
TEACHER'S CORNER

A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data

Craig K. Enders
School of Education
University of Miami

Maximum likelihood algorithms for use with missing data are becoming commonplace in microcomputer packages. Specifically, 3 maximum likelihood algorithms are currently available in existing software packages: the multiple-group approach, full information maximum likelihood estimation, and the EM algorithm. Although they belong to the same family of estimator, confusion appears to exist over the differences among the 3 algorithms. This article provides a comprehensive, nontechnical overview of the 3 maximum likelihood algorithms. Multiple imputation, which is frequently used in conjunction with the EM algorithm, is also discussed.

Until recently, the analysis of data with missing observations has been dominated by listwise (LD) and pairwise (PD) deletion methods (Kim & Curry, 1977; Roth, 1994). However, alternative methods for treating missing data have become increasingly common in software packages, leaving applied researchers with a wide range of data analytic options. In particular, three maximum likelihood (ML) estimation algorithms for use with missing data are currently available: the multiple-group approach (Allison, 1987; Muthén, Kaplan, & Hollis, 1987) can be implemented using existing structural equation modeling (SEM) software; Amos (Arbuckle, 1995) and Mx (Neale, 1995) offer full information maximum likelihood

(FIML) estimation; and at least three packages, SPSS Missing Values, EMCOV (Graham & Hofer, 1993), and NORM (Schafer, 1998), incorporate the expectation maximization (EM) algorithm. The latter two programs also offer multiple imputation, as outlined by Rubin (1987).

The theoretical benefits of ML estimation are widely known (Little & Rubin, 1987), and simulation studies have suggested that ML algorithms may be superior to traditional ad hoc missing-data techniques in many cases (Arbuckle, 1996; Enders & Bandalos, in press; Muthén et al., 1987; Wothke, 2000). Although much of the recent missing-data research has been in the area of SEM, a great deal of confusion apparently exists over the differences among the three ML missing-data algorithms. For example, a search of the SEMNET discussion group archives revealed a large number of threads and requests for clarification during recent years, and the frequency of these threads does not appear to be diminishing. That confusion exists is probably not a surprise and is certainly not unwarranted; the ML algorithms appear fundamentally different in many respects, despite belonging to the same estimation family.

Although an extensive body of technical literature exists on ML missing-data methods (Dempster, Laird, & Rubin, 1977; Finkbeiner, 1979; Hartley & Hocking, 1971; Little & Rubin, 1987), no single reference is available to applied researchers that succinctly summarizes the similarities and differences among the algorithms. Thus, the goal of this article is to provide a thorough, nontechnical primer on three widely available ML estimation algorithms for use with missing data: multiple group analysis, FIML, and the EM algorithm. Multiple imputation algorithms, which are frequently used in conjunction with the EM algorithm, will also be discussed.

THEORETICAL BACKGROUND

To gain a full understanding of the ML algorithms, it is first necessary to review the mechanisms that cause missing data. According to Rubin (1976), data are missing completely at random (MCAR) when the missing values on a variable X are independent of other observed variables as well as the values of X itself. That is, the observed values of X are simply a random subsample of the hypothetically complete data. The missing-at-random (MAR) assumption provides a second, less restrictive, condition under which missing data can be ignored. Under the MAR condition, the probability that an observation is missing on variable X can depend on another observed variable in the data set, but not on the values of X itself. For example, suppose that two exams, X and Y , are administered to a group of examinees, and it is found that low scoring respondents on exam X have a tendency to drop out or refuse to take exam Y . In this case, the propensity to complete exam Y depends on scores from exam X and is unrelated to performance on Y . The MAR assumption is less re-

strictive in the sense that missing observations need not be a random subset of the hypothetically complete data.

The distinction between MCAR and MAR is important, because the most common missing-data methods (LD and PD) will only yield unbiased parameter estimates when MCAR holds, and this assumption is frequently not met in practice (Muthén et al., 1987). In contrast, ML methods should yield unbiased estimates under both the MCAR and MAR assumptions. Furthermore, even when data are MCAR, ML methods should provide more efficient parameter estimates than either LD or PD.

From the previous discussion, the reader may wonder why only ML methods yield unbiased parameter estimates under MAR whereas other methods do not. A simple example involving the estimation of means should clarify this point. The small data set in Table 1 builds on the previous MAR example involving two test administrations, X and Y .

