Published in final edited form as:

Am Stat. 2010 August 1; 64(3): 263–267. doi:10.1198/tast.2010.09203.

Consistency of Normal Distribution Based Pseudo Maximum Likelihood Estimates When Data Are Missing at Random

Ke-Hai Yuan

University of Notre Dame

Peter M. Bentler

University of California, Los Angeles

Abstract

This paper shows that, when variables with missing values are linearly related to observed variables, the normal-distribution-based pseudo MLEs are still consistent. The population distribution may be unknown while the missing data process can follow an arbitrary missing at random mechanism. Enough details are provided for the bivariate case so that readers having taken a course in statistics/probability can fully understand the development. Sufficient conditions for the consistency of the MLEs in higher dimensions are also stated, while the details are omitted.

Keywords

Consistency; maximum likelihood; model misspecification; missing data

1. Introduction

Incomplete or missing data exist in almost all areas of empirical research. They are especially common when data are collected longitudinally and/or by surveys. There can be various reasons for missing data to occur. The process by which data become incomplete was called the missing data mechanism by Rubin (1976). Missing completely at random (MCAR) is a process in which missingness of data is independent of both the observed and the missing values; missing at random (MAR) is a process in which missingness is independent of the missing values given the observed data. When missingness depends on the missing values themselves given the observed data, the process is not missing at random (NMAR). Missing data with an NMAR mechanism are also referred to as non-ignorable non-responses because maximum likelihood estimates (MLE), by ignoring the missing data mechanism, are generally inconsistent. This paper studies the consistency of the normal-distribution-based MLE with MAR data.

In contrast to many ad hoc procedures for missing data analysis, MLEs have the desired property of being consistent even when the specific MAR mechanism is ignored. When modeling real data, however, specifying the correct distribution form to obtain the true MLE is always challenging if not impossible. The normal distribution is often chosen for convenience, not because practical data tend to come from normally distributed populations. Geary (1947, p. 241) observed that “Normality is a myth; there never was, and never will be a normal distribution.” Such an observation was further supported by Micceri (1989), who examined 440 data sets obtained from journal articles, research projects as well as tests, and

found that all were significantly nonnormally distributed. Thus, the normal-distribution-based MLEs for real data typically are pseudo MLEs, whose properties have been obtained by White (1982) and Gourieroux, Monfort and Trognon (1984) in the context of complete data. With missing data, however, according to Laird (1988) and Rotnitzky and Wypij (1994), pseudo MLEs will be inconsistent unless the missing data mechanism is MCAR. The need for a correct likelihood function with an MAR mechanism was also noted by Liang and Zeger (1986) and Little (1993). If a pseudo MLE is not consistent when data are MAR, then only the MCAR mechanism can be ignored when modeling practical multivariate data with unknown population distributions. Thus, in addition to being an important mathematical property, consistency of the normal-distribution-based pseudo MLE with MAR data also has wide implications for many areas of applied statistics where the normal distribution is routinely used to model missing data.

Let $\mathbf{x} = (x_1, x_2, \dots, x_q)'$ be a vector representing a q -variate population, $\mathbf{x}_o$ contain the observed values and $\mathbf{x}_m$ contain the missing values in a realization of $\mathbf{x}$. Let $\mathbf{r} = (r_1, r_2, \dots, r_q)'$, where $r_j = 1$ if x_j is observed and 0 otherwise. The missing data in $\mathbf{x}_m$ are MAR if

$$P(r_j=0|\mathbf{x}_o, \mathbf{x}_m) = P(r_j=0|\mathbf{x}_o) = g_j(\mathbf{x}_o, \gamma_j), j=1, 2, \dots, q \quad (1)$$

and the r_j 's are conditionally/locally independent given $\mathbf{x}$, where γ_j is a vector of parameters. In practice, three popular forms of $g_j(\mathbf{x}_o, \gamma_j)$ are the interval selection model (see e.g., Schafer, 1997, p. 25)

$$g_j(\mathbf{x}_o, \gamma_j) = 1$$

when $\mathbf{x}_o$ falls into certain hyper-rectangles; the probit selection model

$$g_j(\mathbf{x}_o, \gamma_j) = \Phi(\gamma_{j0} + \gamma'_{j1}\mathbf{x}_o),$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$ and γ_{j1} contains the regression coefficients; and the logistic selection model

$$g_j(\mathbf{x}_o, \gamma_j) = \frac{\exp(\gamma_{j0} + \gamma'_{j1}\mathbf{x}_o)}{1 + \exp(\gamma_{j0} + \gamma'_{j1}\mathbf{x}_o)}.$$

