6-2008

A Comparison of Several Algorithms and Models for Analyzing Multivariate Normal Data with Missing Responses

Mojtaba Ganjali
Shaheed Beheshti University

H. Ranji
Statistical Research and Training Center

Recommended Citation

A Comparison of Several Algorithms and Models for Analyzing Multivariate Normal Data with Missing Responses

Mojtaba Ganjali[*]
Department of Statistics, Faculty of Mathematical Sciences
Shaheed Beheshti University, Evin, Tehran, Iran
m-ganjali@sbu.ac.ir

H. Ranji
Statistical Research and Training Center
Dr. Fatemi Avenue, Tehran 1413717911, Iran
E-mail: h.ranji@srtc.ac.ir

Abstract

In this paper we compare some modern algorithms i.e. Direct Maximization of the Likelihood (DML), the EM algorithm, and Multiple Imputation (MI) for analyzing multivariate normal data with missing responses. We also compare two approaches for modeling incomplete data (1) ignoring missing data and (2) joint modeling of response and non-response mechanisms. Several types of Software which can be used to implement the above algorithms are also mentioned. We used these algorithms for a simulation study and to analyze a data set where outliers affect the parameter estimates and final conclusion. As the variance of the estimates cannot be obtained using the available software for some of the algorithms, a bootstrap method is used to find them.

1. Introduction

There are several algorithms to the analysis of a multivariate normal data matrix with an arbitrary pattern of missing values. Three of these modern algorithms are: (1) direct maximization of the likelihood, (2) The EM algorithm which was described by Dempster, et al. (1977) and (3) The general theory of Multiple Imputation (MI), originally proposed by Rubin (1977, 1978). This method uses Data augmentation as part of the Markov Chain Monte Carlo (MCMC) method and shares the same underlying philosophy as the EM; solving an incomplete data problem by repeatedly solving the complete data version. There are also two

---

[*] Correspondence author, Associate Professor of Statistics

approaches to model the data (1) ignoring the missing data mechanism and model only the available data or (2) using the joint modeling of response and non-response. The later has been extensively used recently (Little and Rubin, 2002, Diggle and Kenward, 1994, Crouchley and Ganjali, 2002 and Ganjali and Jolani, 2004) for analyzing cross sectional and longitudinal data.

Rubin (1976) and Little and Rubin (2002) define various missing response mechanisms. Under their definitions, based on the likelihood function, one such mechanism, known as 'missing completely at random' (MCAR) arises when the probability of missing an observation is independent of responses and non-responses. Under 'missing at random' (MAR) this probability, given observed responses, does not depend on the missing responses. In 'missing not at random' (MNAR), which is called by Diggle and Kenward (1994) 'informative dropout' for a monotone pattern of missing responses, this probability depends on missing responses and so, in this case, the mechanism accounting for items being missing (missing mechanism) cannot be ignored.

In this paper we compare the use of the direct maximization, EM and MI for analyzing incomplete data. We also compare two approaches of modeling one based on the MAR assumption (ignoring the missing data mechanism) which does not need any model for non-response mechanism and the joint modeling of responses and non-response mechanism (models such as the model of Diggle and Kenward, 1994, or generalized Heckman model presented by Crouchley and Ganjali, 2002) for analyzing multivariate normal data with missing responses. The model that we will use to model responses is a general multivariate Gaussian model for a vector-valued response, with the data assumed to be an independent random sample from this distribution. This is a sensible choice, as it avoids the need to discuss what kinds of structure might be appropriate in particular settings, for example in longitudinal data analysis. The focus of comparison will be placed firmly on different ways to deal with the missing values, ignoring them or joint modeling of responses with a model for missing data mechanism. In the later case sensitivity of the results to the assumptions of the model for missing mechanism will be also discussed.

In Section 2, direct maximization, the EM and MI algorithms are briefly reviewed. In section 3, joint modeling of response and missing data mechanism is mentioned. In Section 4, we have some comparisons of the three algorithms and some comparisons between the two approaches of modeling missing data, in the course of which the advantages and disadvantages of algorithms and models will be discussed. In Section 5, we use these algorithms and models in an applied example where outliers play an important role. In this Section to find the variance of estimates by the EM or MI a bootstrap method is used. In Section 6 we give our conclusions.

2. Modern Algorithms to the Analysis of Incomplete Normal Data

Assume the data set to be a matrix of n rows and p columns, with rows corresponding to individual observations and columns corresponding to variables measured on each individual. Denote $Y = (Y_{obs}, Y_{mis})$ as pseudo-complete[1] data matrix where Y_{obs} and Y_{mis} represent the observed and missing portions of the data matrix, respectively. Let y_{ij} denote an individual

---

[1] We use the term "pseudo-complete" for the result of putting together the known data item Y_{obs} with a term Y_{mis} representing the unknown value of the missing data item, so nominally "completing" a data set.

element of Y, $i = 1,...,n$, $j = 1,...,p$. The ith row of Y, is $y_i = (y_{i1}, y_{i2},..., y_{ip})'$. We assume $y_1, y_2,..., y_n$ to have independent multivariate normal distributions with mean vector μ and positive definite covariance matrix Σ. Let $\theta = (\mu, \Sigma)$ be the vector of unknown parameters. In the following we use the notation due to Schafer (1997).