The pair of test scores in the first two columns represents the hypothetically complete data set, whereas scores in the last two columns represent an MAR missing-data pattern; notice that data points are missing on Y for those low-scoring individuals on X . Table 2 shows ML estimates of mean and standard errors for the FIML algorithm and complete-data ML following LD; the values in bold are ML estimates taken from the complete data and can be viewed as the “correct” estimates.

An important point is illustrated in Table 2. As seen in the table, the FIML estimates are approximately correct, whereas the LD deletion estimates are positively biased. Although this may seem counterintuitive at first, there is a simple explanation. On the one hand, the two exam scores are dependent on one another ($r =$

TABLE 1
Example of a MAR Missing Data Pattern

<i>Complete Data</i>		<i>Missing Data</i>	
<i>Exam X</i>	<i>Exam Y</i>	<i>Exam X</i>	<i>Exam Y</i>
74	66	74	66
70	58	70	58
66	74	66	74
55	47	55	47
52	61	52	61
47	38	47	38
45	32	45	—
38	46	38	—
33	41	33	41
28	44	28	—

Note. MAR = missing at random.

TABLE 2
Mean Estimation Results (ML vs. LD)

	<i>Mean</i>		<i>SE</i>	
	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
Complete	50.80	50.70	4.95	4.24
LD	56.71	55.00	5.31	4.94
FIML	50.80	50.77	4.95	4.75

Note. Values in bold are ML estimates taken from the complete data and can be viewed as the “correct” estimates. ML = maximum likelihood; LD = listwise deletion.

.734), and under MAR the missing values on *Y* are also dependent on *X*. Under MAR, this implies that the linear relation between *X* and *Y* can be used to infer probable values for the missing-data points on *Y*. In simplistic terms, the ML algorithms “borrow” information from other variables during the estimation of parameters that involve missing values by incorporating information from the conditional distribution of the missing data given the observed data. In contrast, LD utilizes only information from the marginal distributions of *X* and *Y* during parameter estimation. As such, data from low-scoring respondents are discarded in this example, resulting in positive bias of the mean estimates.¹ Note that, under MCAR, the missing values are independent of other variables in the data set, so utilizing information from other observed variables adds nothing to the accuracy of the estimates; it does result in increased efficiency, however. This point is evidenced by the FIML standard error estimates found in Table 2, which are lower than the LD estimates.

Prior to discussing the individual algorithms, it is also important to note that, like complete-data ML methods employed in SEM, the three missing-data algorithms assume multivariate normality. There is virtually no research that has investigated the effects of nonnormality on ML missing-data algorithms, but it is reasonable to expect that these methods are no more robust to the assumption than complete-data ML algorithms. The three algorithms are discussed in detail in the following sections, followed by a discussion of multiple imputation algorithms.

MULTIPLE-GROUP APPROACH

An early method for obtaining ML parameter estimates in the presence of missing data was given by Hartley and Hocking (1971). The application of this method to SEM analyses was outlined by Allison (1987) and Muthén et al. (1987) and has

¹Note that LD and PD would yield the same value for the mean of *Y* in this example. The PD mean of *X* is identical to that generated from the complete data.

since been referred to as the multiple-group method. In this procedure, a sample is divided into G subgroups, such that each subgroup has the same pattern of missing data. That is, observations within each of the G subgroups have the same set of variables present and missing. A likelihood function is computed for each of the G groups, and the groupwise likelihood functions are accumulated across the entire sample and maximized. Although mathematically unrelated, this algorithm is loosely analogous to PD; a subgroup g_i contributes to the estimation of all parameters that involve the observed data points for that group but does not contribute to parameters that involve missing-data points.

