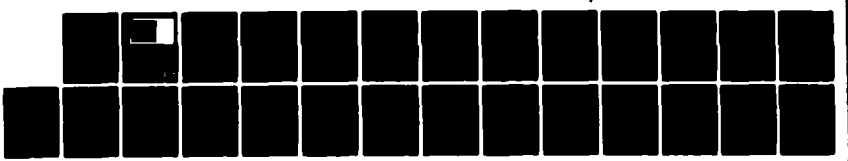
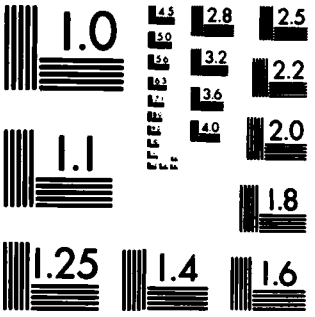
AD-A124 358 RECURSIVE ESTIMATION PROCEDURES FOR MISSING-DATA 1/1
PROBLEMS(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH
CENTER D M TITTERINGTON ET AL. SEP 82 MRC-TSR-2427
UNCLASSIFIED DAAG29-80-C-0041 F/G 12/1 NL



END
DATE FILMED
DTIC



MRC Technical Summary Report #2427

RECURSIVE ESTIMATION PROCEDURES FOR MISSING-DATA PROBLEMS

D. M. Titterington
and J-M. Jiang

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

September 1982

(Received May 12, 1982)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

RECURSIVE ESTIMATION PROCEDURES FOR MISSING-DATA PROBLEMS

D. M. Titterington* and J-M. Jiang**

Technical Summary Report #2427
September 1982

ABSTRACT

Titterington (1982) proposed recursive methods for dealing with incomplete data. The present paper concentrates on versions of these for multiparameter problems involving missing data. Theorems are outlined from which asymptotic properties of the recursive procedures can be established and versions of the recursions are written down for problems in which the missing data are missing at random. After illustration with exponential family models, the case of multivariate Normal data is considered in detail. Numerical comparisons of the various methods are obtained using bivariate Normal data.

AMS (MOS) Subject Classifications: 62A10, 62F10, 62L12, 62H12, 62L20

Key Words: Parameter estimation, missing values, recursive estimation, EM algorithm, exponential family, multivariate Normal, stochastic approximation

Work Unit Number 4 (Statistics and Probability)

*Department of Statistics, University Gardens, Glasgow G12 8QW, Scotland.

**Department of Mathematics and Statistics, SUNY at Albany, 1400 Washington Avenue, Albany, NY 12222.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

This paper is a development of a previous report (#2376) by the first author. The introductory comments to that paper are highly relevant here. Whereas the previous paper discussed incomplete data in general, the present one restricts attention to the problem of missing values. Typically, each experimental unit should have records of the values of several characteristics associated with it. Statistical analysis is made difficult if one or more of those values are missing on some units.

To combat the heavy analysis required for a "proper" analysis of the data, comparatively simple recursive procedures are outlined in which the data are incorporated sequentially into the estimation scheme. Some comments are made about theoretical properties and special emphasis is laid on the case of data from multivariate Normal distributions.

RECURSIVE ESTIMATION PROCEDURES FOR MISSING-DATA PROBLEMS

D. M. Titterington* and J-M. Jiang**

1. INTRODUCTION

Suppose $y_1, y_2, \ldots$ form a sequence of independent observations from a population with parametric probability density function $g(y|\underline{\theta})$, where $\underline{\theta}$ is a vector of s parameters. Let $\underline{S}(y,\underline{\theta})$, the vector of scores corresponding to a single observation, be defined by

$$S_j(y,\underline{\theta}) = \frac{\partial}{\partial \theta_j} \log g(y|\underline{\theta}), \qquad j = 1,\ldots,s ,$$

and let $I(\underline{\theta})$ be the Fisher Information matrix corresponding to one observation. It is assumed that all these quantities exist and that the "usual" regularity conditions hold in order that differentiation with respect to $\underline{\theta}$ and integration over y may be interchanged. Consider the recursion

$$\hat{\underline{\theta}}_{k+1} = \hat{\underline{\theta}}_k + (k+1)^{-1} I(\hat{\underline{\theta}}_k)^{-1} \underline{S}(y_{k+1}, \hat{\underline{\theta}}_k), \qquad k = 0,1,\ldots \tag{1}$$

Under certain extra mild conditions referred to in Titterington (1982), as $k \to \infty$,

$$\sqrt{k}\,(\hat{\underline{\theta}}_k - \underline{\theta}_T) \to N(\underline{0}, I(\underline{\theta}_T)^{-1})$$

in distribution, where $\underline{\theta}_T$ is the true value of $\underline{\theta}$.

If we are dealing with incomplete data, however, the asymptotically efficient stochastic approximation (1) may not be easy to apply, mainly because $I(\underline{\theta})$ can be difficult to evaluate and, if necessary, invert. The former problem arises even with simple, one-parameter models, such as the estimation of the mixing weight in a mixture of two known densities, and both problems appear in, for instance, parameter estimation for a mixture of two univariate Normal densities. These illustrations appear in Titterington (1982) where recursions alternative to (1) are also suggested. One natural proposal is to

*Department of Statistics, University Gardens, Glasgow G12 8QW, Scotland.

**Department of Mathematics and Statistics, SUNY at Albany, 1400 Washington Avenue, Albany, NY 12222.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

replace $(k + 1)I(\hat{\underline{\theta}}_k)$ by the total sample information up to this point. This might still require awkward matrix inversion. Another suggestion is to use, instead, the recursion

$$\tilde{\underline{\theta}}_{k+1} = \tilde{\underline{\theta}}_k + (k + 1)^{-1} I_c(\tilde{\underline{\theta}}_k)^{-1} \underline{S}(y_{k+1}, \tilde{\underline{\theta}}_k), \qquad k = 0,1,... \tag{2}$$

where $I_c(\underline{\theta})$ is the Fisher Information matrix corresponding to a <u>complete</u> observation. $I_c(\underline{\theta})$ is usually easy to evaluate and in many applications, is easy to invert. Again an asymptotic result can be drawn upon and the recursion has strong associations with the EM algorithm (Dempster, Laird and Rubin, 1977) for maximum likelihood estimation; see Titterington (1982).

