PSYCHOMETRIKA-VOL. 47. NO. 1.
MARCH, 1982

EM ALGORITHMS FOR ML FACTOR ANALYSIS

DONALD B. RUBIN AND DOROTHY T. THAYER

EDUCATIONAL TESTING SERVICE

The details of EM algorithms for maximum likelihood factor analysis are presented for both the exploratory and confirmatory models. The algorithm is essentially the same for both cases and involves only simple least squares regression operations; the largest matrix inversion required is for a $q \times q$ symmetric matrix where q is the matrix of factors. The example that is used demonstrates that the likelihood for the factor analysis model may have multiple modes that are not simply rotations of each other; such behavior should concern users of maximum likelihood factor analysis and certainly should cast doubt on the general utility of second derivatives of the log likelihood as measures of precision of estimation.

Key words: factor analysis, EM algorithms, maximum likelihood.

Introduction

Maximum likelihood factor analysis can be conceptualized as maximum likelihood estimation in a multivariate normal model with missing data [Dempster, Laird, & Rubin (1977) section 4.7]. Consequently, there exists a corresponding EM algorithm to find maximum likelihood estimates. This algorithm is iterative, and each cycle, which consists of an E step followed by an M step, increases the likelihood of the parameters. We here explicitly define the E and M steps of the algorithm and present simple matrix expressions for carrying out the computations.

The general theory of EM algorithms given in Dempster, Laird, and Rubin [1977] proves not only that each iteration of EM increases the likelihood, even if starting from a point where the likelihood is not convex, but also that if an instance of the algorithm converges, it converges to a (local) maximum of the likelihood. Experience with EM algorithms suggests that, although the rate of convergence measured by number of steps can be slow, they reliably converge in a wide range of examples. Another advantage of EM algorithms, such as those for factor analysis, is that each iteration is simple to program and computationally inexpensive. Even for confirmatory factor analysis with correlations among factors to be estimated and *a priori* zeros in the factor loadings, each iteration of EM involves only simple matrix manipulations with the most difficult task being the inversion of a $q \times q$ symmetric index, where q is the number of factors. A final advantage of EM algorithms is that they climb the hill of likelihood on which the starting point is located without leaping over valleys in the likelihood; that is, there is a continuous path in the parameter space from the starting point to the stopping point along which the likelihood monotonically increases. This property is important when searching for multiple maxima, which apparently can easily exist in the factor analysis model as the example we present illustrates.

Notation for the Factor Analysis Model

Let Y be the $n \times p$ observed data matrix and Z be the $n \times q$ unobserved factor-score matrix, $q < p$. The rows of (Y, Z) are independently and identically distributed. The margin-

Requests for reprints should be sent to Donald B. Rubin, Educational Testing Service, Rosedale Road, Princeton, New Jersey, 08540.

0033-3123/82/0100-3079$00.75/0

al distribution of each row of Z is normal with mean $(0, \ldots, 0)$, variance $(1, \ldots, 1)$ and correlation matrix R. The conditional distribution of the ith row of Y, Y_i, given Z is normal with mean $\alpha + Z_i \beta$ and residual covariance $\tau^2 = \mathrm{diag}(\tau_1^2, \ldots, \tau_p^2)$, where Z_i is the ith row of Z.

Note that given the factors, the variables are independent. This assumption of conditional independence is the key one in the factor analysis model.

The parameters to be estimated in general consist of α, β, τ^2, and R. Since the marginal distribution of each row of Y is normal with mean α and covariance $\tau^2 + \beta' R \beta$, the maximum likelihood estimate (m.l.e.) of α is $\bar{Y}$. Thus for purposes of maximum likelihood estimation, we may replace $(Y_{ij} - \alpha_j)$ by $(Y_{ij} - \bar{Y}_j)$ and consider the parameters to be (β, τ^2, R). For notational simplicity, we simply suppose $\bar{Y}_j = 0$ (i.e., the observed variables have been centered at the sample means).

Consequently, the marginal distribution of Y given β, τ^2, R is normal with mean 0 and covariance matrix $\tau^2 + \beta' R \beta$. The resulting log likelihood to be maximized is

$$\begin{aligned}\mathrm{LL}(\tau^2, \beta, R) &= -\frac{n}{2} \log \det(\tau^2 + \beta' R \beta) - \frac{1}{2} \sum_1^n Y_i(\tau^2 + \beta' R \beta)^{-1} Y_i' \\ &= -\frac{n}{2} \log \det(\tau^2 + \beta' R \beta) - \frac{n}{2} \mathrm{tr}[C_{yy}(\tau^2 + \beta' R \beta)^{-1}] \qquad (1)\end{aligned}$$

where C_{yy} is the sample covariance of Y.

