A NEW ITERATIVE PROCEDURE FOR THE MISSING-VALUE PROBLEM*

WANG SONG-GUI (王松桂)
(China University of Science and Technology, Hefei, China)

Abstract

For a missing-value problem in linear models, all the iterative procedures employed up to now are entirely within the framework of the EM algorithm. This paper proposes a new iterative procedure which is not the EM algorithm in the most general case. In the procedure nothing is assumed about the error distributions in the proof of its convergence. The convergence rate is also obtained.

1. Introduction and Notations

Let us consider the linear model

$$y_i = x_i'\beta + e_i, \quad i = 1, 2, \cdots, n, \tag{1.1}$$

where y_i are observations of a dependent variable y, $x_i' = (x_{i1}, \cdots, x_{ip})$ are values of p independent variables, β is an unknown p-vector, and e_i are random errors with

$$E(e_i) = 0, \quad \operatorname{cov}(e_i, e_j) = \begin{cases} \sigma^2, & i = j, \\ 0, & i \neq j. \end{cases} \tag{1.2}$$

Suppose that only n_1 of the n intended observations y_i are available, while the other $n - n_1$ of y_i are missing. Without loss of generality, we may take these missing values to be the last $n - n_1$ components of the observation vector $y' = (y_1, \cdots, y_n)$. In general, $x_{n_1+1}, \cdots, x_n$, corresponding to these missing y-values, may or may not lie in $\mu(x_1, \cdots, x_{n_1})$, the space generated by the vectors $x_1, \cdots, x_{n_1}$. Under this circumstance the model (1.1) can be rewritten in the matrix notation

$$y_1 = X_1\beta + e_1, \tag{1.3}$$

$$y_2 = X_2\beta + e_2, \tag{1.4}$$

$$y_3 = X_3\beta + e_3, \tag{1.5}$$

where

$$\mu(X_2') \subset \mu(X_1'), \tag{1.6}$$

$$\mu(X_3') \cap \mu(X_1') = \{0\}, \tag{1.7}$$

and y_i are $n_i \times 1$ vectors, X_i are $n_i \times p$ design matrices, and e_i are $n_i \times 1$ error vectors.

Suppose that y_2 and y_3 are missing. Then, the sub-model (1.3) does not have the advantages of balance properties of the full-model (1.3)—(1.5), and thus its statistical computations are often complicated. It is therefore worthwhile to investigate whether or not we can retain the balance structure of the full-model (1.3)—

* Received May 11, 1983.

(1.5) so as to analyse the sub-model (1.3). All the iterative procedures employed up to now are entirely within the framework of the EM algorithm[1-4]. The purpose of this paper is to propose a new iterative procedure for the most general case (1.3)—(1.5). Following the introduction of the procedure in Section 2, Section 3 is devoted to the proof of its convergence. Included in Section 4 is a brief discussion on the comparison between our procedure and the EM algorithm. The merits of the new procedure are that except for (1.2), nothing is assumed about the error distributions in proving its convergence, and the convergence rate is also obtained. On the contrary, no rate has been obtained for the EM algorithm.

The notations used throughout are as follows: A^- is any g-inverse of A. $\mu(A)$ is the space spaned by the columns of A. $X'=(X_1':X_2':X_3')$, $M_{ij}=X_i(X'X)^-X_j'$, $N_{ij}=X_i(X_1'X_1)^-X_j'$, $i, j=1, 2, 3$, and $\hat\beta=(X_1'X_1)^-X_1'y_1$ is the least squares solution of β in the sub-model (1.3).

2. The Iterative Procedure

Our aim is to compute $c'\hat\beta$, $c\in\mu(X_1')$, by using $(X'X)^-$ rather than $(X_1'X_1)^-$. The iterative procedure we propose is as follows.

Given a starting point $\tilde\beta^{(0)}$, solve the equation

$$X'X\beta=X_1'y_1+X_2'X_2\tilde\beta^{(0)}. \tag{2.1}$$

Denote by $\tilde\beta^{(1)}$ any special solution to (2.1). In general, let $\tilde\beta^{(k-1)}$ be the current value of β after $k-1$ cycles. Solve the equation

$$X'X\beta=X_1'y_1+X_2'X_2\tilde\beta^{(k-1)}, \tag{2.2}$$

and denote by $\tilde\beta^{(k)}$ any special solution. In this way we obtain an estimate sequence $\{\tilde\beta^{(k)}\}$.