The objective of this paper is to develop these recursions for missing-data problems, with emphasis on multivariate Normal data, and to assess their performance on moderately large data-sets. One aspect of importance is the dependence of the results on the <u>order</u> in which the data are incorporated. Given a set of n data points it is intended that $\hat{\underline{\theta}}_n$, as given by (1) or $\tilde{\underline{\theta}}_n$, from (2), be used as the parameter estimate, one pass having been made through the data. It is hoped that such a procedure gives estimators which perform well and yet is inexpensive in computer time in comparison with the iterative numerical procedures, such as the EM algorithm itself, traditionally used with incomplete data. These methods should prove to be very useful with large incomplete data-sets, such as sample surveys with nonresponse.

2. THEORETICAL ASPECTS

The important theoretical questions concern the asymptotic properties of the sequence of estimators generated by the stochastic approximations (1) and (2). We extract the following theorem from the work of Walk (1977), giving the essential features of the result but not detailing all the many regularity conditions.

Theorem 1. Consider the recursion

$$\underline{\theta}_{k+1} = \underline{\theta}_k - (k+1)^{-1}\{\underline{f}(\underline{\theta}_k) + \underline{V}_{k+1}\} ,$$

where $E(\underline{V}_{k+1}|\underline{V}_1,\ldots,\underline{V}_k) = \underline{0}$, almost surely, and, for each k, $E(\underline{V}_k^T\underline{V}_k) < \infty$, where the expected value is based on the true distribution. Suppose

$$\underline{f}(\underline{\theta}) = \underline{f}(\underline{\theta}_T) + A(\underline{\theta}_T)(\underline{\theta} - \underline{\theta}_T) + o|\underline{\theta} - \underline{\theta}_T|$$

and that the eigenvalues of $A_T = A(\underline{\theta}_T)$ are all greater than $\frac{1}{2}$. Then, given certain other regularity conditions, as $k \to \infty$,

$$\sqrt{k}\ (\underline{\theta}_k - \underline{\theta}_T) \to N(\underline{0}, B_T) ,$$

in distribution, where B_T is the unique solution of

$$(A_T - \tfrac{1}{2} I_s)B_T + B_T^T(A_T - \tfrac{1}{2} I_s) = M_T \qquad (3)$$

In this equation, I_s denotes the $s \times s$ identity matrix and $M_T = \lim_{k\to\infty} \mathrm{cov}(\underline{V}_k)$.

In our applications, for $\underline{f}(\underline{\theta}_k) + \underline{V}_{k+1}$ we have

$$-G(\underline{\theta}_k)\underline{S}(y_{k+1}, \underline{\theta}_k)$$

where $G(\underline{\theta}) = I^{-1}(\underline{\theta})$ in (1) and $G(\underline{\theta}) = I_c^{-1}(\underline{\theta})$ in (2).

Now

$$\underline{S}(y, \underline{\theta}_k) = E\ \underline{S}(y, \underline{\theta}_k) + \{\underline{S}(y, \underline{\theta}_k) - E\ \underline{S}(y, \underline{\theta}_k)\}$$

and

$$E\ S(y, \underline{\theta}_k) = E\ S(y, \underline{\theta}_T) - I(\underline{\theta}_T)(\underline{\theta}_k - \underline{\theta}_T) = -I(\underline{\theta}_T)(\underline{\theta}_k - \underline{\theta}_T) .$$

In the theorem, therefore, we have

$$A(\underline{\theta}) = G(\underline{\theta})I(\underline{\theta})$$

and

$$M_T = G(\underline{\theta}_T)I(\underline{\theta}_T)G(\underline{\theta}_T) .$$

For recursion (1), with $G(\underline{\theta}) = I(\underline{\theta})^{-1}$, $A(\underline{\theta}) = I_s$ (so that the eigenvalue condition is certainly satisfied), $M_T = I(\underline{\theta}_T)^{-1}$ and therefore $B_T = I(\underline{\theta}_T)^{-1}$, giving the result stated in Section 1.

When A_T is symmetric, equation (3) can be solved explicitly in terms of the eigenvalues and eigenvectors of A_T, giving the solution obtained by Sacks (1958, p. 399) and Fabian (1968). In (2), however, with

$$A(\underline{\theta}) = I_c(\underline{\theta})^{-1} I(\underline{\theta}) ,$$

symmetry of $A(\underline{\theta})$ is not guaranteed unless $I_c(\underline{\theta})$ and $I(\underline{\theta})$ are both diagonal. Equation (3) does, however, give a linear equation from which B_T may be determined, in principle.

With recursion (2) the eigenvalue condition may come into play. If λ^* is the minimum eigenvalue of A_T, we require $\lambda^* > \frac{1}{2}$. Otherwise $\sqrt{k}$-consistency does not follow for $\{\tilde{\underline{\theta}}_k\}$. Provided, however, $\lambda^* > \frac{1}{2}\beta > 0$, A_T is symmetric and there is sufficient regularity,

$$k^{\beta/2}(\tilde{\underline{\theta}}_k - \underline{\theta}_T) \to N(\underline{0}, B(\beta, \underline{\theta}_T)) , \qquad (5)$$

in distribution as $k \to \infty$, where $B_\beta = B(\beta, \underline{\theta}_T)$ satisfies

$$(A_T - \tfrac{1}{2}\beta I_s)B_\beta + B_\beta(A_T - \tfrac{1}{2}\beta I_s) = M_T .$$

This result comes from Fabian (1968). Although the following result has not been tracked down it is reasonable to suppose that, if A^T is nonsymmetric, (5) holds with B_β satisfying

$$(A_T - \tfrac{1}{2}\beta I_s)B_\beta + B_\beta(A_T^T - \tfrac{1}{2}\beta I_s) = M_T .$$

From (4) we may interpret the eigenvalues of $A(\underline{\theta})$ as giving the amount of information about the parameters available in an incomplete observation relative to a complete one. Were $\lambda^* = 0$ it would mean that not all the parameters are identifiable from the incomplete data and neither (1) nor (2) would be usable. Otherwise the simple recursion (2) does give consistent estimates, if not asymptotically efficient ones.

One-parameter problems were discussed in Titterington (1982) and the multiparameter theorem above could be applied to the Normal mixture problem described in that paper. Here we concentrate on multivariate data with missing values.

3. APPLICATION TO MISSING-DATA PROBLEMS

The problem will be formulated, as in Rubin (1976) and Little (1978), with the help of missing-value indicators. Suppose each observation is r-dimensional and that $\underline{d}_i$ is an indicator vector for observation i such that $\underline{d}_i$ has r-components, zeros to denote missing values and ones elsewhere. Let $\underline{z}_i$ denote the set of observed values, where the symbol 'v' is inserted for a component which is missing. Then the overall observed quantity for observation i is

$$\underline{y}_i = (\underline{z}_i, \underline{d}_i) .$$

A typical complete observation would be

$$(\underline{x}_i, \underline{1}) ,$$

in which $\underline{1}$ is a vector of ones and $\underline{x}_i$ has no "v".