The rows of Y can be grouped according to their missingness patterns. We shall index the missingness pattern by $s = 1, 2,..., S$, where S is the number of patterns in the data matrix with missing values. Let R be an $S \times p$ matrix of binary indicators with element, r_{sj}, where $r_{sj} = 1$ if Y_j is observed in pattern s and $r_{sj} = 0$ if Y_j is missing in pattern s. For each missingness pattern s, let $O(s)$ and $M(s)$ denote the subsets of the column labels $\{1,2,...,p\}$ corresponding to variables that are observed and missing, respectively, $O(s) = \{j : r_{sj} = 1\}$, and $M(s) = \{j : r_{sj} = 0\}$. Finally, let $I(s)$ denote the subset of $\{1,2,...,n\}$ corresponding to the rows of Y that exhibit pattern s. In the following we shall review how the three algorithms can be used to analyze our data with the assumption of MAR.

1. Direct Maximization of the likelihood

Under the ignitability assumption, logarithm of the observed-data likelihood, ignoring some constant values, is

$$\log L(\theta; Y_{obs}) = \sum_{s=1}^{S} \sum_{i \in I(s)} \left[-\frac{1}{2} \log |\Sigma_s^*| - \frac{1}{2} (y_i^* - \mu_s^*)^T \Sigma_s^{*-1} (y_i^* - \mu_s^*) \right], \tag{1}$$

where y_i^* denotes the observed part of the row i of the data matrix and μ_s^* and Σ_s^* denote the subvector of the mean vector μ and the square submatrix of the covariance matrix Σ in pattern s, respectively.

An estimate $\hat{\theta}$ of θ can be obtained as a solution of the observed-data likelihood equation: $\partial \log L(\theta; Y_{obs})/\partial \theta = 0$. The root of this equation which globally maximizes the observed-data likelihood (maximum likelihood estimate, MLE) would be a consistent and efficient estimate of θ under regular conditions. One estimate of the covariance matrix of $\hat{\theta}$ in large samples is $\left[I(\hat{\theta} | Y_{obs}) \right]^{-1}$, where $I(\theta | Y_{obs})$ is the observed Fisher information matrix defined as:

$$I(\theta; Y_{obs}) = -\partial^2 \log L(\theta; Y_{obs})/\partial \theta \partial \theta^T. \tag{2}$$

It is possible to compute iteratively the MLE of θ by using a Newton-Raphson maximization procedure or some other variant of it i.e. Fisher's method of scoring or Quasi-Newton method (see, Everitt, 1987, Thisted, 1988 and McLachlan and Krishnan, 1997).

2. The EM algorithm

When portions of the data matrix Y are missing, we may use iterative computations like EM to find the ML estimates. EM has two steps which are discussed in the following paragraphs.

1. The E-step

In the E-step of the EM algorithm for multivariate normal data, one calculates the expectation of the complete-data sufficient statistics over $P(Y_{mis} | Y_{obs}, \theta)$ for an assumed value of θ. These statistics are of the form $\sum_i y_{ij}$ and $\sum_i y_{ij} y_{ik}$, so to perform the E-step we need to find the expectations of y_{ij} and $y_{ij} y_{ik}$ over $P(Y_{mis} | Y_{obs}, \theta)$. These expectations are obtained by sweeping the θ-matrix on the positions corresponding to the variables in $y_{i(obs)}$, see Schafer (1997, page 164). To see this, let A denote the swept parameter matrix $A = SWP[O(s)]\theta$, and let a_{jk} denote the $(j,k)^{\text{th}}$ element of A, $j, k=0,1,...,p$. The first two moments of $y_{i(mis)}$ with respect to $P(Y_{mis} | Y_{obs}, \theta)$ are given by

$$E(y_{ij} | Y_{obs}, \theta) = a_{0j} + \sum_{k \in O(s)} a_{kj} y_{ik}, \text{ and } Cov(y_{ij}, y_{ik} | Y_{obs}, \theta) = a_{jk},$$

for each $i \in I(s)$ and $j,k \in M(s)$. For any $j \in O(s)$, the moments are

$$E(y_{ij} | Y_{obs}, \theta) = y_{ij}, \text{ and } Cov(y_{ij}, y_{ik} | Y_{obs}, \theta) = 0.$$

Applying the relationship $E(y_{ij} y_{ik} | Y_{obs}, \theta) = Cov(y_{ij}, y_{ik} | Y_{obs}, \theta) + E(y_{ij} | Y_{obs}, \theta) E(y_{ik} | Y_{obs}, \theta)$, it follows that

$$E(y_{ij} | Y_{obs}, \theta) = \begin{cases} y_{ij} & \text{for } j \in O(s), \\ y_{ij}^* & \text{for } j \in M(s), \end{cases} \text{ and}$$

$$E(y_{ij} y_{ik} | Y_{obs}, \theta) = \begin{cases} y_{ij} y_{ik} & \text{for } j,k \in O(s), \\ y_{ij}^* y_{ik} & \text{for } j \in M(s), k \in O(s), \\ a_{jk} + y_{ij}^* y_{ik}^* & \text{for } j,k \in M(s), \end{cases}$$

where

$$y_{ij}^* = a_{0j} + \sum_{k \in O(s)} a_{kj} y_{ik}.$$

The E-step consists of calculating and summing these expected values of y_{ij} and $y_{ij} y_{ik}$ over i for each j and k. The output of an E-step can then be written as $E(T | Y_{obs}, \theta)$, where T is the matrix of pseudo-complete-data sufficient statistics $T = [1, Y]'[1, Y] = \begin{bmatrix} n & 1'Y \\ Y'1 & Y'Y \end{bmatrix}$, where 1 shows a column vector with all elements equal to 1.