The interval selection model is widely used in economics (Amemiya, 1973; Heckman, 1979) while the probit and logistic selection models are commonly used in many other disciplines (Allison 2001; Little & Rubin, 2002; Molenberghs & Kenward, 2007; Daniels & Hogan, 2008). Under the interval selection model, Yuan (2009) showed that the normal-distribution-based pseudo MLEs are consistent and asymptotically normally distributed even when the underlying distribution is unknown. The purpose of this note is to extend the result of Yuan (2009) by showing that the normal-distribution-based MLEs are still consistent when the underlying population is unknown and when the $g_j(\mathbf{x}_o, \gamma_j)$ in (1) is any function of the observed data, including both the probit and logistic selection models.

While all the results in Yuan (2009) can be generalized to an MAR mechanism described by probit and logistic selection models, for brevity we only give the details for the bivariate case in section 2. With missing data and a misspecified likelihood function, little literature

exists that facilitates thorough understanding of issues related to consistency. We choose this simple model with enough details so that readers having taken a course in statistics/probability can fully understand the development. We will also state the results for consistency for a general q in section 3, but the details will be omitted. We conclude the paper by pointing out that not all pseudo MLEs are consistent.

2. Consistency in the bivariate case

Let $\mathbf{x} = (x_1, x_2)'$ with

$$\mu = E(\mathbf{x}) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \Sigma = \text{Cov}(\mathbf{x}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \quad (2)$$

where $\sigma_{11} = \sigma_1^2$, $\sigma_{22} = \sigma_2^2$ and $\sigma_{12} = \sigma_{21} = \sigma_1 \sigma_2 \rho$. A sample from $\mathbf{x}$ with missing values in x_2 can be represented by

$$\begin{aligned} & x_{11}, \dots, x_{n1}, x_{(n+1)1}, \dots, x_{N1} \\ & x_{12}, \dots, x_{n2}, \end{aligned} \quad (3)$$

where $x_{(n+1)2}, \dots, x_{N2}$ are missing. The interest is to infer (2) based on (3) using the possibly wrong assumption $\mathbf{x} \sim N_2(\mu, \Sigma)$. Notice that the number of cases with x_{i2} missing is not controllable so that the n in (3) is a random number.

With two variables, there are a total of 4 possible observed patterns: (x_{i1}, x_{i2}) , $(x_{i1}, \cdot)$, $(\cdot, x_{i2})$ and $(\cdot, \cdot)$. The sample in (3) only contains two of the four. We chose this sample because the MLEs enjoy analytical solutions and the proof of their consistency is simple enough to be understood by a broad audience. We will discuss the consistency of the MLEs with more missing data patterns and more variables in section 3.

Let

$$\begin{aligned} \bar{x}_1 &= \frac{1}{N} \sum_{i=1}^N x_{i1}, \quad s_{11} = \frac{1}{N} \sum_{i=1}^N (x_{i1} - \bar{x}_1)^2, \\ \bar{x}_{1*} &= \frac{1}{n} \sum_{i=1}^n x_{i1}, \quad s_{11*} = \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_{1*})^2, \\ \bar{x}_{2*} &= \frac{1}{n} \sum_{i=1}^n x_{i2}, \quad s_{22*} = \frac{1}{n} \sum_{i=1}^n (x_{i2} - \bar{x}_{2*})^2, \\ s_{21*} &= \frac{1}{n} \sum_{i=1}^n (x_{i2} - \bar{x}_{2*})(x_{i1} - \bar{x}_{1*}), \quad \hat{\beta}_{21} = \frac{s_{21*}}{s_{11*}} = \frac{\sum_{i=1}^n (x_{i2} - \bar{x}_{2*})(x_{i1} - \bar{x}_{1*})}{\sum_{i=1}^n (x_{i1} - \bar{x}_{1*})^2}. \end{aligned}$$