Observe that the full-model being considered involves equation (1.5); therefore it is always possible to augment X_3 such that $X'X$ is nonsingular, or even $X'X=I_p$ in some cases. But in our proof of the convergence theorem, it is not necessary to assume $X'X$ to be nonsingular. By definition,

$$\tilde\beta^{(k)}=(X'X)^-(X_1'y_1+X_2'X_2\tilde\beta^{(k-1)}), \tag{2.3}$$

where $(X'X)^-$ is an arbitrary g-inverse of $X'X$. We will specify an arbitrary one, and denote it by $(X'X)_1^-$ for all k in (2.3).

3. Convergence of the Iteration

Theorem. *For any* $\tilde\beta^{(0)}$,

(1) $\tilde\beta^{(k)}$ *is a convergent sequence;*

(2) *denote* $\lim\limits_{k\to+\infty}\tilde\beta^{(k)}=\beta^*$; *then* $c'\beta^*=c'\hat\beta$ *for all* $c\in\mu(X_1')$;

(3) $|c'\tilde\beta^{(k)}-c'\hat\beta|=O(\max\limits_{1<i<n_2}(\alpha_i/1+\alpha_i)^{k-1})$,

where $\alpha_i\geq0$ *are the eigenvalues of* N_{22}, *defined at the end of Section* 1.

Proof. By (1.6) and (1.7), we get

$$M_{ij}=\begin{cases}X_i(X_1'X_1+X_2'X_2)^-X_j', & \text{for } i, j=1, 2, \quad (3.1)\\ 0, & \text{for } i=1, 2, j=3, \text{ or } i=3, j=1, 2. \quad (3.2)\end{cases}$$

In fact, if we choose a nonsingular matrix $B'=(B_1':B_2':B_3')$ with $\mu(X_1')=\mu(B_1')$ and $\mu(X_3')=\mu(B_2')$, then there exist matrices C_1, C_2 and C_3 such that

$$X=\begin{pmatrix} X_1\\ X_2\\ X_3\end{pmatrix}=\begin{pmatrix} C_1 & 0 & 0\\ C_2 & 0 & 0\\ 0 & C_3 & 0\end{pmatrix}\begin{pmatrix} B_1\\ B_2\\ B_3\end{pmatrix}.$$

Thus

$$\begin{aligned} M_{11}&=(C_1\ 0\ 0)\begin{pmatrix} C_1'C_1+C_2'C_2 & 0 & 0\\ 0 & C_3'C_3 & 0\\ 0 & 0 & 0\end{pmatrix}^{-}\begin{pmatrix} C_1'\\ 0\\ 0\end{pmatrix}\\ &=(C_1\ 0\ 0)\begin{pmatrix} (C_1'C_1+C_2'C_2)^{-} & 0 & 0\\ 0 & (C_3'C_3)^{-} & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix} C_1'\\ 0\\ 0\end{pmatrix}\\ &=(C_1\ 0\ 0)B\left(B'\begin{pmatrix} C_1' & C_2'\\ 0 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix} C_1 & 0 & 0\\ C_2 & 0 & 0\end{pmatrix}B\right)^{-}B'\begin{pmatrix} C_1'\\ 0\\ 0\end{pmatrix}\\ &=X_1(X_1'X_1+X_2'X_2)^{-}X_1', \end{aligned}$$

as the choice of g-inverses $(X'X)^{-}$ and $(X_1'X_1+X_2'X_2)^{-}$ are irrelevant in M_{11}. Other assertions in (3.1) and (3.2) can be proved similarly.