2. The M-step

Once $E(T | Y_{obs}, \theta)$ has been found, the M-step can be carried out. For a given value of T the pseudo-complete-data MLE is $\hat{\theta} = SWP[0] n^{-1} T$, and the M-step carries out this same operation on $E(T | Y_{obs}, \theta)$ rather than on T. A single iteration of EM can, thus, be written as

$$\theta^{(t+1)} = SWP[0] n^{-1} E(T | Y_{obs}, \theta^{(t)}). \tag{3}$$

One can iterate the EM to reach the convergence criteria of interest. By EM the last values of $E(y_{ij} \mid y_{obs}, \theta^{(t)})$ can be imputed for missing values to have a filled in data set.

The EM algorithm does not provide the standard errors of the parameter estimates. Several methods have been proposed in literature to solve this problem, see, for example, Louis (1982), Meilijson (1989) and Meng and Rubin (1991). Louis' formula (Louis, 1982) relates the observed information matrix to the conditional expectation of the second derivatives of complete data log-likelihood function and the covariance of the first derivatives of complete data log-likelihood function. Evaluating the integrals in this formula, in the current setting, may not be easy. We shall use a bootstrap approach, in our application, to find the standard errors of EM estimates.

3. MI

Multiple imputation is a technique that replaces each missing value with two or more accepted values (m values) extracted by, for example, data augmentation (DA). Then each of the imputed data sets is analyzed using standard complete-data procedures and, in the end, results are combined by a flexible method which is given, in details, by Rubin (1987). We shall use Rubin's rules (Little and Rubin, 2002, page 86) in our simulation study in section 4 to find standard errors of the parameter estimates.

DA for incomplete multivariate normal data has two steps: I-step and P-step which are discussed in the following paragraphs.

1. The I-step

For a given vector of parameters at iteration $t\left(\theta^{(t)}\right)$ the I-step simulates $Y_{mis}^{(t+1)} \sim P(Y_{mis} \mid Y_{obs}, \theta^{(t)})$, and the P-step simulates $\theta^{(t+1)} \sim P(\theta \mid Y_{obs}, Y_{mis}^{(t+1)})$. The I-step of data augmentation involves the independent simulation of random normal vectors for each row of the data matrix, with means and covariances given by Schafer (1997, page 181).

2. The P-step

Under the conjugate prior distributions, for μ given Σ, as multivariate normal $[N(\mu_0, \tau^{-1}\Sigma)]$ and inverted Wishart $[W^{-1}(m, \Lambda)]$ for Σ, the complete data posterior $P(\theta \mid Y_{obs}, Y_{mis})$ is a normal inverted Wishart distribution. The P-step of data augmentation, therefore, is merely a simulation of the normal inverted-Wishart distribution, which is: $\mu \mid \Sigma, Y \sim N(\mu'_0, (\tau')^{-1}\Sigma)$, and $\Sigma \mid Y \sim W^{-1}(m', \Lambda')$, for the updated hyperparameter $(\tau', m', \mu'_0, \Lambda')$ (see Schafer, 1997). The missing data $Y_{mis}^{(t)}$ is imputed at last I-step. One can also use Jeffreys' prior, as a non-informative prior, to handle DA.

3. The Joint Modeling of Response and Missing Mechanism

In this section we shall describe two models of many for joint modeling of response and missing mechanism. The first one is called the multivariate generalized Heckman model. The second model is the model used by Diggle and Kenward (1994). The first model includes as a special case the second and some other models which are used for the joint modeling of response and missing mechanism.

1. The Multivariate generalized Heckman selection model

The joint model due to Heckman (1979), for a continuous response y_i and a sample selection mechanism, is generalized by Crouchley and Ganjali (2002) for the situation of repeated responses with dropout. This model is

$$R_{it}^* = \alpha_t W_{it} + v_{it}, \ y_{it}^* = \beta_t X_{it} + \varepsilon_{it}, \tag{4}$$

where t=1,2,...,T . The vectors of covariates W_{it} and X_{it} can include covariates at time s, $s < t$. The vectors of parameters α_t and β_t need to be estimated. The correlated errors for ith individual at time t are v_{it} and ε_{it} . In this model $\mathbf{y}_i = (y_{i1}, y_{i2},..., y_{iT})$, $\mathbf{R}_i = (R_{i2},..., R_{iT})$, where $y_{it} = y_{it}^*$, if $R_{it}^* > 0$, and $y_{it} = 0$, if $R_{it}^* \leq 0$, for i=1,2,...,n, where $y_{it} = 0$ is used to indicate a missing response at time t. An explicit missing data indicator is also defined as $R_{it} = 1$, if $R_{it}^* > 0$, and $R_{it} = 0$, if $R_{it}^* \leq 0$.

It is assumed that all the subjects at the start of the study are observed, i.e. $R_{i1} = 1, \forall i$. This model can be used for monotone and non-monotone missing responses. For the case of monotone pattern (our example in section 5), the observations for the subject i take the form $(\mathbf{y}_i, \mathbf{R}_i) = ([y_{i1}^*,..., y_{ij}^*, 0,...,0], [1,...,1,0,...,0])$, if dropout occurs at time $j+1$ and $(\mathbf{y}_i, \mathbf{R}_i) = ([y_{i1}^*,..., y_{iT}^*], [1,...,1])$, if a subject is completely observed. In equation (4) $var(\varepsilon_i) = \Sigma_{YY}$ and Σ_{YY} can be unstructured so that $var(\varepsilon_{it}) = \sigma_{YY_{t,t}}^2$ and $cov(\varepsilon_{is}, \varepsilon_{it}) = \sigma_{YY_{s,t}}$. It is also assumed that the subjects are independent of each other, so that $cov(\varepsilon_{is}, \varepsilon_{i't}) = 0$ for $i \neq i'$ for all s and t. In equation (4) $var(v_i) = \Sigma_{RR}$, where the diagonal elements of Σ_{RR} are 1, and the off-diagonal elements of $\Sigma_{RR} = [cov(v_{is}, v_{it})] = [\sigma_{RR_{s,t}}]$. Let also $\Sigma_{YR} = [cov(\varepsilon_{is}, v_{it})] = [\sigma_{YR_{s,t}}]$. Both Σ_{YR} and the off-diagonal elements of Σ_{RR} are unstructured. For a sequence without dropout, the variance-covariance matrix of the joint responses $(\mathbf{y}_i, \mathbf{R}_i)$ takes the form $\Sigma = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YR} \\ \Sigma_{RY} & \Sigma_{RR} \end{bmatrix}$.