Since $\mathrm{LL}(\tau^2, \beta, R)$ is viewed as a function of τ^2, β and R for fixed C_{yy}, maximizing $\mathrm{LL}(\tau^2, \beta, R)$ is equivalent to maximizing $(2/n)\,\mathrm{LL}(\tau^2, \beta, R) + \log \det(C_{yy}) + p$ which equals

$$f(\tau^2, \beta, R) = \log \det[C_{yy}(\tau^2 + \beta' R \beta)^{-1}] + p - \mathrm{tr}[C_{yy}(\tau^2 + \beta' R \beta)^{-1}];$$

$f(\tau^2, \beta, R)$ appears in Jöreskog [1969].

The regression coefficient matrix β is commonly called the factor-loading matrix and the residual variances in τ^2 are commonly called the uniquenesses. Three common cases are defined by restrictions on the parameters:

Case 1: $R = I$ (orthogonal factors) and unrestricted β;
Case 2: $R = I$ and *a priori* zeroes in β;
Case 3: R free to be estimated and *a priori* zeroes in β.

Case 1 is sometimes referred to as exploratory factor analysis, and Cases 2 and 3 are sometimes referred to as confirmatory factor analysis.

Computation of M.L.E. Using EM—Overview

The EM algorithm treats the factor matrix Z as missing data, and iteratively maximizes the likelihood supposing Z were observed. If Z were observed, the likelihood would be

$$\left[2\pi \prod_{j=1}^{p} \tau_j^2\right]^{-\frac{n}{2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(Y_{ij} - Z_i \beta_j)^2}{\tau_j^2}\right] \times [2\pi \det R]^{-\frac{n}{2}} \exp\left[-\frac{1}{2} \sum_{i=1}^{n} Z_i R^{-1} Z_i'\right]. \qquad (2)$$

The EM algorithm uses the complete-data likelihood (2) in order to maximize the actual likelihood of the parameters given Y, that is, in order to maximize expression (2) integrated over the missing data Z, which is equal to $\exp[\mathrm{LL}(\tau^2, \beta, R]$.

There are two steps in each cycle of the EM algorithm. First, in the E step, we find the expectation of the logarithm of the complete-data likelihood given the observed data Y and the current estimated value of the parameter. Hence, the E-step requires us to find the expected value (over the distribution of Z given Y and parameters) of:

$$-\frac{n}{2}\sum_{j=1}^{p}\log\tau_j^2-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\frac{(Y_{ij}-Z_i\beta_j)^2}{\tau_j^2}-\frac{n}{2}\log(\det R)-\frac{1}{2}\sum_{i=1}^{n}Z_iR^{-1}Z_i'. \quad (3)$$

The second step of the EM algorithm, the M-step, requires us to maximize the expected log likelihood just as if it were based on complete data. This maximization yields the next value of the parameter. Using the new value of the parameter, we compute the next E-step and continue. As mentioned in the introduction, the general theory of EM algorithms given in Dempster, Laird, and Rubin [1977] proves that each iteration increases the likelihood.

The E-Step

Finding the expectation of (3) given Y and β, τ^2, R requires finding the expectation of the following sufficient statistics (where we recall that the Y_i have been centered at the sample mean $\bar{Y}$):

$$C_{yy}=\sum_{1}^{n}\frac{Y_i'Y_i}{n}, \qquad \text{a } p\times p \text{ observed matrix}$$

$$C_{yz}=\sum_{1}^{n}\frac{Y_i'Z_i}{n}, \qquad \text{a } p\times q \text{ matrix}$$

$$C_{zz}=\sum_{1}^{n}\frac{Z_i'Z_i}{n}, \qquad \text{a } q\times q \text{ matrix.} \quad (4)$$

Given τ^2, R and β, the (Y_i, Z_i) are i.i.d. $(p + q)$-variate normal. Thus, given τ^2, R and β, the conditional distribution of Z_i given Y_i is q-variate normal with mean δY_i and covariance Δ, where the regression coefficient δ and residual covariance matrix Δ are given by:

$$\delta=(\tau^2+\beta'R\beta)^{-1}(\beta'R)$$
$$\Delta=R-(R\beta)(\tau^2+\beta'R\beta)^{-1}(\beta'R). \quad (5)$$