Assuming multivariate normality, the log-likelihood function given by Hartley and Hocking (1971) is

$$-\frac{1}{2} \sum_{g=1}^G n_g \left[\log |\Sigma_g| + tr(S_g \Sigma_g^{-1}) + tr(H_g \Sigma_g^{-1}) + C_g \right] \quad (1)$$

where $H_g = (\bar{x}_g - \mu_g)(\bar{x}_g - \mu_g)'$. For each of the G subgroups, n_g is the number of observations, Σ_g and S_g are the parameter estimates and sample moments, respectively, C_g is a constant that depends on the data, and H_g contains the vector of mean residuals. Because the G subgroups have different patterns of missing data, this implies that the elements of \bar{x}_g , μ_g , S_g , and Σ_g are different for each group. To illustrate, consider a simple model comprising three observed variables: $X1$, $X2$, and $X3$. Furthermore, suppose a subgroup, g_1 , has complete data on $X1$ and $X3$, but is missing $X2$. The μ_g and Σ_g terms in the groupwise likelihood function for g_1 would contain only the parameter estimates that involve $X1$ and $X3$, as follows:

$$\mu = [\mu_1 \quad 0 \quad \mu_3] \text{ and } \Sigma = \begin{bmatrix} \sigma_{11} & 0 & \sigma_{13} \\ 0 & 0 & 0 \\ \sigma_{31} & 0 & \sigma_{33} \end{bmatrix}$$

Similarly, \bar{x}_g and S_g would contain the corresponding sample moments taken from the n_g complete observations in g_1 .

Allison (1987) and Muthén et al. (1987) demonstrated how to implement Hartley and Hocking's (1971) algorithm using the LISREL multiple-group specification, which maximizes the likelihood equation.

$$-\frac{1}{2} \sum_{g=1}^G n_g \left[\log |\Sigma_g| + tr(S_g \Sigma_g^{-1}) + C_g \right] \quad (2)$$

This function is clearly similar to Equation 1, but does not include a term for the vector of mean residuals—LISREL does allow for the addition of a mean vector term, however. In the usual SEM multiple-group analysis, G groups are formed that represent independently sampled subpopulations (e.g., men and women), and it is typically of interest to determine whether some specified set of parameters or parameter values are common to the G groups. In the missing-data application, the subpopulations correspond to the G patterns of missing data required by Hartley and Hocking's algorithm. The additional information from the groups with partially recorded data is incorporated by the specification of parameter equality constraints across the G groups.

Despite the wide availability of the LISREL program at the time, the multiple-group method of missing-data analysis had practical limitations that prevented its widespread use. As pointed out by Arbuckle (1996), the LISREL specification for the multiple-group approach required an exceptional level of expertise and thus was practically limited to situations in which there are only a small number of missing-data patterns. Muthén et al. (1987) and Kaplan (1995) described situations in which this might occur (e.g., BIB spiraled designs), but the number of distinct missing-data patterns is often quite large in applied settings, making the method difficult to implement.

Despite the technical difficulties associated with its implementation, the multiple-group approach does have advantages. First, the method can be used to estimate both just-identified (e.g., correlation, regression) and over-identified (e.g., SEM) model parameters. This is a point of contrast with the EM algorithm, which cannot currently be used to directly estimate linear model parameters. Second, it is important to note that the multiple-group approach does not estimate, or impute, missing observations, but yields direct estimates of model parameters and standard errors. This is an advantage, as additional corrective procedures are not necessary to obtain standard error estimates. Third, the multiple-group approach yields the usual chi-square test statistic for model fit, although the degrees of freedom and accompanying p value are incorrect due to the use of dummy values in the input covariance matrices of subsamples with missing variance/covariance elements. However, this is easily remedied by subtracting the number of pseudo-values from the degrees of freedom term. Finally, as a byproduct of the multiple-group specification, the chi-square statistic can also be used to test the MCAR assumption. If the MCAR assumption holds, parameter estimates across subgroups should be equal. Thus, the chi-square difference test of the equality constraints imposed across the G subgroups is also a test of the MCAR assumption; a statistically significant χ^2 value suggests that data are not MCAR.

FIML

Two structural equation modeling software packages currently offer FIML estimation routines for missing data: AMOS (Arbuckle, 1995) and Mx (Neale, 1995). The FIML approach was originally outlined by Finkbeiner (1979) for use with factor analysis and is similar to the multiple-group method, except that a likelihood function is calculated at the individual, rather than the group, level. For this reason, the FIML approach has been referred to as raw maximum likelihood estimation (Duncan, Duncan, & Li, 1998; Graham, Hofer, & MacKinnon, 1996).