Then it follows from Anderson (1957) that the MLEs of $(\mu_1, \sigma_{11}, \mu_2, \sigma_{22}, \sigma_{12})$ based on the normal distribution assumption by ignoring the missing data mechanism are

$$\hat{\mu}_1 = \bar{x}_1, \quad \hat{\sigma}_{11} = s_{11}, \quad (4a)$$

$$\hat{\mu}_2 = \bar{x}_{2*} + \hat{\beta}_{21} (\bar{x}_1 - \bar{x}_{1*}), \quad (4b)$$

$$\widehat{\sigma}_{22} = s_{22*} + \widehat{\beta}_{21}^2 (\widehat{\sigma}_{11} - s_{11*}), \quad (4c)$$

$$\widehat{\sigma}_{12} = \widehat{\beta}_{21} \widehat{\sigma}_{11}. \quad (4d)$$

Through the work of Rubin (1976) and others, it is widely known that $\widehat{\mu}_2$, $\widehat{\sigma}_{22}$, $\widehat{\sigma}_{12}$ are consistent when all the missing data in (3) are MCAR or MAR and $\mathbf{x} \sim N_2(\mu, \Sigma)$. In the following, we will study the consistency of $\widehat{\mu}_2$, $\widehat{\sigma}_{22}$ and $\widehat{\sigma}_{12}$ using (4) when $\mathbf{x}$ does not follow the bivariate normal distribution and the missing x_{i2} 's in (3) are MAR. For such a purpose, we assume that the population for the data in (3) is

$$x_1 = \mu_1 + \sigma_1 z_1, \quad x_2 = \mu_2 + \sigma_2 \left[\rho z_1 + (1 - \rho^2)^{1/2} z_2 \right], \quad (5)$$

where z_1 and z_2 are independent with $E(z_1) = E(z_2) = 0$ and $\text{Var}(z_1) = \text{Var}(z_2) = 1$. Clearly, (5) follows a bivariate normal distribution only when $z_1 \sim N(0, 1)$ and $z_2 \sim N(0, 1)$. However, the population mean vector and variance-covariance matrix of $\mathbf{x} = (x_1, x_2)'$ remain the same regardless of the distributions of z_1 and z_2 .

Corresponding to (3) there exist independent random variables z_{i1} and z_{i2} such that

$$x_{i1} = \mu_1 + \sigma_1 z_{i1}, \quad x_{i2} = \mu_2 + \sigma_2 \left[\rho z_{i1} + (1 - \rho^2)^{1/2} z_{i2} \right].$$

Let

$$\begin{aligned} \bar{z}_{1*} &= \frac{1}{n} \sum_{i=1}^n z_{i1}, \quad \bar{z}_{2*} = \frac{1}{n} \sum_{i=1}^n z_{i2}, \quad s_{z11*} = \frac{1}{n} \sum_{i=1}^n (z_{i1} - \bar{z}_{1*})^2, \\ s_{z21*} &= \frac{1}{n} \sum_{i=1}^n (z_{i2} - \bar{z}_{2*})(z_{i1} - \bar{z}_{1*}), \quad s_{z22*} = \frac{1}{n} \sum_{i=1}^n (z_{i2} - \bar{z}_{2*})^2. \end{aligned}$$

Then

$$\bar{x}_{1*} = \mu_1 + \sigma_1 \bar{z}_{1*}, \quad s_{11*} = \sigma_1^2 s_{z11*}, \quad s_{21*} = \sigma_2 \sigma_1 \left[\rho s_{z11*} + (1 - \rho^2)^{1/2} s_{z21*} \right], \quad (6a)$$

$$\bar{x}_{2*} = \mu_2 + \sigma_2 \left[\rho \bar{z}_{1*} + (1 - \rho^2)^{1/2} \bar{z}_{2*} \right], \quad s_{22*} = \sigma_2^2 \left[\rho^2 s_{z11*} + 2\rho(1 - \rho^2)^{1/2} s_{z21*} + (1 - \rho^2) s_{z22*} \right]. \quad (6b)$$

The equations in (6) allow us to obtain the probability limits of $\bar{x}_{j*}$ and s_{jk*} through those of $\bar{z}_{j*}$ and s_{zjk*} , which further lead to consistency of $\widehat{\mu}_2$, $\widehat{\sigma}_{12}$ and $\widehat{\sigma}_{22}$ in (4).