If *a priori* $R = I$, (5) reduce to

$$\delta=(\tau^2=\beta'\beta)^{-1}\beta'$$
$$\Delta=I-\beta(\tau^2+\beta'\beta)^{-1}\beta'. \quad (6)$$

Thus, the conditional expectations of the sufficient statistics (4) given parameters τ^2, β, R observed data $Y_1, \ldots., Y_n$ are:

$$E(C_{yy}\mid Y,\phi)=C_{yy}$$
$$E(C_{yz}\mid Y,\phi)=C_{yy}\delta$$
$$E(C_{zz}\mid Y,\phi)=\delta'C_{yy}\delta+\Delta. \quad (7)$$

Equations (7) and (5) define the E step from observed data Y and current estimated parameters τ^2, β and R. Equations (7) and (6) define the E-step in the orthogonal factor model from observed data Y and current estimated parameters τ^2 and β.

Although the inversion of a $p \times p$ symmetric matrix appears in (5) and (6), in fact, only a $q \times q$ symmetric matrix need be inverted. Using Woodbury's identity we have:

$$(\tau^2 + \beta' R \beta)^{-1} = \tau^{-2} - (\tau^2 \beta')(R^{-1} + \beta \tau^{-2} \beta')^{-1}(\beta \tau^{-2}) \tag{8}$$

and when $R = I$,

$$(\tau^2 + \beta' \beta)^{-1} = \tau^{-2} - (\tau^{-2} \beta')(I + \beta \tau^{-2} \beta')^{-1}(\beta \tau^{-2}). \tag{9}$$

The M-Step Case 1—Unrestricted β and $R = I$

The M-step of the EM algorithm obtains the next value of parameters by maximizing the expected log likelihood found in the E – step, just as if the expected values of the sufficient statistics were the observed values of the sufficient statistics. First suppose that β is unrestricted and that $R = I$.

By (7) and standard regression arguments, the m.l.e.'s of β and τ^2 are given by

$$\begin{aligned} \beta^* &= [\delta' C_{yy} \delta + \Delta]^{-1} (C_{yy} \delta)' \\ \tau^{*2} &= \operatorname{diag}\{C_{yy} - C_{yy} \delta [\delta' C_{yy} \delta + \Delta]^{-1} (C_{yy} \delta)'\}. \end{aligned} \tag{10}$$

Using the estimators in (10) in place of β and τ^2 in (6) gives the next values of δ and Δ; these new values of δ and Δ are then used in the right hand sides of (10) to obtain new values of β and τ^2, and so forth.

The iteration between (10) and (6) is almost the method for maximum likelihood factor analysis described by Lawley and Maxwell [1963]. If in (10) we replace $\delta' C_{yy} \delta + \Delta$, which is the conditional expectation of C_{zz} given Y, β and τ^2, by its unconditional expectation, i.e., the identity matrix, we obtain

$$\tilde{\beta}^* = (C_{yy} \delta)' = \beta (\tau^2 + \beta' \beta)^{-1} C_{yy} \qquad \text{(Lawley and Maxwell equation 2.6)}$$

$$\tilde{\tau}^{*2} = \operatorname{diag}(C_{yy} - \beta^{*\prime} \beta^*) \qquad \text{(Lawley and Maxwell equation 2.8).}$$

These Lawley and Maxwell equations do not define an EM algorithm, and so this method does not necessarily enjoy the general convergence properties of EM algorithms. Lawley and Maxwell [1971, 2nd ed., chap. 7] no longer propose their 1963 method.

The M-Step Cases 2 and 3 : a priori Zeroes in β

When there are *a priori* zeroes in β, different Y-variables have different collections of "relevant" factors (that is, factors having nonzero β's). Because the Y_j $j = 1, \ldots, p$ given Z are conditionally independent, we can deal with each Y variable separately, although in practice all Y-variables with the same pattern of *a priori* zeroes in β will be handled together.