Like the multiple-group approach, the FIML algorithm is conceptually analogous to PD (although mathematically unrelated) in the sense that all available data is used for parameter estimation. An examination of the individual-level likelihood function illustrates this point. Assuming multivariate normality, the casewise likelihood of the observed data is obtained by maximizing the function

$$\log L_i = K_i - \frac{1}{2} \log |\Sigma_i| - \frac{1}{2} (x_i - \mu_i)' \Sigma_i^{-1} (x_i - \mu_i) \quad (3)$$

where x_i is the vector of complete data for case i , μ_i contains the corresponding mean estimates derived from the entire sample, and K_i is a constant that depends on the number of complete-data points for case i . Like μ_i , the determinant and inverse of Σ_i are based only on those variables that are observed for case i . The overall discrepancy function value is obtained by summing the n casewise likelihood functions as follows:

$$\log L(\mu, \Sigma) = \sum_{i=1}^N \log L_i \quad (4)$$

To illustrate, suppose ML parameter estimates are sought for a model comprised of three observed variables: $X1$, $X2$, and $X3$. The parameters of interest are

$$\mu = [\mu_1 \ \mu_2 \ \mu_3] \text{ and } \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

The likelihood value for an observation with $X2$ missing would be a function of the two complete observations as well as the parameter estimates that involved $X1$ and $X3$. The relevant parameters are shown in the following.

$$\mu = [\mu_1 \quad 0 \quad \mu_3] \text{ and } \Sigma = \begin{bmatrix} \sigma_{11} & 0 & \sigma_{13} \\ 0 & 0 & 0 \\ \sigma_{31} & 0 & \sigma_{33} \end{bmatrix}$$

Based on the previous examples, the mathematical similarities between the multiple-group and FIML algorithms should be apparent; the primary difference is that FIML fitting function is the sum of n casewise likelihood values, whereas the multiple-group function is the sum of G groupwise likelihood values.

Several points should be made about the FIML algorithm. First, like the multiple-group approach, one of the advantages of the FIML algorithm is its applicability to both just-identified and over-identified models. In the latter case, the likelihood equation in Equation 3 is extended such that the first- and second-order moments (μ and Σ , respectively) are expressed as functions of some parameter vector, γ (Arbuckle, 1996). As such, the method is quite general and can be applied to a wide variety of analyses, including the estimation of means, covariance matrices, multiple regression, and SEM. Second, when used in SEM applications, FIML yields a chi-square test of model fit. However, the chi-square statistic generated by FIML does not take the usual form $F(N - 1)$, where F is the value of the fitting function. Clearly, the chi-square test cannot be calculated in the normal fashion, as there is no single value of N that is applicable to the entire sample. Also, unlike the usual SEM fitting functions, there is no minimum value associated with the FIML log-likelihood function, although the value of this statistic will increase as model fit worsens. Instead, a chi-square test for model fit is calculated as the difference in log-likelihood functions between the unrestricted (H_0) and restricted (H_1) models with degrees of freedom equal to the difference in the number of estimated parameters between the two models. Third, although many popular fit indexes can be computed under FIML, the specification of a means structure (required for estimation) renders certain fit indexes undefined (e.g., GFI). Fourth, similar to PD, indefinite covariance matrices are a potential byproduct of the FIML approach. However, Wothke (2000) suggested that indefiniteness problems are less pervasive with FIML than with PD. Fifth, unlike the EM algorithm (discussed in the following), standard error estimates are obtained directly from the analysis, and bootstrapping is not necessary. Finally, it is important to note that the FIML algorithm does not impute missing values; only model parameters are estimated.

EM ALGORITHM

At least three packages currently implement the EM algorithm: SPSS Missing Values, EMCOV (Graham & Hofer, 1993), and NORM (Schafer, 1998). An early work by Orchard and Woodbury (1972) explicated the underlying method, which they called the “missing information principle.” Dempster et al. (1977) provided an extensive generalization and illustration of the method and named it the EM algorithm.

The EM algorithm uses a two-step iterative procedure where missing observations are filled in, or imputed, and unknown parameters are subsequently esti-

mated. In the first step (the E step), missing values are replaced with the conditional expectation of the missing data given the observed data and an initial estimate of the covariance matrix. That is, missing values are replaced by the predicted scores from a series of regression equations where each missing variable is regressed on the remaining observed variables for a case i . Using the observed and imputed values, the sums and sums of squares and cross products are calculated.