We also need to connect the observations in (3) to the MAR mechanism. Let $r_i = 1$ if the x_{i2} in (3) is observed and $r_i = 0$ if the x_{i2} in (3) is missing. Because x_{i1} and z_{i1} are uniquely determined by each other, we can rewrite the MAR mechanism in equation (1) as

$$P(r_i=0|x_{i1}, x_{i2})=P(r_i=0|x_{i1})=P(r_i=0|z_{i1})=g(x_{i1}, \gamma)=h(z_{i1}), \quad (7)$$

where the parameter vector γ is omitted from $h(\cdot)$. Let the probability density functions (pdf) of z_1 and z_2 be $f_1(t)$ and $f_2(t)$, respectively. Then n , the number of complete cases in (3), follows the binomial distribution $B(N, p_o)$, where, with I being an indicator function,

$$\begin{aligned} p_o &= P(r_i=1) = E(I_{\{r_i=1\}}) = E(1 - I_{\{r_i=0\}}) \\ &= 1 - E[E(T_{\{r_i=0\}}|x_{i1})] = 1 - E[P(r_i=0|x_{i1})] \\ &= 1 - E[h(z_{i1})]. \end{aligned}$$

Let t_{i1} be the realized value of z_{i1} and f be a generic notation for the probability distribution/density function of the involved random variables. It follows from (7) and

$$f(r_i=1, t_{i1}) = p_o f(t_{i1}|r_i=1) = P(r_i=1|t_{i1}) f_1(t_{i1})$$

that

$$f(t_{i1}|r_i=1) = \frac{1}{p_o} [1 - h(t_{i1})] f_1(t_{i1}).$$

Thus, the z_{i1} corresponding to the observed x_{i2} are independent, identically distributed (iid), and each follows the distribution with pdf

$$f_{1*}(t) = \frac{1}{p_o} [1 - h(t)] f_1(t).$$

Notice that, due to the MAR mechanism, the missingness in (3) has nothing to do with z_{i2} . Each z_{i2} corresponding to either the observed x_{i2} or missing x_{i2} still has the same distribution as $z_2 \sim f_2(t)$.

Let the mean and variance of $u \sim f_{1*}(t)$ be μ_{z1*} and σ_{z1*} . Let $u_1, u_2, \dots, u_N$ be iid with pdf $f_{1*}(t)$; $\omega_1, \omega_2, \dots, \omega_N$ be iid with $P(\omega_i = 1) = p_o$ and $P(\omega_i = 0) = 1 - p_o$; and the ω_i 's be independent of u_i 's and z_{i2} 's. Then n and $\sum_{i=1}^N \omega_i$ have the same distribution; $\sum_{i=1}^n z_{i1}$ and $\sum_{i=1}^N \omega_i u_i$ have the same distribution; $\sum_{i=1}^n z_{i1}^2$ and $\sum_{i=1}^N \omega_i u_i^2$ have the same distribution; $\sum_{i=1}^n z_{i1} z_{i2}$ and $\sum_{i=1}^N \omega_i u_i z_{i2}$ have the same distribution; $\sum_{i=1}^n z_{i2}^2$ and $\sum_{i=1}^N \omega_i z_{i2}^2$ have the same distribution. A brief proof for the above relationships using characteristic functions is provided in Appendix A of Yuan and Lu (2008).

We are now ready to show the result of consistency. Applying the law of large numbers to the average of ω_i yields

$$\frac{n}{N} = \frac{1}{N} \sum_{i=1}^N \omega_i \xrightarrow{wpl} p_o,$$

where the equal sign follows from equivalence in distribution and $\xrightarrow{wp1}$ is the notation for convergence with probability one. Continuously applying equivalence in distribution and the law of large numbers to the averages of $\omega_i u_i$ and $\omega_i z_{i2}$, respectively, we have

$$\bar{z}_{1*} = \frac{\sum_{i=1}^N \omega_i u_i / N}{n/N} \xrightarrow{wp1} \frac{p_o E(u)}{p_o} = \mu_{z1*} \quad (8)$$

and

$$\bar{z}_{2*} = \frac{\sum_{i=1}^N \omega_i z_{i2} / N}{n/N} \xrightarrow{wp1} \frac{p_o E(z_2)}{p_o} = 0. \quad (9)$$

