ has S-explicit representation. At the r-th iteration of the EM algorithm, the E step is exactly as given in the preceeding section. The M step computations of $\underset{\sim}{\beta}^{(r)}$, $\omega^{(r)}$ and $\underset{\sim}{C}_1^{(r)}$ are given by equations (22), (23), and (17) and

$$\underset{\sim}{\tau}_1^{(r)} = \underset{\sim}{D}_p^{-1}[(Tr\{\underset{\sim}{G}_s\underset{\sim}{C}_1^{(r)}\})_s] \qquad (26)$$

where

$$\underset{\sim}{D}_p = \begin{cases} Diag(p,\ 2p,\ 2p,\ .\ .\ .\ ,\ 2p) & \text{if } p \text{ is odd} \\ Diag(p,\ 2p,\ 2p,\ .\ .\ .\ ,\ 2p,\ p) & \text{if } p \text{ is even.} \end{cases}$$

This problem is overparameterized; τ_1 and ω are not separately identifiable. However, the parameter being estimated is $\sigma_1 = \tau_1 + \omega$, which is identifiable. The practical consequence is that one should check for convergence of $\sigma_1^{(r)} = \tau_1^{(r)} + \omega^{(r)}$ rather than separately checking convergence of $\tau_1^{(r)}$ and $\omega^{(r)}$. After convergence,

$$\hat{\sigma}_j = \begin{cases} \hat{\tau}_1 + \hat{\omega}, & j=1 \\ \hat{\tau}_j, & j=2,\ .\ .\ .\ ,\ INT[(p+2)/2] \\ \hat{\tau}_{p-j+2}, & j>INT[(p+2)/2]. \end{cases}$$

As an illustration consider the case p=6, in which

$$\Sigma_1 = \begin{bmatrix} \tau_1 & \tau_2 & \tau_3 & \tau_4 & \tau_3 & \tau_2 \\ \tau_2 & \tau_1 & \tau_2 & \tau_3 & \tau_4 & \tau_3 \\ \tau_3 & \tau_2 & \tau_1 & \tau_2 & \tau_3 & \tau_4 \\ \tau_4 & \tau_3 & \tau_2 & \tau_1 & \tau_2 & \tau_3 \\ \tau_3 & \tau_4 & \tau_3 & \tau_2 & \tau_1 & \tau_2 \\ \tau_2 & \tau_3 & \tau_4 & \tau_3 & \tau_2 & \tau_1 \end{bmatrix}.$$

Notice that $\underset{\sim}{C}_1^{(r)}$ is a symmetric matrix with the same dimensions as $\underset{\sim}{\Sigma}_1$ and that each $\tau_j^{(r)}$ is the unweighted mean of those elements in $\underset{\sim}{C}_1^{(r)}$ which are in the same row and column of $\underset{\sim}{C}_1^{(r)}$ as a τ_j in Σ_1:

$$\begin{aligned} \tau_1^{(r)} &= Tr(\underset{\sim}{C}_1^{(r)})/6 = (\Sigma_i C_{1;i,i}^{(r)})/6 \\ \tau_2^{(r)} &= (C_{1;1,6}^{(r)} + \Sigma_i C_{1;i,i+1}^{(r)})/6 \\ \tau_3^{(r)} &= (C_{1;1,5}^{(r)} + C_{1;2,6}^{(r)} + \Sigma_i C_{1;i,i+2}^{(r)})/6 \\ \tau_4^{(r)} &= (\Sigma_i C_{1;i,i+3}^{(r)})/3 \end{aligned} \tag{27}$$

We have proven that this pattern--that $\tau_j^{(r)}$ is the unweighted mean of the elements of $\underset{\sim}{C}_1^{(r)}$ in "corresponding positions"--holds for integers $2 \leq p \leq 8$ and we conjecture that the pattern holds for all integers $p \geq 2$. We chose an even value of p for the example because for even p, $\tau_{(p+2)/2}^{(r)}$ is a mean of p/2 terms while each other $\tau_j^{(r)}$ is a mean of p terms. If p is odd every $\tau_j^{(r)}$ is a mean of p terms.