By using (3.1) and the fact that

$$(X_1'X_1+X_2'X_2)^{-}=(X_1'X_1)^{-}-(X_1'X_1)^{-}X_2'[I_{n_2}+X_2(X_1'X_1)^{-}X_2']^{-}X_2(X_1'X_1)^{-}-(X_1'X_1)^{-}X_2'C-DX_2(X_1'X_1)^{-}, \tag{3.3}$$

where C and D are two arbitrary matrices with $X_1'X_1D=0$, $CX_1'X_1=0$, it is easy to show that

$$M_{22}=X_2(X_1'X_1+X_2'X_2)^{-}X_2'=N_{22}(I+N_{22})^{-1}. \tag{3.4}$$

From this, it can be shown that the eigenvalues of M_{22} are strictly less than 1[3]; thus

$$M_{22}^k\to 0 \quad \text{as} \quad k\to\infty. \tag{3.5}$$

On the other hand, by the definition of $\tilde{\beta}^{(k)}$, one can verify that

$$\tilde{\beta}^{(k)}-\tilde{\beta}^{(k-1)}=(X'X)_1^{-}X_2'M_{22}^{k-1}X_2(\tilde{\beta}^{(1)}-\tilde{\beta}^{(0)}), \tag{3.6}$$

where $(X'X)_1^{-}$ is a specified g-inverse of $(X'X)$, mentioned earlier. Combining (3.5) and (3.6), we get the proof of (1).

Taking the limit in both sides of (2.3), and denoting it by β^*, we get

$$\beta^*=(X'X)_1^{-}X_1'y_1+(X'X)_1^{-}X_2'X_2\beta^*. \tag{3.7}$$

For any $c\in\mu(X_1')$, c can be written as $c=X_1'\alpha$ for some α. Using (3.3) and (3.7) and observing that $X_i(X'X)^{-}X_j$ is independent of the choice of $(X'X)^{-}$, we have

$$\begin{aligned} c'\beta^*&=\alpha'X_1(X'X)_1^{-}X_1'y_1+\alpha'X_1(X'X)_1^{-}X_2'X_2\beta^*\\ &=\alpha'X_1(X_1'X_1)^{-}X_1'y_1+\alpha'X_1(X_1'X_1)^{-}X_2'X_2\beta^*\\ &\quad-\alpha'X_1(X_1'X_1)^{-}X_2'X_2(X'X)^{-}X_1'y_1\\ &\quad-\alpha'X_1(X_1'X_1)^{-}X_2'X_2(X'X)^{-}X_2'X_2\beta^*\\ &=c'\hat{\beta}+\alpha'N_{12}X_2[\beta^*-(X'X)_1^{-}X_1'y_1-(X'X)_1^{-}X_2'X_2\beta^*]=c'\hat{\beta}. \end{aligned}$$

The last equality follows from (3.7). This proves (2).

Now we prove (3). After simple and tedious calculations, we rewrite $\tilde{\beta}^{(k)}$ as

$$\tilde{\beta}^{(k)}=(X'X)_1^- X_1'y_1+(X_1'X_1)_1^- X_2(I-M_{22}^{k-1})(I-M_{22})^{-1}M_{21}y_1.$$

Again by (3.1) and (3.3), we have

$$\begin{aligned} c'\tilde{\beta}^{(k)}-c'\hat{\beta}&=\alpha'M_{12}(I-M_{22}^{k-1})(I-M_{22})^{-1}M_{21}y_1\\ &\quad-\alpha'X_1(X_1'X_1+X_2'X_2)^- X_1'y_1-\alpha'N_{11}y_1\\ &=\alpha'M_{12}(I-M_{22}^{k-1})(I-M_{22})^{-1}M_{21}y_1-\alpha'N_{12}(I+N_{22})^{-1}N_{21}y_1\\ &=\alpha'N_{12}(I+N_{22})^{-1}(I-M_{22}^{k-1}-(I+N_{22})(I-M_{22}))\\ &\quad\times(I-M_{22})^{-1}(I+N_{22})^{-1}N_{21}y_1\\ &=-\alpha'N_{12}(I+N_{22})^{-1}M_{22}^{k-1}(I-M_{22})^{-1}(I+N_{22})^{-1}N_{21}y_1. \end{aligned}$$

In view of the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} |c'\tilde{\beta}^{(k)}-c'\hat{\beta}|&\leqslant(\alpha'N_{12}(I+N_{22})^{-1}M_{22}^{k-1}(I-M_{22})^{-2}M_{22}^{k-1}(I+N_{22})^{-1}N_{12}\alpha)^{1/2}\\ &\quad\times(y_1'N_{12}(I+N_{22})^{-2}N_{21}y_1)^{1/2}. \end{aligned}\tag{3.8}$$