Similarly,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n z_{i1}^2 &= \frac{\sum_{i=1}^N \omega_i u_i^2 / N}{n/N} \xrightarrow{wp1} E(u^2); \\ \frac{1}{n} \sum_{i=1}^n z_{i1} z_{i2} &= \frac{\sum_{i=1}^N \omega_i u_i z_{i2} / N}{n/N} \xrightarrow{wp1} E(u) E(z_2) = 0; \\ \frac{1}{n} \sum_{i=1}^n z_{i2}^2 &= \frac{\sum_{i=1}^N \omega_i z_{i2}^2 / N}{n/N} \xrightarrow{wp1} E(z_2^2) = 1;. \end{aligned}$$

Thus,

$$s_{z11*} = \frac{1}{n} \sum_{i=1}^n z_{i1}^2 - \bar{z}_{1*}^2 \xrightarrow{wp1} E(u^2) - \mu_{z1*}^2 = \sigma_{z11*}; \quad (10)$$

$$s_{z21*} = \frac{1}{n} \sum_{i=1}^n z_{i1} z_{i2} - \bar{z}_{1*} \bar{z}_{2*} \xrightarrow{wp1} 0; \quad (11)$$

$$s_{z22*} = \frac{1}{n} \sum_{i=1}^n z_{i2}^2 - \bar{z}_{2*}^2 \xrightarrow{wp1} 1. \quad (12)$$

It is obvious that

$$\widehat{\mu}_1 = \bar{x}_1 \xrightarrow{wp1} \mu_1 \text{ and } \widehat{\sigma}_{11} = s_{11} \xrightarrow{wp1} \sigma_{11}. \quad (13)$$

Regarding μ_1 , μ_2 , σ_1 , σ_2 and ρ as constants, $\bar{x}_{1*}$, $\bar{x}_{2*}$, s_{11*} , s_{21*} and s_{22*} in (6) are just linear combinations of $\bar{z}_{1*}$, $\bar{z}_{2*}$, s_{z11*} , s_{z21*} and s_{z22*} , whose probability limits have already been obtained. Combining (6a), (8), (10) and (11) yields

$$\bar{x}_{1*} \xrightarrow{wp1} \mu_{1*} = \mu_1 + \sigma_1 \mu_{z1*}, \quad s_{11*} \xrightarrow{wp1} \sigma_{11*} = \sigma_{11} \sigma_{z11*}, \text{ and } s_{21*} \xrightarrow{wp1} \sigma_2 \sigma_1 \sigma_{z11*} \rho. \quad (14)$$

Combining (6b) and (8) to (12) yields

$$\bar{x}_* \xrightarrow{wp1} \mu_2 + \sigma_2 \rho \mu_{z1*} \text{ and } s_{22*} \xrightarrow{wp1} \sigma_{22} \left[\rho^2 \sigma_{z11*} + (1 - \rho^2) \right]. \quad (15)$$

Thus,

$$\widehat{\beta}_{21} = \frac{s_{21*}}{s_{11*}} \xrightarrow{wp1} \frac{\sigma_2}{\sigma_1} \rho. \quad (16)$$

It follows from (4b) and (13) to (16) that

$$\widehat{\mu}_2 \xrightarrow{wp1} \mu_2 + \sigma_2 \rho \mu_{z1*} + \frac{\sigma_2}{\sigma_1} \rho [\mu_1 + (\mu_1 + \sigma_1 \mu_{z1*})] = \mu_2.$$

So $\widehat{\mu}_2$ is consistent. It follows from (4c) and (13) to (16) that

$$\widehat{\sigma}_{22} \xrightarrow{wp1} \sigma_{22} \left[\rho^2 \sigma_{z11*} + (1 - \rho^2) \right] + \frac{\sigma_{22} \rho^2}{\sigma_{11}} (\sigma_{11} - \sigma_{11} \sigma_{z11*}) = \sigma_{22}.$$

So $\widehat{\sigma}_{22}$ is also consistent. It follows from (4d), (13) and (16) that

$$\widehat{\sigma}_{12} \xrightarrow{wp1} \frac{\sigma_2}{\sigma_1} \rho \sigma_{11} = \sigma_{12}.$$

So $\widehat{\sigma}_{12}$ is, again, consistent.

Notice that the $g(\cdot, \cdot)$ or $h(\cdot)$ in (7) can be any function of the observed data. Thus, the normal-distribution-based pseudo MLEs are consistent for any MAR process.