For conditions we need for MCAR and MAR in this model see Ganjali and Rezaei (2005, page 290). For a check of goodness of fit, Crouchley and Ganjali (2002), using model fitting of equation (4), introduced a modified Pearson residual for detecting outliers where only the goodness of fit of the observed data, given being observed, are examined. For the goodness of fit of the model one has to assume that missing data follow the same distribution as observed data. Ganjali and Jolani (2004) propose a two-stage method to find the parameters estimates of the generalized Heckman model.

2. Form of the Diggle and Kenward model

To account for dependence between the response and non-response processes in a monotone pattern, Diggle and Kenward (1994) use the current and previous value of the response process in their model for the non-response process. Assuming the first response is observed for all individuals, the form of their model is:

$$R_{it}^* = \alpha_t W_{it} + \gamma_1 y_{it-1} + \gamma_2 y_{it} + v_{it}, \quad t = 2,...,T; \ y_{it}^* = \beta_t X_{it} + \varepsilon_{it}, \quad t = 1,...,T , \tag{5}$$

where vectors of covariates are W_{it} and X_{it}. The vectors of parameters α_t, β_t and the scale parameters γ_1 and γ_2 need to be estimated. The correlated errors for i^{th} individual at time t are ν_{it} and ε_{it}. In this model Diggle and Kenward (1994) assume $\text{cov}(\nu_{is}, \nu_{it}) = 0$, and $\text{cov}(\nu_{it}, \nu_{i't}) = 0$. They use a logistic distribution for ν_{it}. However, One may assume that ν_{it} is *i.i.d* normal with mean zero and variance 1, then use a probit model. In Diggle and Kenward (1994) model we have MCAR if $\gamma_1 = \gamma_2 = 0$. We have MAR if $\gamma_2 = 0$, and when $\gamma_2 \neq 0$ we have MNAR.

At first sight the Diggle and Kenward (1994) model looks very different from the generalized Heckman model. However, Crouchley and Ganjali (2002) found a reduced form of the Diggle and Kenward (1994) model which has the same form as a generalized Heckman model. For a study limited to $T = 2$, both the generalized Heckman model and the Diggle and Kenward (1994) model will produce the same inference for β and the nature of the dropout. But, for longer series, the Diggle and Kenward (1994) model is more restrictive than the generalized Heckman model. Besides, the Diggle and Kenward (1994) model can only be used for monotone pattern of missing responses, but generalized Heckman model can be used for monotone or non-monotone missing responses. Some other models that have a special structure of the generalized Heckman model are the random coefficient model proposed by Follman and Wu (1995) and Wu and Carroll (1988), models proposed by Ridder (1990) and Hausman and Wise (1979) model (see Crouchley and Ganjali, 2002).

4. Some comparisons of three algorithms and two ways of modeling

Generally in direct maximization of the likelihood Newton-Raphson algorithm converges faster than quasi-Newton, but it is less stable and more sensitive to the choice of initial values, whereas Quasi Newton is more stable but converges much more slowly (Tang et. al., 2003). Use of complete case analysis estimates may help Newton-Raphson or quasi-Newton to converge faster.

The EM algorithm is a general-purpose algorithm which, with its simplicity and stability, can be implemented, for multivariate normal model, on BMDP, Norm, Mix, Cat, SAS and, recently, SPSS statistical software. All of these software packages provide the EM parameter estimates, but, due to a lack of ready formula, they do not give the variance of the estimates. In the following section we shall use, in our application, a bootstrap approach to find the variance of the estimates. EM imputes only one value for each missing value and can not reflect uncertainty about the true values of the missing data. Multiple imputations replace each missing value by $m > 1$ simulated values and can reflect the uncertainty mentioned above. MI uses MCMC, an approach that, like EM, provides point estimates of parameters. Furthermore, in contrast with EM, MCMC approach can provide measures of uncertainty associated with parameter estimates. MCMC methods are stochastic and easier to implement conceptually and computationally when the number of unknown parameters is large. They can also converge to the probability distribution we need, but EM is slow and may not converge to a unique global maximum. However, MCMC needs high performance computers and, in using it, monitoring the convergence could be difficult. Furthermore, in this method, the use of a prior distribution can be regarded as subjective.

Now, let us point out some comments in using joint modeling. Joint modeling is computationally intensive and you may need to write your own program, for the data you have, to use it (unless you use the two-stage method proposed by Ganjali and Jolani (2004),

which gives consistent estimates, but estimates that are not as efficient as those given by ML approach). In contrast with EM or MI, in using joint modeling you do not need to make the MAR assumption. In fact, joint modeling is the only way you can test for MAR. However, in an applied example with two responses Molenberghs et al. (2001), using a sensitivity approach, and Crouchley and Ganjali (2002), using the generalized Heckman selection model, found some influential outlier points as the reason for MNAR. They found a MNAR mechanism when they used data including the outliers and a MCAR mechanism when the outliers were removed. Kenword (1998), in the same example, using a joint model, without removing the outliers, used a 't' distribution for the second response given the first response and found a MAR mechanism for missing responses. So, joint modeling can test MAR for the data on hand, but for this it has to adopt strong assumptions that are not testable, and in this method results may be sensitive to the modeling assumptions. As mentioned, EM and MI are based on the MAR assumption and if this assumption does not apply to the data on hand, results obtained by EM or MI can be misleading. Therefore, we need to test for MAR in one place where we have to make strong assumptions (i.e. in joint modeling). In other places, if we ignore the missing mechanism and use MI or EM, when the data are in fact missing not at random, we reach incorrect estimates.