To illustrate, suppose a mean vector and covariance matrix, $\theta = (\mu, \Sigma)$, is sought for an $n \times K$ data matrix, Y , that contains sets of observed and missing values (Y_{obs} and Y_{mis} , respectively). Using the observed values (Y_{obs}) and current parameter estimates ($\theta^{(t)}$), the calculations for the sufficient statistics at the t th iteration of the E step are

$$\begin{aligned} E\left(\sum_{i=1}^n y_{ij} | Y_{\text{obs}}, \theta^{(t)}\right) &= \sum_{i=1}^n y_{ij}^{(t)} \quad j = 1, \dots, K \\ E\left(\sum_{i=1}^n y_{ij} y_{ik} | Y_{\text{obs}}, \theta^{(t)}\right) &= \sum_{i=1}^n (y_{ij}^{(t)} y_{ik}^{(t)} + c_{jki}^{(t)}) \quad j, k = 1, \dots, K \end{aligned} \quad (5)$$

where

$$y_{ij}^{(t)} = \begin{cases} y_{ij}, & \text{if } y_{ij} \text{ is observed} \\ E(y_{ij} | y_{\text{obs}}, \theta^{(t)}), & \text{if } y_{ij} \text{ is missing} \end{cases}$$

and

$$c_{jki}^{(t)} = \begin{cases} 0 & \text{if } y_{ij} \text{ or } y_{ik} \text{ is observed} \\ \text{Cov}(y_{ij}, y_{ik} | y_{\text{obs}}, \theta^{(t)}), & \text{if } y_{ij} \text{ and } y_{ik} \text{ are missing} \end{cases}$$

Thus, missing values of y_{ij} are replaced with conditional means and covariances given the observed data and the current set of parameter estimates.² It should be noted that the preceding formulas can be found in Little and Rubin (1987).

In the second step (the M step), ML estimates of the mean vector and covariance matrix are obtained just as if there were no missing data using the sufficient statistics calculated at the previous E step. Thus, the M step is simply a complete-data ML estimation problem. The resulting covariance matrix and regression coefficients from the M step are then used to derive new estimates of the missing values

²As pointed out by Little and Rubin (1987), missing values are not necessarily replaced with actual data points, but are replaced by the condition functions of the missing values in the complete-data log-likelihood.

at the next E step, and the process begins again. The algorithm repeatedly cycles through these two steps until the difference between covariance matrices in subsequent M steps falls below some specified convergence criterion. Readers are encouraged to consult Little and Rubin (1987) for further technical details.

Several points should be noted concerning the EM algorithm. First, unlike the multiple-group and FIML approaches, the EM algorithm cannot be used to obtain direct estimates of linear model parameters (e.g., regression, SEM); as currently implemented, the EM algorithm can only be used to obtain ML estimates of a mean vector and covariance matrix. Obviously, this matrix can be used for input in subsequent linear model analyses. Additionally, the covariance matrix can be used to estimate, or impute, missing-data points at the final iteration. The latter approach may, at first glance, be appealing due to the illusion of a complete data set, but there is a notable drawback associated with this practice. Although the imputed values are optimal statistical estimates of the missing observations, they lack the residual variability present in the hypothetically complete data set; the imputed values fall directly on a regression line and are thus imputed without a random error component. As a result, standard errors from subsequent analyses will be negatively biased to some extent, and bootstrap (Efron, 1981) procedures must be employed to obtain correct estimates. Alternatively, multiple imputation procedures designed to recover residual variability are available in the EMCOV (Graham & Hofer, 1993) and NORM (Schafer, 1998) packages and are discussed next. However, it is important to note that a correction factor is added to the conditional expectation of the missing data at each E step to correct for this negative bias in the output covariance matrix; this is seen in the $c_{jkl}^{(t)}$ term in Equation 5. Although no studies have compared the impact of these two EM methods in the context of SEM, it seems reasonable to run analyses using the output covariance matrix rather than the singly imputed dataset.