3. Consistency in general

Parallel to (5), let

$$\mathbf{x} = (x_1, x_2, \dots, x_q)' = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}, \quad (17)$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_q)'$,

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ & & \cdots & & \\ a_{q1} & a_{q2} & a_{q3} & \cdots & a_{qq} \end{pmatrix}$$

and satisfies $\Sigma = \mathbf{A}\mathbf{A}'$, and $\mathbf{z} = (z_1, z_2, \dots, z_q)'$ with $z_1, z_2, \dots, z_q$ being independent and standardized random variables. Then $E(\mathbf{x}) = \boldsymbol{\mu}$, $\text{Cov}(\mathbf{x}) = \Sigma$ and the distribution of $\mathbf{x}$ is determined by those of the z_j 's. When $\mathbf{z} \sim N_q(\mathbf{0}, \mathbf{I})$, $\mathbf{x} \sim N_q(\boldsymbol{\mu}, \Sigma)$. We do not know the distribution form of $\mathbf{x}$ in general even when the distributions of z_j 's are known. Notice that

each element x_j of $\mathbf{x}$ in either (5) or (17) is a linear combination of independent random components in $\mathbf{z}$, which is a necessary condition for the consistency of the pseudo MLE $\widehat{\mu}$ and $\widehat{\Sigma}$.

Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ be a random sample drawn from $\mathbf{x}$ with $\mathbf{x}_i = \mu + \mathbf{A}\mathbf{z}_i$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iq})'$ and $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iq})'$. Consider the case $\mathbf{x}_i$ in which d variables $x_{ij_1}, x_{ij_2}, \dots, x_{ij_d}$ are missing and the event is related in probability to the c values of $z_{il_1}, z_{il_2}, \dots, z_{il_c}$. Let $J = \min(j_1, j_2, \dots, j_d)$ and $L = \max(l_1, l_2, \dots, l_c)$. When $L < J$ and $(x_{i1}, x_{i2}, \dots, x_{iL})$ are observed, all the information related to missing values is observed. Let $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{iq})'$ be the vector with $r_{ij} = 1$ if x_{ij} is observed and zero otherwise, $\mathbf{x}_{im} = (x_{ij_1}, x_{ij_2}, \dots, x_{ij_d})'$, and $\mathbf{x}_{io}$ be the vectors of observed variables. There exists

$$\begin{aligned} P(r_{ij}=0|\mathbf{x}_{io};\mathbf{x}_{im}) &= P(r_{ij}=0|\{x_{i1}, x_{i2}, \dots, x_{iL}, \dots\};\mathbf{x}_{im}) \\ &= P(r_{ij}=0|\{z_{i1}, z_{i2}, \dots, z_{il}, \dots\};\mathbf{x}_{im}) \\ &= P(r_{ij}=0|\{z_{i1}, z_{i2}, \dots, z_{il}, \dots\}) \\ &= P(r_{ij}=0|\mathbf{x}_{io}) \\ &= g_j(\mathbf{x}_{io}, \gamma_j) = h_j(z_{il}). \end{aligned} \quad (18)$$

Thus, the probability of missing only depends on the observed values and the missing data mechanism is MAR (Rubin, 1976). Notice that $\mathbf{x}_{iL} = (x_{i1}, x_{i2}, \dots, x_{iL})'$ is a subset of $\mathbf{x}_{io}$, and thus, the probability function $g_j(\mathbf{x}_{io}, \gamma_j)$ in (18) can also be written as $g_j(\mathbf{x}_{iL}, \gamma_j)$. Also notice that, due to a random process, the number d and the subscripts $j_1, j_2, \dots, j_d$ may change from case to case while $l_1, l_2, \dots, l_c$ are held constant across the sample.

Unlike the problems considered in the previous section, the MLEs do not possess analytical forms when the observed data patterns are not monotonic. So we cannot directly show that the MLEs are consistent as was done in the previous section. We cannot use the established theory of maximum likelihood as in Rubin (1976) either, because the MLEs are obtained based on an incorrect likelihood function. By showing that the normal estimating equation is unbiased at the true population values, Yuan (2009) proved that the normal-distribution-based MLEs are consistent when the missingness of x_{im} is due to $(z_{il_1}, z_{il_2}, \dots, z_{il_c})$ falling into certain hyper-rectangles. When (17) and (18) hold and $(x_{i1}, x_{i2}, \dots, x_{iL})$ are observed, using essentially the same approach as in Yuan (2009) we can show that the normal-distribution-based MLEs are consistent regardless of the distribution shape of $\mathbf{x}$ and the form of $g_j(\mathbf{x}_{io}, \gamma_j)$ in (18). The MLEs are also asymptotically normally distributed and the covariance matrix of the MLEs can be consistently estimated by a sandwich-type covariance matrix.