Following simulated example illustrates some of our above discussions. Consider a bivariate data in which Y_1 is always observed but Y_2 is sometimes missing. We would like to compare the performance of the complete case analysis (CC), available case analysis using direct maximization of the likelihood (DML), EM, MI and joint modeling by simulation. Let us define $R = (r_1,...,r_n)$ to be the vector of response indicators where $r_i = 1$ if Y_2 is observed and $r_i = 0$ if Y_2 is missing for unit $i, i = 1,...,n$. We consider the following ignorable and nonignorable mechanisms:

(1) $p(r_i = 1 | y_{i1}, y_{i2}) = a_1, \ 0 \le a_1 \le 1,$
(2) $p(r_i = 1 | y_{i1}, y_{i2}) = \Phi(a_2 + b_2 y_{i1}),$
(3) $p(r_i = 1 | y_{i1}, y_{i2}) = \Phi(a_3 + b_3 y_{i2}),$ \hfill (6)

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. In above mechanisms, (1) is MCAR, (2) is MAR and (3) is MNAR. A simulation is conducted in which samples of size n=100 and n=1000 are drown from bivariate normal distribution with $\mu_1 = 0, \mu_2 = 0, \sigma_1^2 = 1, \sigma_2^2 = 1$ and $\sigma_{12} = \rho = 0$ and 0.5, respectively. Random samples are drawn with $a_1 = 0.75$, $a_2 = a_3 = 0.70$, $b_2 = b_3 = 1$. These constants are chosen to yield an expected response rate of 25% under each mechanism, a level close to our applied example in the next subsection.

We find the estimate of parameters in each repetition by writing our program in R (using function 'optim') for CC, DML, EM, JM and we used SAS, version 9.1, for MI (where Jeffreys' prior is used, m=5 is used and estimate of standard errors are found by Rubin's rules, 1987). The means of estimates for CC, DML, EM, MI and joint modeling over 1000 repetitions are shown in Tables 1 and 2 for n=100 and n=1000, respectively. The standard errors of estimates of CC, DML and joint modeling are means of standard errors (obtained by using Fisher information matrix) over 1000 repetitions. The standard error of estimates of EM and MI approaches are the sampling standard error.

Table 1.   Estimated parameter by using the complete cases (CC), available cases using direct maximization of the likelihood (DML), EM, MI and joint modeling (JM) for simulated data (n=100) with MCAR, MAR and MNAR mechanisms (*: mean of the standard errors over 1000 iteration obtained using information matrix, **: sampling standard error)

Mec.	Par.	True	CC Es.	CC Se.*	DML Es.	DML Se.*	EM Es.	EM Se.**	MI Es.	MI Se.**	JM Es.	JM Se.*
True value of $\rho=0$										
MCAR	μ_1	0.00	0.00	0.11	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10
	μ_2	0.00	0.00	0.11	0.01	0.10	-0.01	0.11	0.00	0.11	0.00	0.11
	σ_1	1.00	0.99	0.08	0.99	0.07	0.99	0.07	1.00	0.07	0.99	0.07
	σ_2	1.00	0.99	0.08	0.98	0.07	0.99	0.08	1.00	0.08	0.99	0.08
	ρ	0.00	0.00	0.11	0.00	0.10	0.00	0.12	0.00	0.11	0.00	0.11
MAR	μ_1	0.00	0.36	0.10	0.00	0.10	0.01	0.10	0.00	0.10	0.00	0.10
	μ_2	0.00	0.00	0.12	0.01	0.11	0.00	0.13	0.00	0.14	0.00	0.13
	σ_1	1.00	0.85	0.07	0.99	0.07	0.99	0.07	1.00	0.07	0.99	0.07
	σ_2	1.00	1.00	0.08	0.99	0.07	0.99	0.08	1.00	0.09	0.99	0.09
	ρ	0.00	0.00	0.12	0.00	0.11	0.00	0.14	0.00	0.15	-0.01	0.14
MNAR	μ_1	0.00	0.00	0.12	-0.01	0.10	0.00	0.10	0.00	0.10	0.00	0.10
	μ_2	0.00	0.36	0.10	0.35	0.10	0.36	0.10	0.36	0.11	0.12	0.28
	σ_1	1.00	0.98	0.08	1.00	0.07	0.99	0.07	1.00	0.07	0.99	0.07
	σ_2	1.00	0.85	0.07	0.85	0.07	0.85	0.07	0.87	0.08	1.01	0.16
	ρ	0.00	0.00	0.12	0.00	0.12	0.00	0.12	0.00	0.13	0.00	0.11
True value of $\rho=0.5$										
MCAR	μ_1	0.00	0.00	0.11	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10
	μ_2	0.00	0.00	0.11	0.00	0.11	0.00	0.11	0.00	0.11	0.00	0.11
	σ_1	1.00	0.99	0.08	1.00	0.07	0.99	0.07	1.00	0.07	0.99	0.07
	σ_2	1.00	0.99	0.08	0.99	0.08	0.99	0.08	1.00	0.08	0.99	0.08
	ρ	0.50	5.00	0.09	0.50	0.08	0.51	0.08	0.50	0.08	0.50	0.08
MAR	μ_1	0.00	0.37	0.10	0.00	0.10	-0.01	0.10	0.00	0.10	0.00	0.10
	μ_2	0.00	0.18	0.12	0.00	0.11	-0.01	0.13	0.00	0.13	-0.01	0.12
	σ_1	1.00	0.86	0.07	0.99	0.07	0.99	0.07	1.00	0.07	0.99	0.07
	σ_2	1.00	0.96	0.08	1.00	0.08	0.99	0.09	1.00	0.09	0.99	0.09
	ρ	0.50	0.45	0.10	0.50	0.08	0.50	0.10	0.50	0.10	0.50	0.10
MNAR	μ_1	0.00	0.18	0.12	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.10
	μ_2	0.00	0.36	0.10	0.29	0.10	0.29	0.10	0.29	0.11	0.00	0.15
	σ_1	1.00	0.96	0.08	1.00	0.07	0.99	0.07	1.00	0.07	0.99	0.07
	σ_2	1.00	0.85	0.07	0.86	0.07	0.86	0.08	0.87	0.08	1.01	0.12
	ρ	0.50	0.45	0.10	0.46	0.10	0.455	0.10	0.45	0.10	0.50	0.08