The computations for this special case are straightforward.

5. A COVARIANCE COMPONENTS EXAMPLE

Dempster, Rubin, and Tsutakawa (1981) and Laird and Ware (1982) have shown that the EM algorithm can be applied to find maximum likelihood estimates of the parameters in the covariance components model

$$\underset{\sim}{y}_i = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{Z}\underset{\sim}{\delta}_i^* + \underset{\sim}{\varepsilon}_i$$

where

$$\begin{pmatrix} \underset{\sim}{\delta}_i^* \\ \underset{\sim}{\varepsilon}_i \end{pmatrix} \sim \text{NID}\left[\underset{\sim}{0}, \begin{pmatrix} \underset{\sim}{\Sigma}_1^* & \underset{\sim}{0} \\ \underset{\sim}{0} & \omega \underset{\sim}{I}_p \end{pmatrix}\right]$$

Z is a known $p \times q_2$ matrix of rank $q_2 \leq p$,

$$\underset{\sim}{\Sigma}_1^* = \sum_{s=1}^{m_1} \tau_s \underset{\sim}{G}_s^*$$

$$\underset{\sim}{\delta}_i = \underset{\sim}{Z}\underset{\sim}{\delta}_i^*$$

$$\underset{\sim}{\Sigma}_1 = \text{Var}(\underset{\sim}{\delta}_i) = \underset{\sim}{Z}\underset{\sim}{\Sigma}_1^*\underset{\sim}{Z}^T$$

From Section 3 results we see that the EM algorithm can be applied to this problem by letting $\underset{\sim}{\delta}_i^*$ be the part of the complete data which is not available in the incomplete data. At the r-th iteration let $\underset{\sim}{d}_i = \underset{\sim}{y}_i - \underset{\sim}{\delta}_i = \underset{\sim}{y}_i - \underset{\sim}{Z}\underset{\sim}{\delta}_i^*$; the first two expectations of sufficient statistics are given by (19) and (20), with $\underset{\sim}{\delta}_i^{(r)} = \underset{\sim}{Z}\underset{\sim}{\delta}_i^{*(r)}$, the third by:

$$E[\Sigma_i \underset{\sim}{\delta}_i \underset{\sim}{\delta}_i^{*T} \mid \#] = NV^{*(r)} + \Sigma_i \underset{\sim}{\delta}_i^{*(r)} \underset{\sim}{\delta}_i^{*(r)T} \qquad (28)$$

where

$$\underset{\sim}{\delta}_i^{*(r)} = \underset{\sim}{\Sigma}_1^{*(r-1)} \underset{\sim}{Z}^T [\underset{\sim}{Z}\underset{\sim}{\Sigma}_1^{*(r-1)}\underset{\sim}{Z}^T + \omega^{(r-1)}\underset{\sim}{I}]^{-1}[\underset{\sim}{y}_i - \underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}]$$

$$= [\underset{\sim}{Z}^T\underset{\sim}{Z} + \omega^{(r-1)}(\underset{\sim}{\Sigma}_1^{*(r-1)})^{-1}]^{-1}\underset{\sim}{Z}^T[\underset{\sim}{y}_i - \underset{\sim}{X}\underset{\sim}{\beta}^{(r-1)}] \qquad (29)$$

$$\underset{\sim}{V}^{*(r)} = \underset{\sim}{\Sigma}_1^{*(r-1)} - \underset{\sim}{\Sigma}_1^{*(r-1)} \underset{\sim}{Z}^T [\underset{\sim}{Z} \underset{\sim}{\Sigma}_1^{*(r-1)} \underset{\sim}{Z}^T + \omega^{(r-1)} \underset{\sim}{I}]^{-1} \underset{\sim}{Z} \underset{\sim}{\Sigma}_1^{*(r-1)}$$

$$= \omega^{(r-1)} [\underset{\sim}{Z}^T \underset{\sim}{Z} + \omega^{(r-1)} (\underset{\sim}{\Sigma}_1^{*(r-1)})^{-1}]^{-1} \qquad (30)$$