Introduce the density functions $g(\underline{z}, \underline{d} | \underline{\theta})$ and $f(\underline{x}, \underline{1} | \underline{\theta})$ and make the following assumptions.

(i) $g(\underline{z}, \underline{d} | \underline{\theta}) = g_1(\underline{z} | \underline{\theta}_1) g_2(\underline{d} | \underline{\theta}_2)$.

(ii) $f(\underline{x}, \underline{1} | \underline{\theta}) = f_1(\underline{x} | \underline{\theta}_1) g_2(\underline{1} | \underline{\theta}_2)$.

(iii) $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2)$ where $\underline{\theta}_1$ and $\underline{\theta}_2$ are distinct sets of parameters.

If $\underline{z}$ is now interpreted as representing just the non-missing components, it is also assumed that

$$g_1(\underline{z} | \underline{\theta}_1) = \int_{X(\underline{z})} f_1(\underline{x} | \underline{\theta}_1) d\underline{x} ,$$

where $X(\underline{z})$ is the set of $\underline{x}$ which can be regarded as completions of $\underline{z}$.

Under these assumptions we may ignore the missing-data process when making inferences about $\underline{\theta}_1$, see Rubin (1976). As a result of (i) we may write the score vector and information matrix as

$$\underline{S}(\underline{y}, \underline{\theta}) = \begin{pmatrix} \underline{S}_1(\underline{z}, \underline{\theta}_1) \\ \underline{S}_2(\underline{d}, \underline{\theta}_2) \end{pmatrix}$$

and

$$I(\underline{\theta}) = \begin{pmatrix} I_1(\underline{\theta}_1) & 0 \\ 0 & I_2(\underline{\theta}_2) \end{pmatrix} ,$$

so that

$$I(\underline{\theta})^{-1}\underline{S}(\underline{y},\underline{\theta}) = \begin{pmatrix} I_1(\underline{\theta}_1)^{-1}\underline{S}_1(\underline{z},\underline{\theta}_1) \\ I_2(\underline{\theta}_2)^{-1}\underline{S}_2(\underline{d},\underline{\theta}_2) \end{pmatrix} .$$

This separation means that, in the recursions, $\underline{\theta}_1$ and $\underline{\theta}_2$ can be updated more or less separately and indeed updating $\underline{\theta}_2$ reduces simply to the estimation of multinomial parameters using relative frequencies. The separation is not quite complete because

$$I_1(\underline{\theta}_1) = \sum_{\underline{d}} g_2(\underline{d}|\underline{\theta}_2) I_1(\underline{\theta}_1|\underline{d}) ,$$

where $I_1(\underline{\theta}_1|\underline{d})$ denotes the information matrix obtained from missing-data pattern $\underline{d}$. Typically $I_1(\underline{\theta}_1|\underline{d})$ will be calculable, as it is associated with a complete-data problem, of lower dimension than r. Thus, in this class of problems it is often possible to calculate $I_1(\underline{\theta}_1)$ explicitly, although it does depend on $\underline{\theta}_2$. In this respect recursion (1) should be more feasible here than it is for mixture problems, say.

It is clear, from Section 2, that $I_1(\underline{\theta}_1)$ has to be positive definite.

Example. Exponential family models.

Suppose

$$\log f_1(\underline{x}|\underline{\theta}_1) = \text{const} + \underline{t}(\underline{x})^T\underline{\theta}_1 - a(\underline{\theta}_1)$$

and define $\underline{\phi} = E(\underline{t}(\underline{x})|\underline{\theta}_1)$. Given complete data $\underline{x}_1,\ldots,\underline{x}_n$, yielding $\underline{t}_1,\ldots,\underline{t}_n$, the maximum likelihood estimator for $\underline{\phi}$ is $\hat{\underline{\phi}}_n = n^{-1}\sum_{i=1}^{n}\underline{t}_i$, which can be calculated recursively from

$$\hat{\underline{\phi}}_{k+1} = \hat{\underline{\phi}}_k + (k+1)^{-1}(\underline{t}_{k+1} - \hat{\underline{\phi}}_k) , \qquad (6)$$

$k = 0,1,\ldots,n-1$, with $\hat{\underline{\phi}}_0 = \underline{0}$.

Suppose $\underline{t}^{(d)}$ denotes the components of $\underline{t}$ observed when the missing-data pattern in $\underline{z}$ is $\underline{d}$. Then, as was shown in Titterington (1982), where the relationship with a

recursive version of the EM algorithm was pointed out, recursion (2) can be written

$$\tilde{\underline{\phi}}_{k+1} = \tilde{\underline{\phi}}_k + (k+1)^{-1}\{E(\underline{t}_{k+1}|\underline{t}_{k+1}^{(d)}, \tilde{\underline{\phi}}_k) - \tilde{\underline{\phi}}_k\} . \tag{7}$$

For special exponential family problems this recursion takes another interesting form. Suppose we write $\underline{t}^T = (\underline{t}^{(d)^T}, \underline{t}^{(\bar{d})^T})$ and suppose that $\underline{t}_d$ is a *cut*; see Barndorff-Nielsen (1978, p. 50). Then $\underline{t}^{(\bar{d})}$ has linear regression on $\underline{t}^{(d)}$ (Barndorff-Nielsen, 1978, p. 197), so that, for some matrix $H(\underline{\phi})$,

$$E(\underline{t}^{(\bar{d})}|\underline{t}^{(d)}, \underline{\phi}) - \underline{\phi}^{(\bar{d})} = H(\underline{\phi})(\underline{t}^{(d)} - \underline{\phi}^{(d)}) \tag{8}$$

Thus, partitioning (7), we have

$$\left.\begin{aligned} \tilde{\underline{\phi}}_{k+1}^{(d)} &= \tilde{\underline{\phi}}_k^{(d)} + (k+1)^{-1}(\underline{t}_{k+1}^{(d)} - \tilde{\underline{\phi}}_k^{(d)}) \\ \tilde{\underline{\phi}}_{k+1}^{(\bar{d})} &= \tilde{\underline{\phi}}_k^{(d)} + (k+1)^{-1}H(\tilde{\underline{\phi}}_k)(\underline{t}_{k+1}^{(d)} - \tilde{\underline{\phi}}_k^{(d)}) \end{aligned}\right\} .$$

This pattern will be apparent in the multivariate Normal examples considered later.