Table 2. Estimated parameter by using the CC, available cases using direct maximization (DML), EM, MI and joint modelling (JM) for a simulated data (n=1000) with MCAR, MAR and MNAR mechanisms (*: mean of the standard errors over 1000 iteration obtained using information matrix, **: sampling standard error)

Mec.	Par.	True	CC		DML		EM		MI		JM	
			Es.	Se.*	Es.	Se.*	Es.	Se.**	Es.	Se.**	Es.	Se.*
True value of $\rho = 0$												
MCAR	μ_1	0.00	0.00	0.04	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
	μ_2	0.00	0.00	0.04	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
	σ_1	1.00	1.00	0.03	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	σ_2	1.00	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.03	1.00	0.03
	ρ	0.00	0.00	0.04	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
MAR	μ_1	0.00	0.36	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
	μ_2	0.00	0.00	0.04	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
	σ_1	1.00	0.86	0.02	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	σ_2	1.00	1.00	0.03	1.00	0.02	1.00	0.03	1.00	0.03	1.00	0.03
	ρ	0.00	0.00	0.04	0.00	0.04	0.00	0.04	0.00	0.04	0.00	0.04
MNAR	μ_1	0.00	0.00	0.04	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
	μ_2	0.00	0.36	0.03	0.36	0.03	0.36	0.03	0.36	0.03	0.04	0.11
	σ_1	1.00	1.00	0.03	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	σ_2	1.00	0.86	0.02	0.86	0.02	0.86	0.02	0.86	0.02	0.99	0.07
	ρ	$\tilde{0}00$	0.00	0.04	0.00	0.04	0.00	0.04	0.00	0.04	0.00	0.04
True value of $\rho = 0.5$												
MCAR	μ_1	0.00	0.00	0.04	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
	μ_2	0.00	0.00	0.04	0.00	0.03	0.00	0.03	0.00	0.04	0.00	0.04
	σ_1	1.00	1.00	0.03	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	σ_2	1.00	1.00	0.03	1.00	0.03	1.00	0.03	1.00	0.02	1.00	0.03
	ρ	0.50	0.50	0.03	0.50	0.03	0.50	0.03	0.50	0.03	0.50	0.03
MAR	μ_1	0.00	0.36	0.03	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
	μ_2	0.00	0.18	0.04	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
	σ_1	1.00	0.86	0.02	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	σ_2	1.00	0.97	0.03	1.00	0.02	1.00	0.03	1.00	0.03	1.00	0.03
	ρ	0.50	0.44	0.03	0.50	0.03	0.50	0.03	0.50	0.03	0.50	0.03
MNAR	μ_1	0.00	0.18	0.04	0.00	0.03	0.00	0.03	0.00	0.03	0.00	0.03
	μ_2	0.00	0.36	0.03	0.29	0.03	0.29	0.03	0.29	0.03	0.00	0.05
	σ_1	1.00	0.97	0.03	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.02
	σ_2	1.00	0.86	0.02	0.87	0.02	0.87	0.02	0.87	0.02	1.00	0.04
	ρ	0.50	0.445	0.03	0.46	0.03	0.46	0.03	0.46	0.03	0.50	0.03

Under the non-MCAR mechanisms, the CC estimate is biased whenever $\rho \neq 0$. The DML estimates, however, is unbiased and consistent under the MCAR and MAR mechanisms. From standpoint of bias and consistency, the ML estimate by DML has a clear advantage over the CC estimate. Under the more restrictive condition of MCAR, both the DML and CC estimates are unbiased, but DML still has an advantage over CC for $\rho = 0.5$, because its variance is lower. This reduction in variance occurs because Y_1 becomes an increasingly valuable predictor of the missing values of Y_2 as ρ increases. The EM and MI estimates, like

DML estimates, are unbiased and consistent under the MCAR and MAR mechanisms. However, MI estimates, comparing with EM estimates, are closer to the true values when $n = 100$.