Denote by α_i the eigenvalues of N_{22}, $i=1, 2, \cdots, n_2$. It follows from (3.4) that the eigenvalues of $M_{22}^{k-1}(I-M_{22})^{-2}M_{22}^{k-1}$ are $\alpha_i^{2(k-1)}/(1+\alpha_i)^{2(k-2)}$, $i=1, \cdots, n_2$. By a well-known property of eigenvalues (see, for example, Rao [5]), the first factor on the right hand side of (3.8) is less than or equal to $\max\limits_{1<i<n_2}\alpha_i^{k-1}/(1+\alpha_i)^{k-2}$; the second factor, however, is independent of k. Therefore

$$|c'\tilde{\beta}^{(k)}-c'\hat{\beta}|\leqslant C\max_{1<i<n_2}(\alpha_i^{k-1}/(1+\alpha_i)^{k-2}),$$

which completes the proof of (3).

4. Discussion

We conclude by comparing the procedure proposed here with the EM algorithm.

For the missing-value problem considered in this paper, since the EM algorithm does not distinguish equation (1.5) from equation (1.4), if we denote all the missing values by y_u and the corresponding design matrix by X_u, the model (1.3)—(1.5) can be rewritten as

$$y_1=X_1\beta+e_1,\tag{4.1}$$

$$y_u=X_u\beta+e_u.\tag{4.2}$$

Starting from a specific $\beta^{(0)}$, we can obtain an EM sequence $\{\beta^{(k)}\}$ by a two-step cycle:

E-step: Estimate the missing values by their expectations, given current parameter values, $\tilde{y}_u^{(k-1)}=X_u\beta^{(k-1)}$,

M-step: Determine $\beta^{(k)}$ as the solution to the equation

$$X'X\beta=X_1'y_1+X_u'\tilde{y}_u^{(k-1)}.\tag{4.3}$$

Comparing (2.2) and (4.3), we can see that if the missing values y_u correspond only to X_2, then the two procedures are identical. For this special case, the iterative procedure has already been discussed in detail by Preece[4] and Xiang[8]; the latter has given a concise proof of its convergence. For the most general case (1.3)—(1.5), however, the convergence proof of our procedure is independent of the error distributions, as we saw in Section 3. The EM algorithm is a general iterative approach to getting the maximum likelihood estimate when the observations can be viewed as incomplete data. In nature, its derivation and convergence proof depend largely upon some supplementary assumptions[1,2,6]. Further, the convergence rate of our procedure has been obtained, while no results of this kind are known for the EM algorithm.

In addition, it may be noted that since the full-model contains equation (1.5), it is always possible to add to a given design matrix some rows which are not present in our original design. In such way $X'X$ can be assumed to be nonsingular. In particular, if the side conditions for the least squares estimate are $H\beta=0$, where H is a known matrix with $R(X)+R(H)=p$ and $\mu(X')\cap\mu(H')=\{0\}$, $R(A)$ denotes the rank of A, then we might regard H as a part of X_2. Hence (2.2) becomes

$$(X'X+H'H)\beta=X_1'y_1+X_2'X_2\beta^{(\alpha-1)} \tag{4.4}$$

Now the coefficient matrix $X'X+H'H$ is nonsingular, and $(X'X+H'H)^{-1}$ is a candidate for $(X'X)_1^-$. From the iterative point of view, we prefer (4.4) to (2.2), because H is usually readily found.

A non-iterative method for the missing-value problem in the model (1.3)—(1.5) can also be obtained, and will be the subject of another paper.

Acknowledgement. The author is grateful to Professor Chen Xi-ru for his valuable suggestions.

References

[1] A. P. Dempster, et al., Maximum likelihood from incomplete data via the EM algorithm, J. R. Statist. Soc., B. 39, 1977, 1—22.

[2] Wang Song-gui, EM algorithm, *Applied Mathematics and Mathematics of Computation*, 6 (1983), 43—48.

[3] Xiang Ke-feng, The statistical treatment of non-orthogonal data, *Acta Math. Appl. Sinica*, 6, 1983, 70—75.

[4] D. A. Preece, Iterative procedure for missing values in experiments, *Technometrics*, 13, 1971, 743—753.

[5] C. R. Rao, Linear Statistical Inferences and its Applications. 2nd edition, John Wiley, 1973.

[6] Wu C. F. Jeff, On the convergence of EM algorithm, *Ann. Statist*, 11: 1 (1983), 95—103.