A formula for $H(\underline{\phi})$ can be obtained in terms of the complete-data information matrix for $\underline{\phi}$, $I_c(\underline{\phi})$.

Since $\underline{t}^{(d)}$ is a cut, it follows, as in Barndorff-Nielsen (1978, p. 128), that $\underline{t}^{(d)}$ itself has an exponential family distribution. Given the above partitioning, the appropriate score vector for use in (1) or (2) is, therefore,

$$\begin{pmatrix} I_c(\underline{\phi}^{(d)})(\underline{t}^{(d)} - \underline{\phi}^{(d)}) \\ \underline{0} \end{pmatrix} ,$$

where $\underline{\phi}^{(d)} = E(\underline{t}^{(d)})$ and $I_c(\underline{\phi}^{(d)})^{-1}$ is the leading square submatrix of $I_c(\underline{\phi})^{-1}$. Suppose that $\underline{\phi}^{(d)}$ is q-dimensional and that the first q columns of $I_c(\underline{\phi})^{-1}$ are given by

$$\begin{pmatrix} I_c(\underline{\phi}^{(d)})^{-1} \\ I_c^{(\bar{d},d)} \end{pmatrix}$$

Then because of the form of recursion (2), the matrix $H(\underline{\phi})$ in (8) is given by

$$I_c^{(\bar{d},d)} I_c(\underline{\phi}^{(d)})^{-1} .$$

This can be written as

$$-I_c(\bar{d},\bar{d})^{-1}I_c(\bar{d},d) ,$$

where these two matrices come from an appropriate partitioning of $I_c(\underline{\phi})$; see, for instance Barndorff-Nielsen (1978, p.3).

To investigate possible consistency of recursions (1) and (2), nonsingularity of the incomplete-data information matrix for $\underline{\phi}$, $I(\underline{\phi})$, must be sought. It is convenient to use the notation of Hartley and Hocking (1971). Suppose $\underline{t}$ and $\underline{\phi}$ are s-dimensional and that $\underline{t}^{(d)}$ is q-dimensional, with $q < s$. Then, for a certain $q \times s$ matrix, D_d, of zeros and ones,

$$\underline{t}^{(d)} = D_d\underline{t}; \quad \underline{\phi}^{(d)} = D_d\underline{\phi}; \quad I_c(\underline{\phi}^{(d)})^{-1} = D_dI_c(\underline{\phi})^{-1}D_d^T .$$

Also, $D_dD_d^T = I_q$, the $q \times q$ identity matrix. In the special partitioning of $\underline{t}$ considered earlier, the first q columns of D_d make up I_q. For brevity, let

$$\pi(\underline{d}) = g_2(\underline{d}|\underline{\theta}_2) ,$$

the probability of obtaining incompleteness pattern $\underline{d}$.

Then if, whatever $\underline{d}$ is, the corresponding $\underline{t}^{(d)}$ is a cut,

$$I(\underline{\phi}) = \sum_{\underline{d}} \pi(\underline{d})D_d^TI_c(\underline{\phi}^{(d)})D_d .$$

Suppose, for $j = 1,\ldots,s$, π_j denotes the sum of $\pi(\underline{d})$ over all $\underline{d}$ from which the jth component of $\underline{t}$ can be calculated. Then, for nonsingularity of $I(\underline{\phi})$, we require $\pi_j > 0$, for all j.

Whether or not (7) leads to $\sqrt{k}$-consistent estimators depends on the eigenvalues of $I_c(\underline{\phi}_T)^{-1}I(\underline{\phi}_T)$. In the very simple case of the s-dimensional multivariate exponential distribution with independent components, we require that the probability of observation of each component be at least $\frac{1}{2}$.

The rest of the paper will be devoted to the multivariate Normal distribution. Since this belongs to the exponential family we already have some insight into this case. For instance, consistency at even the weakest rate is possible only if there is positive probability of observing each element in the appropriate sufficient statistic $\underline{t}(\underline{x})$. For the multivariate Normal case there must be positive probability of observing each pair of components of $\underline{x}$.

4. MULTIVARIATE NORMAL DATA

In spite of the aforementioned relationship with the exponential family, it is helpful to look at recursions for the more familar parameters, to be denoted by $(\underline{\mu},\Sigma)$. Given complete observations $\underline{x}_1,\ldots,\underline{x}_n$, which are independent and identically distributed $N(\underline{\mu},\Sigma)$, the maximum likelihood estimators of $(\underline{\mu},\Sigma)$ can be generated recursively by

$$\hat{\underline{\mu}}_{k+1} = \hat{\underline{\mu}}_k + (k+1)^{-1}(\underline{x}_{k+1} - \hat{\underline{\mu}}_k) \tag{8}$$

$$\hat{\Sigma}_{k+1} = \hat{\Sigma}_k + (k+1)^{-1}\left\{\frac{k}{k+1}(\underline{x}_{k+1} - \hat{\underline{\mu}}_k)(\underline{x}_{k+1} - \hat{\underline{\mu}}_k)^T - \hat{\Sigma}_k\right\}, \tag{9}$$

starting from $\hat{\underline{\mu}}_0 = \underline{0}$, $\hat{\Sigma}_0 = 0$. These recursions can be obtained from the complete-data version of (1) for the appropriate parameters in the exponential family representation, that is, recursion (6), and subsequent re-expression in terms of $(\underline{\mu},\Sigma)$. The direct version of (1) in terms of $(\underline{\mu},\Sigma)$ gives (8) along with

$$\hat{\Sigma}_{k+1} = \hat{\Sigma}_k + (k+1)^{-1}\{(\underline{x}_{k+1} - \hat{\underline{\mu}}_k)(x_{k+1} - \hat{\underline{\mu}}_k)^T - \hat{\Sigma}_k\}, \tag{10}$$

which is an apparently very minor variant of (9).

Further recourse to the notation of Hartley and Hocking (1971) will help the consideration of the incomplete-data problem.