Consider the jth Y-variable with regression coefficient β_j on Z. Reorder the factors so that $\beta_j = (\beta_{lj}, \beta_{0j})'$ where β_{0J} are *a priori* zero and β_{1j} are to be estimated. Similarly partition the matrices $(\delta' C_{yy} \delta + \Delta)$ and $\delta' C_{yy}$ so that $(\delta' C_{yy} \delta + \Delta)_{1j}$ and $(\delta' C_{yy})_{1j}$ correspond to the factors with nonzero coefficients for the jth variable; in case 2 when $R = I$, δ and Δ are defined by (6), and in case 3 when R is to be estimated, δ and Δ are defined by (5). The m.l.e. of β_j from estimated sufficient statistics is

$$\beta_j^* = (\beta_{1j}^*, \beta_{0j}^*)' \tag{11}$$

where

$$\beta_{0j}^* = (0, \ldots, 0)$$

and

$$\beta^{*\prime}_{1j} = [(\delta' C_{yy}\delta + \Delta)_{1j}]^{-1}(\delta' C_{yy})_{1j}$$

and the m.l.e. of τ_j^2 from estimated sufficient statistics is

$$\tau_j^{*2} = C_{yyj} - (C_{yy}\delta)_{1j}[(\delta' C_{yy}\delta + \Delta)_{1j}]^{-1}(\delta' C_{yy})_{1j}$$

where C_{yyj} is the jth diagonal element of C_{yy}. Then, $\beta^* = (\beta_1^*, \ldots, \beta_p^*)$ and $\tau^{*2} = (\tau_1^{*2}, \ldots, \tau_p^{*2})$.

In case 2, the M-Step leaves R fixed at I. In case 3, with R to be estimated, the m.l.e. of R at the M-step is simply the expectation of C_{zz} normed to be a correlation matrix:

$$R^* = \text{normed}[\delta' C_{yy}\delta + \Delta]. \tag{12}$$

Consider an example with 9 variables, 4 factors, and 2 patterns of *a priori* zeroes among the coefficients: Suppose that variables 1–4 have *a priori* zero coefficients on factor 4, variables 5–9 have *a priori* zero coefficients on factor 3, and otherwise there are no restrictions. Hence, in the notation of (11), for $j = 1, \ldots, 4, \beta^*_{1j}$ consists of the coefficients on factors 1, 2, 3, β^*_{0j} consists of the zero coefficient on factor 4, $(\delta' C_{yy}\delta + \Delta)_{1j}$ is the 3×3 submatrix of $\delta' C_{yy}\delta + \Delta$ consisting of its first three rows and columns, and $(\delta' C_{yy})_{1j}$ is the 3×1 submatrix of $\delta' C_{yy}$ consisting of rows 1, 2, 3 in column j of $\delta' C_{yy}$. For $j = 5, \ldots, 9, \beta^*_{1j}$ consists of the coefficients on factors 1, 2, 4, β^*_{0j} consists of the zero coefficient on factor 3, $(\delta' C_{yy}\delta + \Delta)_{1j}$ is the 3×3 submatrix of $\delta' C_{yy}\delta + \Delta$ consisting of rows and columns 1, 2, 4, and $(\delta' C_{yy})_{1j}$ is the 3×1 submatrix of $(\delta' C_{yy})$ consisting of rows 1, 2, 4 in column j of $\delta' C_{yy}$.

An Example

The primary advantage of the EM algorithm for factor analysis over methods such as that described in Jöreskog [1969], occurs with *a priori* zeros in the factor loadings because the EM algorithm does not require second derivatives to be calculated and so, in principle, requires substantially less storage. The C_{yy} in Table 1 is from Jöreskog [1969]. Following Jöreskog, we look for a four factor solution where, *a priori*, variables 1–4 have zero factor loadings on factor 4 and variables 5–9 have zero factor loading on factor 3; also, *a priori*, $R = I$.

Starting values for the residual variances τ^2 and regression coefficient matrix β are

TABLE 1

C_{yy} Matrix for Example

	Variables								
Variables	1	2	3	4	5	6	7	8	9
1	1.0	.554	.227	.189	.461	.506	.408	.280	.241
2		1.0	.296	.219	.479	.530	.425	.311	.311
3			1.0	.769	.237	.243	.304	.718	.730
4				1.0	.212	.226	.291	.681	.661
5					1.0	.520	.514	.313	.245
6						1.0	.473	.348	.290
7							1.0	.374	.306
8								1.0	.672
9									1.0

given in Table 2 for three different starting solutions, all with $R = I$. Also given in Table 2 are the results after 50 iterations of EM.

Table 3 gives the values of $f(\tau^2, \beta, R)$ and the rate of covergence for the slowest component of τ^2 for every fifth iteration.