Despite the difficulties previously noted, the EM algorithm may be preferred in situations where the missing-data mechanism (i.e., the variables are assumed to influence missingness) is not included in the linear model being tested. This is because the MAR assumption discussed previously is defined relative to the analyzed variables in a given data set. For example, if the missing values on a variable Y are dependent on the values of another variable X , the MAR assumption no longer holds if X is not included in the ultimate analysis. This is clearly problematic for the two direct estimation algorithms, as X must be incorporated in the substantive model for MAR to be tenable. However, this is not the case with the EM algorithm, as the input covariance matrix used to estimate substantive model parameters may be a subset of a larger covariance matrix produced from an EM analysis. In this case, the EM mean vector and covariance matrix are estimated using the full set of observed variables, and the elements that are of substantive interest are extracted for future analyses. Of course, the application of the EM algorithm in

this scenario assumes that the researcher has explicit knowledge of the missing-data mechanism, which may not likely be the case in practice. Nevertheless, the use of the EM algorithm in the manner described previously may make the MAR assumption more plausible in certain circumstances.

MULTIPLE IMPUTATION

The primary problem associated with EM algorithm is that the variability in the hypothetically complete data set is not fully captured during the imputation process. Multiple imputation, as outlined by Rubin (1987), creates $m > 1$ imputed data sets that are analyzed using standard complete-data methods. The m sets of parameter estimates are subsequently pooled into a single set of estimates using formulas provided by Rubin. The logic of multiple imputation is based on the notion that two sources of variability are lost during the EM imputation process. As described previously, the first occurs due to regression imputation; imputed values fall directly on the regression line and thus lack residual variability. The second source of lost variability is due to the fact that the regression equations are derived from a covariance matrix that is, itself, estimated with error due to the missing data. That is, the covariance matrix used to impute values is one of many plausible covariance matrices. The multiple imputation process attempts to restore the lost variability from both of these sources. Currently, there are at least two widely available multiple imputation programs based on the EM algorithm: EMCOV (Graham & Hofer, 1993) and NORM (Schafer, 1998).³ Although conceptually similar, the multiple imputation algorithms are quite different: EMCOV generates m imputed data sets using the bootstrap technique, whereas NORM does so using Bayesian simulation.

Following an initial EM analysis, EMCOV (Graham & Hofer, 1993) restores residual variability by adding a randomly sampled (with replacement) residual term to each of the imputed data points. For every nonmissing-data point in the original data set, a vector of residuals for each variable is calculated as the difference between the actual and predicted values from the regression equations (all other variables serving as predictors) used to impute missing values. Next, m data sets are created by repeatedly imputing missing values to the original data set such that $m - 1$ imputations are based on new estimates of the covariance matrix. In the first step, a bootstrap is performed on the original data, yielding a new data matrix of the same dimensions as the original. Next, the bootstrapped data are analyzed using the EM algorithm, and a new estimate of the covariance matrix is obtained. Finally, missing values in the original data set are imputed using regression equations generated from the new covariance matrix. This bootstrap process is repeated $m - 1$ times (the imputed data matrix from the original EM analysis serves as the

³EMCOV and NORM are available at <http://methcenter.psu.edu/software.html>

first of the m data sets), and residual variation is restored to the $m - 1$ sets of imputed data points using randomly sampled residual terms, as described previously.

In contrast, NORM (Schafer, 1998) uses iterative Bayesian simulation to generate m imputed data sets. Like the EM algorithm, the NORM algorithm repeatedly cycles through two steps: Missing observations are imputed (the imputation, or I step) and unknown parameters are estimated (the posterior, or P step). However, unlike EM, the data augmentation (DA) algorithm implemented in NORM uses a stochastic rather than a deterministic process. In the first step, missing-data points are replaced by randomly drawn values from the conditional distribution of the missing data given the observed data and a current estimate of the parameter vector θ ; parameter estimates from an EM analysis provide start values for the first iteration. Next, new parameter estimates are randomly drawn from a Bayesian posterior distribution conditioned on the observed and imputed values from the first step. These new parameter values are used to impute values in the subsequent I step, and the process begins again. This two-step procedure is iterated until convergence occurs, at which point the first of m imputed data matrices is created from a final I step. Additional imputed data sets are obtained by repeating the DA process $m - 1$ times. Finally, it should be noted that the stochastic nature of the DA process requires a different convergence criterion than the EM algorithm. Because DA parameter estimates are drawn randomly from a posterior probability distribution, values will naturally vary between successive iterations, even after convergence occurs. Thus, the DA algorithm converges when the *distribution* of the parameter estimates no longer changes between contiguous iterations. Readers are encouraged to consult Schafer (1997) and Schafer and Olsen (1998) for further details.