Suppose $\underline{d}$ represents a particular missing-data pattern with $q(<s)$ observed components in $\underline{x}$. Suppose $\underline{\mu}_d$ $(q \times 1)$ and $\Sigma_d (q \times q)$ are the corresponding mean vector and covariance matrix and let D_d be as in Section 3. Then

$$\underline{\mu}_d = D_d\underline{\mu}; \quad \Sigma_d = D_d\Sigma D_d^T.$$

Let $\underline{\sigma}$ be the vector, of length $\frac{1}{2}s(s+1)$, of the elements of Σ written as a vector. Let $\underline{\sigma}_d$ be the vector, of length $\frac{1}{2}q(q+1)$, containing the elements of Σ_d and let C_d be a matrix of zeros and ones such that

$$\underline{\sigma}_d = C_d\underline{\sigma}.$$

$C_dC_d^T$ will be the identity matrix of order $\frac{1}{2}q(q+1)$.

The complete-data information matrix, per observation, is

$$I_c(\underline{\mu},\underline{\sigma}) = \begin{pmatrix} I_c(\underline{\mu}) & 0 \\ 0 & I_c(\underline{\sigma}) \end{pmatrix},$$

where $I_c(\underline{\mu}) = \Sigma^{-1}$ and $I_c(\underline{\sigma}) = U^{-1}$, where U is a symmetric matrix of order $\frac{1}{2}s(s+1)$. The element in row (i,j) and column (u,v) of U is

$$\sigma_{iu}\sigma_{jv} + \sigma_{iv}\sigma_{ju}, \quad i \leq j,\ u \leq v .$$

The incomplete-data information matrix is also block-diagonal.

$$I(\underline{\mu},\underline{\nu}) = \begin{pmatrix} I(\underline{\mu}) & 0 \\ 0 & I(\underline{\sigma}) \end{pmatrix},$$

in which

$$I(\underline{\mu}) = \sum_{\underline{d}} \pi(\underline{d}) D_{\underline{d}}^T \Sigma_{\underline{d}}^{-1} D_{\underline{d}}$$

$$= \sum_{\underline{d}} \pi(\underline{d}) D_{\underline{d}}^T (D_{\underline{d}} \Sigma\, D_{\underline{d}}^T)^{-1} D_{\underline{d}} \qquad (11)$$

and

$$I(\underline{\sigma}) = \sum_{\underline{d}} \pi(\underline{d}) C_{\underline{d}}^T (C_{\underline{d}} U C_{\underline{d}}^T)^{-1} C_{\underline{d}} . \qquad (12)$$

The explicit forms for (11) and (12) allow recursion (1) to be considered without the need for numerical integration. A recursion using sample information could also be considered but it will be more complicated since the block-diagonal form noted above will not obtain. This is why this method is not included in the simulation study in Section 5.

Calculation of the score vector associated with a single observation also follows Hartley and Hocking (1971). Suppose $g_1(\underline{x}_{\underline{d}})$ denotes the p.d.f. for an observed value corresponding to missing-data pattern $\underline{d}$. Then

$$\frac{\partial}{\partial \underline{\mu}} \log g_1(\underline{x}_{\underline{d}}) = D_{\underline{d}}^T \Sigma_{\underline{d}}^{-1} (\underline{x}_{\underline{d}} - \underline{\mu}_{\underline{d}}) = \underline{s}_{\mu}(\underline{x}_{\underline{d}}), \text{ say;}$$

$$\frac{\partial}{\partial \underline{\sigma}} \log g_1(\underline{x}_{\underline{d}}) = C_{\underline{d}}^T (C_{\underline{d}} U C_{\underline{d}})^{-1} (\underline{s}_{\underline{d}} - \underline{\sigma}_{\underline{d}}) = \underline{s}_{\sigma}(\underline{x}_{\underline{d}}), \text{ say.}$$

In the second equation, $\underline{s}_{\underline{d}} = C_{\underline{d}}\underline{s}$, where $\underline{s}$ is the suitably-ordered vector version of $(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^T$.

For recursion (2) we must premultiply this score vector by $I_c(\underline{\mu},\underline{\sigma})^{-1}$. In the case of a complete observation, for which C_d is the identity matrix of order $\frac{1}{2}s(s+1)$, we obtain the recursions (8) and (10). Even for an incomplete observation we obtain an appealing result. Partition the score vector, as above, as

$$\underline{S}(\underline{x}_d,\underline{\mu},\underline{\sigma}) = \begin{pmatrix} \underline{S}_\mu(\underline{x}_d) \\ \underline{S}_\sigma(\underline{x}_d) \end{pmatrix} .$$

Then

$$I_c(\underline{\mu},\underline{\sigma})^{-1}\underline{S}(\underline{x}_d,\underline{\mu},\underline{\sigma}) = \begin{pmatrix} \Sigma\, \underline{S}_\mu(\underline{x}_d) \\ U\, \underline{S}_\sigma(\underline{x}_d) \end{pmatrix}$$

Suppose, without loss of generality, the first q components of $\underline{x}$ are observed and are denoted by $\underline{x}_1$. Write $\underline{x}^T = (\underline{x}_1^T\ \underline{x}_2^T)$ and partition $\underline{\mu}$ and Σ correspondingly. Then

$$\Sigma\, \underline{S}_\mu(\underline{x}_d) = \begin{pmatrix} \Sigma_{11}\Sigma_{12} \\ \Sigma_{12}^T\Sigma_{22} \end{pmatrix}\begin{pmatrix} I_q \\ 0 \end{pmatrix}\Sigma_{11}^{-1}(I_q\ 0)\begin{pmatrix} \underline{x}_1 - \underline{\mu}_1 \\ \underline{x}_2 - \underline{\mu}_2 \end{pmatrix} = \begin{pmatrix} \underline{x}_1 - \underline{\mu}_1 \\ \Sigma_{12}^T\Sigma_{11}^{-1}(\underline{x}_1 - \underline{\mu}_1) \end{pmatrix} . \quad (13)$$

It is more helpful to express the vector $U\underline{S}_\sigma(\underline{x}_d)$ as a $k \times k$ symmetric matrix at this point. If S_{11} denotes the matrix $(\underline{x}_1 - \underline{\mu}_1)(\underline{x}_1 - \underline{\mu}_1)^T$ then we obtain

$$\begin{pmatrix} S_{11} - \Sigma_{11} & (S_{11} - \Sigma_{11})\Sigma_{11}^{-1}\Sigma_{12} \\ \Sigma_{12}^T\Sigma_{11}^{-1}(S_{11} - \Sigma_{11}) & \Sigma_{12}^T\Sigma_{11}^{-1}(S_{11} - \Sigma_{11})\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix} \quad (14)$$

Expressions (13) and (14) can now be used in the second term on the right hand side of recursion (2).