The estimates for τ^2 found by EM agree with those found by Jöreskog's program from the same starting values. The relative performance of the two algorithms was not carefully monitored since it was not our objective to develop a competing packaged program. In fact, for this example, we did not optimize code at all, but simply used existing regression operators to perform the calculations on square symmetric matrices. Nevertheless, it may

Table 2

Results of Three EM Factor Analyses for Data of Table 1

	Variables	Estimates for Factor Loadings, β				Estimates for τ^2
		1	2	3	4	
SOLUTION 1	1	0.31(0.70)	0.26(0.60)	-0.59(0.30)	0.00(0.00)*	0.49(0.06)
	2	0.35(0.70)	0.30(0.60)	-0.61(0.30)	0.00(0.00)*	0.41(0.06)
	3	0.66(0.70)	0.57(0.60)	0.20(0.30)	0.00(0.00)*	0.19(0.06)
	4	0.62(0.70)	0.53(0.60)	0.25(0.30)	0.00(0.00)*	0.27(0.06)
	5	0.29(0.70)	0.25(0.60)	0.00(0.00)*	0.66(0.30)	0.42(0.06)
	6	0.31(0.70)	0.27(0.60)	0.00(0.00)*	0.55(0.30)	0.53(0.06)
	7	0.34(0.70)	0.29(0.60)	0.00(0.00)*	0.53(0.30)	0.53(0.06)
	8	0.62(0.70)	0.53(0.60)	0.00(0.00)*	0.01(0.30)	0.34(0.06)
	9	0.61(0.70)	0.52(0.60)	0.00(0.00)*	-0.09(0.30)	0.34(0.06)
SOLUTION 2	1	0.16(0.70)	0.65(0.60)	0.22(0.35)	0.00(0.00)*	0.50(0.17)
	2	0.21(0.65)	0.67(0.50)	0.39(0.50)	0.00(0.00)*	0.36(0.27)
	3	0.90(0.80)	0.08(0.50)	0.13(0.30)	0.00(0.00)*	0.16(0.16)
	4	0.84(0.70)	0.07(0.55)	-0.00(0.40)	0.00(0.00)*	0.29(0.21)
	5	0.19(0.60)	0.65(0.50)	0.00(0.00)*	0.15(0.55)	0.52(0.27)
	6	0.20(0.65)	0.73(0.50)	0.00(0.00)*	0.02(0.55)	0.42(0.16)
	7	0.29(0.65)	0.55(0.30)	0.00(0.00)*	0.69(0.70)	0.14(0.08)
	8	0.78(0.80)	0.24(0.40)	0.00(0.00)*	0.02(0.40)	0.33(0.24)
	9	0.79(0.80)	0.18(0.40)	0.00(0.00)*	-0.04(0.40)	0.35(0.25)
SOLUTION 3	1	0.70(0.70)	-0.12(-0.12)	0.15(0.15)	0.00(0.00)*	0.48(0.48)
	2	0.74(0.74)	-0.08(-0.08)	0.22(0.22)	0.00(0.00)*	0.40(0.41)
	3	0.39(0.39)	0.81(0.81)	0.33(0.33)	0.00(0.00)*	0.09(0.09)
	4	0.36(0.37)	0.75(0.75)	0.08(0.08)	0.00(0.00)*	0.30(0.31)
	5	0.65(0.65)	-0.03(-0.03)	0.00(0.00)*	0.37(0.37)	0.44(0.44)
	6	0.72(0.72)	-0.05(-0.05)	0.00(0.00)*	0.15(0.15)	0.46(0.46)
	7	0.60(0.60)	0.09(0.09)	0.00(0.00)*	0.35(0.35)	0.52(0.52)
	8	0.51(0.51)	0.65(0.65)	0.00(0.00)*	0.02(0.02)	0.32(0.32)
	9	0.47(0.48)	0.67(0.67)	0.00(0.00)*	-0.12(-0.12)	0.32(0.32)

Note: Starting values are given in parentheses and asterisked values are fixed at zero. Starting values are: ad hoc, near Jöreskog solution and based on principal components analysis.

TABLE 3

Convergence of EM for Three Solutions of Table 2:

	Solution 1		Solution 2		Solution 3	
Iteration	-f	τ^2 - ratio	-f	τ^2 - ratio	-f	τ^2 - ratio
5	0.84402	1.4749	0.21636	1.0806	0.00951	0.9989
10	0.49283	1.2112	0.08304	1.0392	0.00950	0.9996
15	0.45383	1.0423	0.03803	1.0267	0.00949	0.9998
20	0.44856	1.0146	0.02344	1.0193	0.00949	0.9998
25	0.44680	1.0085	0.01866	1.0136	0.00949	0.9999
30	0.44604	1.0062	0.01692	1.0095	0.00949	0.9999
35	0.44568	1.0048	0.01620	1.0078	0.00949	0.9999
40	0.44551	1.0038	0.01586	1.0062	0.00949	0.9999
45	0.44542	1.0030	0.01569	1.0050	0.00949	0.9999
50	0.44537	1.0024	0.01560	1.0041	0.00949	0.9999