All ignorable approaches (CC, DML, EM and MI) are biased under MNAR mechanism, but JM estimate is unbiased and consistent. Under this mechanism missing values are related to Y_2 rather than Y_1, but as ρ increases Y_1 becomes an increasingly useful proxy for Y_2 in JM approach and that decreases the variance of estimates in this approach. The variance of estimates in this approach may seem to be larger than those of EM or MI approaches MNAR mechanism, but one expect this to occurs as JM reflects the uncertainly due to the missing values by larger standard errors for estimates related to the parameters of Y_2 distribution. This uncertainly decreases as the number of individuals increase or the correlation between two responses increases. The JM estimates and their standard errors are also unbiased and consistent under MCAR and MAR mechanisms. It is also worth to mention that JM is the only approach that gives some information about the missing mechanism, information like the rate of response in our example or dependence of missing mechanism on some explanatory variables in general.

5. Empirical application: Mastitis data

Mastitis can reduce the milk yield of infected animals. Diggle and Kenward (1994) used the total milk yield (in thousands of litters) for 107 cows from a single herd, in two consecutive years, to investigate the relationship between yield and mastitis. Of the 107 animals, 27 were infected in their second year. The 27 animals with mastitis in their second year were treated as missing.

A bivariate normal model was used to display milk yield as a response during two years, i.e., yields that might be observed in the absence of any mastitis infection. We have (for $i = 1,2,...,107$),

$$\mathbf{Y_i} = \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix} \sim N\left(\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \ \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right).$$

Pseudo-complete data are in the $Y_{n\times2} = (Y_{obs}, Y_{mis})$ term. y_{ij} is the milk yield for the ith cow in the jth year ($i=1,2,...,107$, j=1,2). The parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ need to be estimated.

Diggle and Kenward (1994), Crouchley and Ganjali (2002) and Ganjali and Jolani (2004) analyze the same data. However, none of these analyses gives EM and MI estimates. For comparative purposes we reanalyze these data by all DML, EM, MI, and JM approaches. To obtain an estimate of the variance of the estimates in EM and MI we use the following bootstrap approach 200 times (Little and Rubin, 2002):
 (a) Generate a bootstrap sample from the original unimputed sample,
(b) Supply information where the data are missing by applying the EM or MI to the bootstrap sample,
(c) Compute estimates using the filled-in data from (b).
Then, sample variances of these 200 estimates are used as bootstrap estimates of the variances of the parameter estimates (we give the square root of these variances as BSe. in Table 3). In Table 3 the parameters estimated by EM and MI (m=10, Jeffreys prior) are given using package SAS, version 9.1.

Table 3. Estimated parameter by using the complete cases (CC), available cases using direct maximization (DML), EM, MI and joint modeling (BSe.: bootstrap standard error)

Parameter	CC		DML		EM		MI		JM	
	Es.	Se.	Es.	SE.	Es.	BSe.	Es.	BSe.	Es.	Se.
μ_1	5.708	0.107	5.766	0.090	5.765	0.094	5.765	0.089	5.765	0.090
μ_2	6.444	0.128	6.523	0.107	6.483	0.119	6.468	0.125	6.080	0.146
σ_1	0.955	0.076	0.931	0.064	0.931	0.062	0.935	0.064	0.931	0.064
σ_2	1.148	0.091	1.094	0.074	1.138	0.103	1.169	0.116	1.274	0.113
ρ	0.592	0.073	0.519	0.069	0.581	0.111	0.596	0.097	0.470	0.087

By viewing this table we conclude that estimates of the values of μ_1 and μ_2, due to the EM and MI methods, are very close to each other, but the estimated covariance matrices (Σ) are different. Using the EM algorithm the estimates of σ_1 and σ_2 are less than those using the MI method. The estimate of ρ obtained using the MI method is slightly greater than that found using the EM algorithm. The DML estimates are close to the estimates of EM or MI, but the estimates of standard errors of DML is less than those of bootstrap estimate of EM or MI approaches.

Crouchley and Ganjali (2002), whose conclusions are based on the likelihood ratio test, reject the MAR assumption in these data. They also found 3 outliers (cows 4, 5 and 66) in the data. MI and EM can be used when the MAR assumption is true. If data show a MNAR mechanism one should use a joint model. Generalized Heckman selection model for the mastitis data is formed as below:

$$y_{i1}^* = \mu_1 + \varepsilon_{i1}, \ \ y_{i2}^* = \mu_2 + \varepsilon_{i2}, \ R_{i2}^* = \alpha_0 + \nu_{i2}. \tag{7}$$

The results using this model, obtained by minimizing the negative logarithm of the likelihood using the NAG (1996) routine EO4UCF, are also given in table 3. Using this method, as the data are MNAR, the estimate of μ_2 is less than those using the EM or MI and the estimate of the covariance matrix is completely different (except for the element σ_1^2).

After eliminating outliers, the results using CC, AC, the EM algorithm, MI and generalized Heckman selection model, are given in Table 4.

Table 4. Estimated parameters after eliminating the 4th, 5th and 66th cows by using the complete cases (CC), available cases using direct maximization (DML), EM, MI and joint modeling (BSe.: bootstrap standard error)

Parameter	CC		DML		EM		MI		JM	
	Es.	Se.	Es.	Se.	Es.	BSe.	Es.	BSe.	Es.	Se.
μ_1	5.750	0.100	5.798	0.086	5.798	0.083	5.798	0.086	5.798	0.086
μ_2	6.358	0.119	6.447	0.101	6.398	0.111	6.386	0.118	6.400	0.110
σ_1	0.880	0.071	0.872	0.060	0.872	0.056	0.864	0.051	0.872	0.060
σ_2	1.047	0.084	1.030	0.073	1.042	0.073	1.071	0.077	1.042	0.080
ρ	0.730	0.053	0.742	0.045	0.726	0.052	0.735	0.058	0.727	0.052

This table shows that the results using different methods (after removing outliers) are close. We expect this because the EM and MI methods are based on the MAR assumption and, when data are MAR, joint modeling should give close results.