One disadvantage of recursion (2) is that a different form of (13) and (14) is required for each pattern, $\underline{d}$, of missing values. Expression (13), however, can be written as

$$\begin{pmatrix} \underline{x}_1 - \underline{\mu}_1 \\ \tilde{\underline{x}}_2 - \underline{\mu}_2 \end{pmatrix} , \qquad (15)$$

where $\tilde{\underline{x}}_2 = \underline{\mu}_2 + \Sigma_{12}^T \Sigma_{11}^{-1}(\underline{x}_1 - \underline{\mu}_1)$, the regression function for $\underline{x}_2$ given $\underline{x}_1$.

If, at stage $k + 1$, we can observe $\underline{x}_1(k + 1)$ and have obtained, currently, $\tilde{\underline{\mu}}_k$ and $\tilde{\Sigma}_k$, then, with these quantities substituted in the right hand side of (15), we obtain a regression imputation $\tilde{\underline{x}}_2(k + 1)$ for the missing values $\underline{x}_2(k + 1)$. The use of (13) in recursion (2) is then equivalent to (8) using the imputed $\tilde{\underline{x}}_2(k + 1)$. The use of (15) removes the need for separate versions of recursion (2) for $\underline{\mu}$. Hope that the completed $\underline{x}_{k+1}$ can be used in (10) instead of (14) are ill-founded. If $\tilde{\underline{x}}_2$ from (15) is used in this way then, instead of (14), we obtain a matrix which differs in the bottom right hand block. This block is

$$\Sigma_{12}^T \Sigma_{11}^{-1}(S_{11} - \Sigma_{11})\Sigma_{11}^{-1}\Sigma_{12} - (\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1}\Sigma_{12}) .$$

Since $\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1}\Sigma_{12}$ is nonnegative definite, this block is "smaller than it should be", and the resulting recursion would lead to "negative bias" in the estimator of Σ_{22}. This is because the imputed $\tilde{\underline{x}}_2$ is a conditional expected value and the residual variance is ignored; see also Beale and Little (1975). Perhaps a more satisfactory approach based on recursive imputation and use of (8) and (10) is to use, not (15), but a pseudo-random imputation

$$\tilde{\underline{x}}_2 = \underline{\mu}_2 + \Sigma_{12}^T \Sigma_{11}^{-1}(\underline{x}_1 - \underline{\mu}_1) + \underline{\varepsilon}_2 , \qquad (16)$$

where

$$\underline{\varepsilon}_2 \sim N(\underline{0}, \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1}\Sigma_{12}) .$$

Such imputation and updating does not correspond exactly to (13) and (14) but, conditionally on $\underline{x}_1$ as well as Σ_{11}, Σ_{12} and Σ_{22}, the updating terms do, on average, give (13) and (14).

We now illustrate these approaches for the case of bivariate data.

5. ILLUSTRATION FOR BIVARIATE NORMAL DATA

We consider the special case of bivariate Normal data in which the first component of any observation is always available but the second component may be missing, with probability $(1 - \pi)$. This is special in that explicit solutions are available for the maximum likelihood estimates of the parameters, because of the "nested" missing-data structure; see Morrison (1971), Rubin (1974) and Hocking and Marx (1979). It will be possible, therefore, to compare our recursive methods easily with non-recursive maximum likelihood.

Recursion (2) takes the following form. At stage $(k + 1)$, if $\underset{\sim}{x}_{k+1}$ is complete then (8) and (10) are used to update. If only the first component $x_1(k + 1)$ is given, then (8) is used to update $\tilde{\underset{\sim}{\mu}}_k$, using the regression imputation $\tilde{x}_2(k + 1)$ for $x_2(k + 1)$, where

$$\tilde{x}_2(k + 1) = \tilde{\mu}_2(k) + \tilde{\beta}_k \delta_{k+1} , \tag{17}$$

in which $\tilde{\beta}_k = \tilde{\Sigma}_{12}^{(k)} / \tilde{\Sigma}_{11}^{(k)}$ and

$$\delta_{k+1} = x_1(k + 1) - \tilde{\mu}_1(k) .$$

$\tilde{\Sigma}_k$ is updated according to

$$\tilde{\Sigma}_{11}(k + 1) = \tilde{\Sigma}_{11}(k) + (k + 1)^{-1} \Delta_{k+1}$$

$$\tilde{\Sigma}_{12}(k + 1) = \tilde{\Sigma}_{12}(k) + (k + 1)^{-1} \tilde{\beta}_k \Delta_{k+1}$$

$$\tilde{\Sigma}_{22}(k + 1) = \tilde{\Sigma}_{22}(k) + (k + 1)^{-1} \tilde{\beta}_k^2 \Delta_{k+1} ,$$

where $\Delta_{k+1} = \{x_1(k + 1) - \tilde{\mu}_1(k)\}^2 - \tilde{\Sigma}_{11}(k)$.

In order to illustrate the resulting bias in estimating Σ_{22}, a regression-imputation recursion was tried, using (8) and (10) along with the imputation formula (17). Also, a regression-imputation with residual was studied, in which a pseudo-random ϵ_{k+1} was added on to (17) before it was plugged into (8) and (10).

$$\epsilon_{k+1} \sim N(0, \tilde{\Sigma}_{22}(k) - \tilde{\Sigma}_{12}(k)^2 / \tilde{\Sigma}_{11}(k)) .$$

Another characteristic of this example is that recursion (1) will be easy to apply because of the explicit invertability of the incomplete-data information matrix, $I(\underset{\sim}{\mu}, \underset{\sim}{\sigma})$.

From Section 4 it is clear that the two blocks, $I(\underline{\mu})$ and $I(\underline{\sigma})$, in the block diagonal form of $I(\underline{\mu},\underline{\sigma})$ may be considered separately. For this example

$$I(\underline{\mu}) = \pi\Sigma^{-1} + (1 - \pi)\underline{v}\underline{v}^T ,$$

where $\underline{v}^T = (\Sigma_{11}^{-1/2},0)$, so that

$$I(\underline{\mu})^{-1} = \pi^{-1}\Sigma - \pi^{-1}(1 - \pi)\Sigma\underline{v}\underline{v}^T\Sigma = \Sigma + \pi^{-1}(1 - \pi)\begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{22} - \Sigma_{12}^2/\Sigma_{11} \end{pmatrix} . \qquad (18)$$

Similarly, if $\underline{\sigma}^T = (\Sigma_{11},\Sigma_{12},\Sigma_{22})$,

$$I(\underline{\sigma}) = \pi U^{-1} + (1 - \pi)\underline{w}\underline{w}^T ,$$

where U is as described in Section 4 and $\underline{w}^T = (1/\sqrt{2}\ \Sigma_{11},0,0)$. Thus

$$I(\underline{\sigma})^{-1} = \pi^{-1}U - \pi^{-1}(1 - \pi)U\underline{w}\underline{w}^TU , \qquad (19)$$

which can be rearranged to some extent for easiest programming.