The latter forms of (29) and (30) are obtained using a matrix identity from Dempster, Rubin, and Tsutakawa (1981). For the M step $\underset{\sim}{\beta}^{(r)}$ is given by (22), $\omega^{(r)}$ by (23), $\underset{\sim}{\tau}_1^{(r)}$ by (14) modified by adding a superscript asterisk to each instance of $\underset{\sim}{\Sigma}_1^{(r)}$, $\underset{\sim}{G}_{1k}$, and $\underset{\sim}{C}^{(r)}$, and

$$\underset{\sim}{C}_1^{*(r)} = \underset{\sim}{V}^{*(r)} + \Sigma_i \underset{\sim}{\delta}_i^{*(r)} \underset{\sim}{\delta}_i^{*(r)^T} / N.$$

Either of the following conditions is sufficient for $\underset{\sim}{\tau}_1^{(r)}$ to have explicit representation:

(1) There are no restrictions on the form of the positive definite symmetric $\underset{\sim}{\Sigma}_1^*$ or

(2) $\underset{\sim}{\Sigma}_1^* = \underset{\sim}{I} \otimes \underset{\sim}{\Delta}$ where there are no restrictions on the form of the positive definite symmetric $\underset{\sim}{\Delta}$.

One of these two conditions is commonly assumed in covariance components models. If $\underset{\sim}{\tau}_1^{(r)}$ has S-explicit representation,

$$\underset{\sim}{\tau}_1^{(r)} = [(Tr\{\underset{\sim}{G}_{1k}^* \underset{\sim}{G}_{1k'}^*\})_{kk'}]^{-1} [(Tr\{\underset{\sim}{G}_{1k}^* \underset{\sim}{C}_1^{*(r)}\})_k],$$

which is a straightforward computation.

ACKNOWLEDGEMENTS

Mr. Andrade was a Ph.D. student in the Department of Biostatistics at the University of North Carolina when this research was performed. This research was supported in part by Pediatric Pulmonary Specialized Center of Research Grant P50-19171, National Heart, Lung, and Blood Institute, and by a Fellowship from EMBRAPA.

REFERENCES

Anderson, T.W. (1970). "Estimation of Covariance Matrices which are Linear Combinations or whose Inverses are Linear Combinations of Given Matrices." Essays in Probability and Statistics, 1-24. Ed. R.C. Bose, I.M. Chakravarti, P.C. Mahalanobis, C.R. Rao, and K.J.C. Smith. University of North Carolina Press, Chapel Hill, NC.

Anderson, T.W. (1973). "Asymptotic Efficient Estimation of Covariances Matrices with Linear Structure." Ann. Statist. 1, 135-41.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). J.R.Statist.Soc. B39, 1-39.

Dempster, A.P., Rubin, D.B. and Tsutakawa, R.K, (1981). "Estimation in Covariance Components Models." J.Am.Statist.Assoc. 76, 341-53.

Laird, N.M. and Ware, J.H. (1982). "Random-effects Models for Longitudinal Data." Biometrics 38, 963-74.

Olkin, I. and Press, S.J. (1969). "Testing and Estimation for a Circular Stationary Model." Ann. Math. Statist. 40, 1358-73.

Rubin, D.B. and Szatrowski, T.H. (1982). "Finding Maximum Likelihood Estimates of Patterned Covariance Matrices by the EM Algorithm." Biometrika 69, 656-60.

Szatrowski, T.H. (1980). "Necessary and Sufficient Conditions for Explicit Solutions in the Multivariate Normal Estimation Problem for Patterned Means and Covariances." Ann. Statist. 8, 802-10.