After implementing EMCOV or NORM, complete-data analyses are performed on each of the m imputed data sets, and the parameter estimates from these analyses are stored in a new file. Using rules provided by Rubin (1987), a single set of point estimates and standard error values can be obtained; both EMCOV and NORM include routines that will perform the necessary calculations. Two final points should be made regarding multiple imputation. First, Schafer (1997) suggested that adequate results could be obtained using as few as five imputed data sets. Second, a straightforward method of obtaining SEM goodness-of-fit tests is not currently available, although work on the topic is ongoing (Schafer & Olsen, 1998).

SUMMARY

Recent software advances have provided applied researchers with powerful options for analyzing data with missing observations. Specifically, three ML algorithms (multiple-group analysis, FIML, and the EM algorithm) are widely avail-

able in existing software packages. However, the wide array of data analytic options has resulted in some confusion over the differences among the three algorithms. As such, the goal of this article was to provide a brief overview of ML algorithms in hopes that applied researchers can make informed decisions regarding the use of ML algorithms in various data analytic settings.

To briefly summarize, the ML algorithms are built on the assumption that the observed data contain information that can be used to infer probable values for the missing data. In simplistic terms, information is borrowed from the complete data during the estimation of parameters that involve missing values. Like complete-data ML methods, the missing-data algorithms assume multivariate normality. To contrast the algorithms, it may be useful to dichotomize the methods into direct and indirect approaches. The multiple-group and FIML algorithms are direct approaches in the sense that linear model parameter estimates are obtained directly from the available raw data without a preliminary data preparation step (e.g., imputation). These algorithms are applicable to a wide variety of analyses (e.g., means estimation, regression, SEM) and are computationally less expensive than the EM algorithm because no additional steps are required to obtain correct standard error estimates. Additionally, tests of model fit are readily available in SEM settings.

In contrast, the EM algorithm is an indirect approach in the sense that an additional data preparation phase is necessary to obtain linear model parameter estimates. Currently, the EM algorithm can only be used to obtain ML estimates of a mean vector and covariance matrix. The covariance matrix can subsequently be used to estimate linear model parameters, or missing values can be imputed using the covariance matrix from the final iteration. However, additional analyses (e.g., bootstrapping, multiple imputation) are necessary to recover the residual variability lost during the imputation process. Despite the additional data analytic steps, the EM algorithm may be preferable when the missing-data mechanism does not appear in the substantive model.

REFERENCES

- Allison, P. D. (1987). Estimation of linear models with incomplete data. In C. C. Clogg (Ed.), *Sociological methodology, 1987* (pp. 71–103). San Francisco: Jossey-Bass.
- Arbuckle, J. L. (1995). Amos user's guide [Computer software]. Chicago: Smallwaters.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B*, 39, 1–38.

- Duncan, T. E., Duncan, S. C., & Li, F. (1998). A comparison of model- and multiple imputation-based approaches to longitudinal analyses with partial missingness. *Structural Equation Modeling*, *5*, 1–21.
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap, and other resampling methods. *Biometrika*, *68*, 589–599.
- Enders, C. K., & Bandalos, D. L. (in press). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*.
- Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika*, *44*, 409–420.
- Graham, J. W., & Hofer, S. M. (1993). EMCOV reference manual [Computer software]. Los Angeles: University of Southern California, Institute for Prevention Research.
- Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*, 197–218.
- Hartley, H. O., & Hocking, R. R. (1971). The analysis of incomplete data. *Biometrics*, *27*, 783–823.
- Kaplan, D. (1995). The impact of BIB spiraling-induced missing data patterns on goodness-of-fit tests in factor analysis. *Journal of Educational and Behavioral Statistics*, *20*, 69–82.
- Kim, J., & Curry, J. (1977). The treatment of missing data in multivariate analyses. *Sociological Methods & Research*, *6*, 215–240.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431–462.
- Neale, M. C. (1995). Mx: Statistical Modeling (3rd ed.) [Computer software]. Richmond, VA: Department of Psychiatry, Medical College of Virginia, Virginia Commonwealth University.
- Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and applications. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 697–715.
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, *47*, 537–560.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall.
- Schafer, J. L. (1998). NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2 [Computer software]. Software for Windows 95/98/NT. Available: <http://www.stat.psu.edu/~jls/misoftwa.html>
- Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, *33*, 545–571.
- Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples* (pp. 219–240). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Copyright of Structural Equation Modeling is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.