Expressions (18) and (19) combine with the score vectors to give very simple versions of recursion (1).

For this example, checking of the eigenvalue condition for the optimal convergence rate of recursion (1) is quite easy.

$$A(\underline{\mu},\underline{\sigma}) = I_c(\underline{\mu},\underline{\sigma})^{-1}I(\underline{\mu},\underline{\sigma}) = \begin{pmatrix} I_c(\underline{\mu})^{-1}I(\underline{\mu}) & 0 \\ 0 & I_c(\underline{\sigma})^{-1}I(\underline{\sigma}) \end{pmatrix}$$

$$I_c(\underline{\mu})^{-1}I(\underline{\mu}) = \pi I_2 + (1 - \pi)\Sigma\begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \beta(1 - \pi) & \pi \end{pmatrix} ,$$

where $\beta = \Sigma_{12}/\Sigma_{11}$. Also,

$$I_c(\underline{\sigma})^{-1}I(\underline{\sigma}) = \pi I_3 + (1 - \pi)U\underline{w}\underline{w}^T = \begin{pmatrix} 1 & 0 & 0 \\ \beta(1 - \pi) & \pi & 0 \\ \beta^2(1 - \pi) & 0 & \pi \end{pmatrix} .$$

Thus, the distinct eigenvalues of $A(\underline{\mu},\underline{\sigma})$ are 1 and π, so that $\sqrt{k}$-consistency obtains if $\pi > \frac{1}{2}$. Note that, unless $\beta = 0$, A is not symmetric. If $\beta = 0$, the asymptotic covariance matrix, B, for $\{\sqrt{k}\ \tilde{\underline{\theta}}_k\}$ is

$$\mathrm{diag}\{1, \gamma, 2, \gamma, 2\gamma\} ,$$

where $\gamma = \pi(2\pi - 1)^{-1}$. For $\beta \neq 0$, the linear equations (3) must be solved for B.

It should be stressed that we are unusually lucky in this example to be able to use recursion (1) and the proper maximum likelihood procedure so easily. In many examples the versions of recursion (2) should dominate as far as computation time is concerned. Apart from this consideration, however, the following comparisons should give a guide to relative effectiveness of the different methods.

In the simulations, bivariate Normal distributions with zero means and unit component variances are used. Two values of the correlation coefficient, $\rho = 0$ and $\rho = 0.6$ are considered. In each simulation the first 10 observations were complete. Thereafter, the second component of each observation was treated as missing with probability $1 - \pi = 3/4$, independently for each observation. Since $\pi < \frac{1}{2}$ this means that we cannot expect the optimal rate of consistency for recursion (2). Extensive results are given in an unpublished M.Sc report at the State University of New York at Albany by J-M. Jiang, but the results reported here were obtained on a VAX/780 computer at the University of Wisconsin in Madison. The estimation procedures are denoted as follows.

I. Standard recursion (2).

II. Recursive procedure using (8) and (10) with regression imputation (17).

III. As II but incorporating pseudo-random residuals into the imputation.

IV. Standard recursion (1).

V. True maximum likelihood from the available data.

VI. Recursive treatment of the complete data.

Table 1 gives root mean-squared errors (RMSE's) for the five parameters from batches of 100 simulations, each with total sample size 100.

The results in Section (a) of the table correspond to the use of recursions like (10) for Σ, those in Section (b) to (9). Thus, VI in Section (b) corresponds to maximum likelihood estimation on the complete data. The natural superiority of VI(b) is clear from the results. Other major features of the table are as follows.

TABLE 1

RMSE's FOR BIVARIATE NORMAL PARAMETERS 100 SIMULATIONS, EACH OF 100 OBSERVATIONS

Method	$\rho = 0.0$					$\rho = 0.6$				
	μ_1	μ_2	Σ_{11}	Σ_{12}	Σ_{22}	μ_1	μ_2	Σ_{11}	Σ_{12}	Σ_{22}
(a)										
I	0.106	0.207	0.166	0.265	0.320	0.108	0.168	0.170	0.230	0.314
II	0.106	0.207	0.166	0.265	0.619	0.108	0.168	0.170	0.230	0.461
III	0.106	0.254	0.166	0.310	0.442	0.108	0.211	0.170	0.269	0.374
IV	0.106	0.171	0.166	0.210	0.270	0.108	0.174	0.170	0.190	0.256
V	0.106	0.169	0.157	0.191	0.254	0.108	0.147	0.162	0.183	0.253
VI	0.106	0.093	0.166	0.106	0.136	0.108	0.098	0.170	0.128	0.138
(b)										
I	0.106	0.207	0.157	0.258	0.316	0.108	0.168	0.162	0.224	0.309
II	0.106	0.207	0.157	0.258	0.626	0.108	0.168	0.162	0.224	0.469
III	0.106	0.248	0.157	0.295	0.382	0.108	0.206	0.162	0.257	0.342
IV	0.106	0.171	0.166	0.211	0.270	0.108	0.174	0.170	0.190	0.256
VI	0.106	0.093	0.157	0.103	0.130	0.108	0.098	0.162	0.124	0.132

(i) Qualitatively, the picture is the same for $\rho = 0.0$ and $\rho = 0.6$.

(ii) Apart from estimation of Σ_{22} by method II (regression imputation) the results in (b) are better than those in (a).

(iii) Particularly in (a), results in IV (recursion (1)) are better than those in I-III (recursion (2)).

(IV) Method IV in (b) does almost as well as method V, the actual maximum likelihood.

(V) Method III (regression imputation plus residual) is the least efficient of the variants of recursion (1) for estimating μ_2 but, as expected, it is better than method II for the estimation of Σ_{22}. This illustrates the fact that method II should underestimate Σ_{22} on average, according to previous remarks. To check this the frequency of underestimation of Σ_{22} was obtained for each method, as reported in Table 2. The results point to the strong negative bias in method II.

The whole exercise was carried out also for samples of size 50 with qualitatively the same results as above.