Note: Column labeled -f is $-f(\tau^2, \beta, R)$. Column labeled τ^2 - ratio is the ratio of $\tau^2_{j\ell}$ to $\tau^2_{j,\ell-1}$ where ℓ refers to the iterations and j refers to the ratio farthest from unity. Thus, for solution 1, 1.0024 means that the most rapidly changing τ^2 from iteration 49 to 50 had a ratio of 1.0024.

be of interest to note that in this small example (9 variables, 4 factors), 50 iterations of EM took but 30% longer than LISREL to reach about the same solution, but used about 20% less storage. These figures would no doubt change for larger examples and for optimized EM code, and also would vary with the shape of the likelihood being maximized, LISREL being very efficient for normal likelihoods (i.e., quadratic log-likelihoods).

The fact that three different starting values lead to three essentially different (i.e., different τ^2) solutions is quite interesting, and complicates the interpretation of any solution. The first set of starting values is simply an ad hoc choice we make in order to test the coding for EM. Evidently, there is a broad but low maximum in the likelihood around the fixed values from this start because other ad hoc starting values converge to this solution. The second set of starting values is from Jöreskog [1969], and although the peak in the likelihood is much higher than for the first solution, it is apparently narrow in the sense that we must start near it to reach it. The third set of starting values is based on an initial principal components analysis, and also appears to be a relatively higher, narrow peak.

We have no reason to believe that these are the only maximums for this problem. In fact, one referee points out that a two factor solution (i.e., all β's in factors 3 and 4 set to zero) also appears to result in local maximum of the likelihood! Such behavior of the likelihood function should make any user of maximum likelihood factor analysis very uneasy. Minimally, the use of standard errors based on the second derivative matrix evaluated at a mode to measure precision of estimation should be viewed as being entirely questionable.

Discussion

A general issue that deserves commentary is that when the sample sizes are moderate, we do not expect maximum likelihood estimates, no matter how obtained, to be very good. The reason is that the m.l.e. is a joint modal estimate of many location (β) and scale (τ^2 and possibly R) parameters which does not necessarily yield a good estimate of τ^2 unless the likelihood is unimodal and approximately symmetric. From the behavior of algorithms discussed earlier, we see that the likelihood can be multimodal and presumably, asymmetric. We have no reason to suspect that this problem is rare.

Even with complete data in the linear model, maximum likelihood estimation is blind to the degree of freedom adjustments for variance estimates that correspond to adjusting for the dimensionality of the subspace of location parameters, i.e., the estimation of variances should be from their marginal likelihood having integrated over location parameters. The resultant systematic underestimation of the uniqueness by maximum likelihood will tend to yield Heywood cases (i.e., some $\tau_j^2 = 0$) too often. In standard normal problems this adjustment for variances is simple and does not affect the location estimates since location and scale parameters appear in separate factors of the likelihood (i.e., are *a posteriori* independent). However, in the factor analysis model, the β and τ^2 likelihoods do not factor, and consequently the systematic underestimation of τ^2 by maximum likelihood will affect both the estimates of τ^2 and of β. A better analysis, however, which estimates τ^2 after integrating out the location parameters, does not appear to be computationally straightforward. Nevertheless, it might eliminate the multiple modes in the likelihood, and so may be a worthwhile area for study.

Of course, the entire issue of the sensitivity of results to the assumption of multivariate normality is important for the wise application of the technique in practice. It is quite possible that new latent trait (or hidden variable) models, rather than factor analysis, would be more useful in practice. As pointed out in Aitkin, et.al. [1981], the EM algorithm is a generally powerful computational tool for estimation in such models, and so the detailed presentation of the EM algorithm here for factor analysis can be viewed as simply an illustration of its use when searching for structure in multivariate data with the hidden or latent variables treated as missing data.

REFERENCES

Aitkin, M., Anderson, D., & Hinde, J. Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Series A*, 1981, *144*.

Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, *39*, 1–38.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, *34*, 183–202.

Lawley, D. N. & Maxwell, A. E. *Factor analysis as a statistical method.* London: Butterworth, 1963.

Lawley, D. N. & Maxwell, A. E. *Factor analysis as a statistical method.* Second edition. London: Butterworth, 1971.

Manuscript received 3/18/81
Final version received 11/3/81