6. Conclusions

In this paper, the results from using complete cases (CC), direct maximization (DML), the EM, MI and joint modeling methods, to analyze multivariate normal data with missing responses, are reviewed and compared in a simulation study. The CC analysis gives biased and inconsistent estimates under not-MCAR mechanism. To use DML, EM or MI one has to make the assumption of MAR. If this assumption is not valid for the data, we may reach incorrect inferences. On the other hand, to test for MAR we have to use joint modeling of responses and of the missing mechanism where the test for normality is impossible, due to the fact that part of the data being missing, and the results may be sensitive to the model assumptions.

Two points are worth mentioning; (1) detection of outliers is crucial in reaching to the final conclusion and (2) if joint modeling gives results which are considerably different from those of DML, EM or MI, it may be because of the fact that responses are MNAR. In the latter case, if you do not trust the results of joint modeling, use a method with the MAR assumption, but sensitivity analysis to explore potential deviation from the MAR assumption needs to be done (Janson et. al., 2006; Ganjali and Rezaei, 2005).

REFERENCES

Crouchley, R. and M. Ganjali (2002). The Common Structure of Several Recent Statistical Models For Dropout In Repeated Continuous Responses. Statistical Modeling, **2**, pp. 39-62.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum Likelihood From Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society, Series B, **39**, pp. 1-38.

Diggle, P. J. and M. G. Kenward (1994). Informative Dropout in Longitudinal Data Analysis. Applied Statistics, **4**3, pp. 49-93.

Everitt, B. S. (1987). Introduction to Optimization Methods and Their Application in Statistics. Chapman & Hall, London.

Follmann, D. and M. Wu (1995). An Approximation Generalized Linear Model with Random Effects for Informative Missing Data. Biometrics, **51**, pp. 151-168.

Ganjali, M. and S. Jolani (2004). Moments of the Truncated Multivariate Normal Distribution and Their Applications to Censored Regression Models. Journal of Statistical Theory and Applications, **3**, pp. 189-200.

Ganjali, M. and M. Rezaei (2005). An Influence Approach for Sensitivity Analysis of Non-Random Dropout Based on the Covariance Structure. Iranian Journal of Science and Technology, Transaction A, **29**, No. A2, pp. 287-294.

Hausman, J. A. and D. A. Wise (1979). Attrition Bias in Experimental and Panel Data: The

Gary Income Maintenance Experiment, Econometrica, **47**, pp. 455-473.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. Econometrica, **47**, pp. 153-161.

Jansen, I., N. Hens, G. Molenberghs, M. Aerts, G. Verbeke, and G. Kenward(2006). The Nature of Sensitivity in Monotone Missing Not At Random Models. Computational Statistics and Data Analysis, **50**, pp. 830-858.

Little, R. J. A. and D. B. Rubin(2002). Statistical Analysis with Missing Data. Wiley, New York.

Kenward, M. G. (1998). Selection Models for Repeated Measurements with Nonrandom Dropout: An Illustration of Sensitivity. Statistics in Medicine, **17**, 2, pp. 723-2732.

Louis, T.A. (1982). Finding the Observed Information Matrix When Using the EM Algorithm. J. Roy. Statist. Soc. B, **44**, pp. 226-232.

McLachlan, G. J. and T. Krishnan(1997). The EM Algorithm and Extensions. Wiley, New York.

Meilijson, I. (1989). A Fast Improvement to the EM Algorithm on Its Own Terms. J. Roy. Statist. Soc. B, **51**, pp. 127-138.

Meng, X. L., and D. B. Rubin (1991). Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm. J. Amer. Statist. Assoc., **86**, pp. 899-909.

Molenberghs G., G. Verbeke, H. Thijs, E. Lesaffre, and M. G. Kenward (2001). Mastitis in Dairy Cattle: Influence Analysis To Assess Sensitivity Of The Dropout Process. Computational Statistics and Data Analysis, **37**, pp. 93-113.

NAG (1996). Numerical Algorithms Group Manual, Mark 16. Oxford, U.K.

Ridder, G. (1990). Attrition in Multi-Wave Panel Data. In Panel Data in Labor Market Studies (Hattog. J., Ridder. G. and Theeuwes. J., Editors), pp. 45-76, Elsevier Science B.V., Amsterdam, North Holland.

Rubin, D. B. (1976). Inference and Missing Data. Biometrika, 63, pp. 581-592.

Rubin, D. B. (1977). The Design of a General and Flexible System for Handling Nonresponse In Sample Surveys. Manuscript prepared for the US Social Security Administration.

Rubin, D. B. (1978). Multiple Imputation in Sample Surveys - A Phenomenological Bayesian Approach to Nonresponse. Proceedings of the Survey Research Methods Section of the American Statistical Association, pp. 20-34. Also in: Imputation and Editing of Faulty or Missing Survey Data. US Dept. of Commerce, Bureau of the Census, pp. 1-23.

Rubin, D. B., Multiple Imputation for Nonresponse in Surveys. Wiley, New York, (1987).

Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. Chapman & Hall, New York.

Tang, G., R. J. A. Little, and T. E. Raghunathan (2003). Analysis of Multivariate Missing Data with Nonignorable Nonresponse. Biometrika, **90**, 4, pp. 747-764.

Thisted, R. A. (1988). Elements of Statistical Computing: Numerical Computation. Chapman & Hall, London.

Wu, M. C. and R. J. Carroll (1988). Estimation and Comparison of Changes in the Presence of Informative Censoring by Modeling the Censoring Process, Biometrics, **44**, pp. 175-188.