We have already pointed out the criticism that the results for the recursive methods are order-dependent, an obvious disadvantage for the treatment of a finite data-set. A very limited assessment of the order effect was obtained by comparing the results gathered from one ordering of the data and its reverse. (Consideration of all possible orderings would be out of the question although a more ambitious project would be to compare a moderately large number of "random" orderings.) Table 3 gives, from the 100 simulations, the root mean squared difference in parameter estimates obtained by the two orderings for methods I-IV. For some of the parameter estimates, as indicated, the ordering does not affect the value. The most variation occurs with methods I, II and III particularly in the estimation of Σ_{22}. The results favor section (b) of the table very slightly.

Further numerical investigation is recorded in the M.Sc report of J-M. Jiang. In particular he investigated the effect of a double-pass through the data. He considered a pass through (forwards) followed by another pass with the reverse ordering (backwards). He also considered the corresponding backwards-forwards double-pass. Among the conclusions

TABLE 2

FREQUENCIES (OUT OF 100) OF UNDERESTIMATION OF Σ_{22}

	Method					
	I	II	III	IV	V	VI
(ρ = 0.0)						
(a)	54	99	59	45	58	44
(b)	56	99	64	45	58	53
(ρ = 0.6)						
(a)	55	94	50	45	58	40
(b)	58	94	55	45	58	47

TABLE 3

RMS DIFFERENCES IN PARAMETER ESTIMATES BETWEEN ONE ORDERING OF THE DATA AND ITS REVERSE

	$\rho = 0.0$					$\rho = 0.6$				
Method	μ_1	μ_2	Σ_{11}	Σ_{12}	Σ_{22}	μ_1	μ_2	Σ_{11}	Σ_{12}	Σ_{22}
(a)										
I	0	0.094	0.004	0.127	0.154	0	0.078	0.004	0.100	0.141
II	0	0.094	0.004	0.127	0.090	0	0.078	0.004	0.100	0.127
III	0	0.167	0.004	0.212	0.294	0	0.131	0.004	0.159	0.287
IV	0	0.041	0.004	0.074	0.059	0	0.083	0.004	0.052	0.073
(b)										
I	0	0.094	0	0.119	0.145	0	0.077	0	0.093	0.132
II	0	0.094	0	0.119	0.086	0	0.077	0	0.093	0.120
III	0	0.160	0	0.193	0.247	0	0.126	0	0.144	0.249
IV	0	0.041	0.005	0.074	0.059	0	0.082	0.004	0.052	0.073

were the following, both of which support intuition.

(i). Double-passes produced smaller RMSE's than single passes.

(ii). The differences between the forwards-backwards double-pass and the backwards-forwards double-pass were small and even more so than the effects exemplified in Table 3 for different single-pass procedures.

Jiang also considered the performance of these recursive methods for the simpler problem of estimating the mixing weight in a mixture of two known densities. He found, in particular, that there the "order effect" seemed less substantial than the among-replications effect. Comparison of Tables 1 and 3 suggests that this may not be the case in our example, particular in the estimation of Σ_{22}.

6. DISCUSSION

Although this paper has concentrated on missing-value problems, the general procedures defined by (1) and (2) could be applied to many other incomplete data problems. Some were looked at by Titterington (1982) and other illustrations could be drawn from Dempster, Laird and Rubin (1977). In missing value problems, recursion (1) is comparatively easy to use and particularly so in the illustration of Section 5. It seems likely that, in spite of their inferior asymptotic performance, recursion (2) and the imputation-based versions thereof will be appealing for their comparative simplicity in application. As far as imputation is concerned, Section 5 illustrates once more the disadvantages that can arise from mean-imputation as compared with pseudo-random imputation, a point stressed by Sedransk and Titterington (1980). They also describe methods, such as the hot-deck, for dealing with incomplete data from sample surveys. It is hoped that this is one area in particular where these recursive procedures will be very useful.

Acknowledgements

I am grateful to Professor A. P. Dawid for helpful comments and to Professor V. Fabian for pointing out the reference to Walk. Part of the work was carried out at the State University of New York at Albany, with the encouragement of Professor J. Sedransk and with the support of the Bureau of the Census, under contract number JSA 80-12.

REFERENCES

Barndorff-Nielsen, O. (1978). Information and Exponential Families. Wiley: New York.

Beale, E. M. L. and Little, R. J. A. (1975). Missing values in multivariate analysis. J. R. Statist. Soc. B, 37, 129-145.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. R. Statist. Soc. B, 39, 1-38.

Fabian, V. (1968). On asymptotic normality in stochastic approximation. Ann. Math. Statist., 39, 1327-1332.

Hartley, H. O. and Hocking, R. R. (1971). The analysis of incomplete data. Biometrics, 27, 783-823.

Hocking, R. R. and Marx, D. L. (1979). Estimation with incomplete data: an improved computational method and the analysis of nested data. Commun. Statist., A8, 1155-1181.

Little, R. J. A. (1978). Consistent regression methods for discriminant analysis with incomplete data. J. Amer. Statist. Assoc., 73, 319-322.

Morrison, D. F. (1971). Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data. J. Amer. Statist. Assoc., 66, 602-604.

Rubin, D. B. (1974). Characterizing the estimation of parameters in incomplete-data problems. J. Amer. Statist. Assoc., 69, 467-474.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63, 581-592.

Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. Ann. Math. Statist., 29, 373-405.

Sedransk, J. and Titterington, D. M. (1980). Non-response in sample surveys. Report to U. S. Bureau of Census.

Titterington, D. M. (1982). Recursive parameter estimation using incomplete data. Univ. of Wisc. MRC Tech. Report #2376.

Walk, H. (1977). An invariance principle for the Robbins-Monro process in a Hilbert space. Z. Wahrscheinlechkeitsth., 23, 135-150.

DMT/JMJ/ed

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2427	2. GOVT ACCESSION NO. AD-A124358	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Recursive Estimation Procedures for Missing-Data Problems		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. M. Titterington and J-M. Jiang		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE September 1982 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Parameter estimation, missing values, recursive estimation, EM algorithm, exponential family, multivariate Normal, stochastic approximation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Titterington (1982 proposed recursive methods for dealing with incomplete data. The present paper concentrates on versions of these for multiparameter problems involving missing data. Theorems are outlined from which asymptotic properties of the recursive procedures can be established and versions of the recursions are written down for problems in which the missing data are missing at random. After illustration with exponential family models, the case of multivariate Normal data is considered in detail. Numerical comparisons of the various methods are obtained using bivariate Normal data.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

DATE
FILMED

3-83

DTIC