4. Conclusions

It has been argued that, in any statistical modeling, the distribution specification is at best only an approximation to the real world (Box, 1979). Thus, all MLEs in practice are pseudo MLEs. In the context of missing data, it is nice to know that pseudo MLEs can remain consistent when (17) and (18) hold. We need to note that data model (17) does not include nonnormal distributions created by nonlinear functions of independent random variables $z_1, z_2, \dots, z_q$; although it includes an infinite number of nonnormal distributions. Yuan (2009) describes an example in which the MLEs are not consistent when $\mathbf{x} = (x_1, x_2)'$ with $x_1 = z_1$ and $x_2 = az_1^2 + bz_2$, where z_1 and z_2 are independent, and x_2 is missing when z_1 falls into a certain interval. It can be shown that the MLEs also fail to be consistent for this data model when the MAR mechanism obeys either a logit or a probit selection process. We also would

like to note that pseudo MLEs based on an incorrect distributional specification other than $N_q(\mu, \Sigma)$ may not be consistent when missing values are MAR. Actually, even without missing data, Gourieroux et al. (1984) showed that pseudo MLEs are consistent only when the assumed distribution belongs to a quadratic exponential family.

For the purpose of allowing missingness to depend on all the linear combinations of the previously observed variables, we specified $\mathbf{A}$ as a lower triangular matrix in (17) so that $(z_1, z_2, \dots, z_L)$ and $(x_1, x_2, \dots, x_L)$ are determined by each other. In practice, a participant may join the study after missing a few times and then be missing again. The missingness at the later stage may depend on all the previously observed variables. We can match such a case with (17) by specifying that the rows of $\mathbf{A}$ that correspond to the observed variables form the upper-left part of a lower triangular matrix, then the consistency result still holds.

As with a general MAR missing data mechanism, the MAR condition in (18) cannot be tested. Without extra information beyond the observed sample, it is impossible to distinguish between MAR and NMAR mechanisms (Molenberghs et al., 2008). Similarly, the data model (17) cannot be tested either because the distribution of $\mathbf{z}$ is arbitrary.

Acknowledgments

We would like to thank the editor, an associate editor, and a referee for comments that lead to a significant improvement of the paper.

This research was supported by Grants DA00017 and DA01070 from the National Institute on Drug Abuse and a grant from the National Natural Science Foundation of China (30870784).

References

- Allison, PD. Missing data. Sage; Thousand Oaks, CA: 2001.
- Amemiya T. Regression analysis when the dependent variable is truncated normal. *Econometrica* 1973;41:997–1016.
- Anderson TW. Maximum likelihood estimates for the multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association* 1957;52:200–203.
- Box, GEP. Robustness in the strategy of scientific model building. In: Launer, RL.; Wilkinson, GN., editors. *Robustness in statistics*. Academic Press; New York: 1979. p. 201–236.
- Daniels, MJ.; Hogan, JW. Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis. Chapman & Hall; Boca Raton, Florida: 2008.
- Geary RC. Testing for normality. *Biometrika* 1947;34:209–242. [PubMed: 18918691]
- Gourieroux C, Monfort A, Trognon A. Pseudo maximum likelihood methods: Theory. *Econometrica* 1984;52:681–700.
- Heckman JJ. Sample selection bias as a specification error. *Econometrica* 1979;47:153–161.
- Laird NM. Missing data in longitudinal studies. *Statistics in Medicine* 1988;7:305–315. [PubMed: 3353609]
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
- Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993;88:125–134.
- Little, RJA.; Rubin, DB. *Statistical analysis with missing data*. 2nd ed.. Wiley; New York: 2002.
- Micceri T. The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin* 1989;105:156–166.
- Molenberghs G, Beunckens C, Sotito C, Kenward MG. Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society B* 2008;70:371–388.
- Molenberghs, G.; Kenward, MG. *Missing data in clinical studies*. Wiley; Chichester, England: 2007.

- Rotnitzky A, Wypij D. A note on the bias of estimators with missing data. *Biometrics* 1994;50:1163–1170. [PubMed: 7786998]
- Rubin DB. Inference and missing data (with discussions). *Biometrika* 1976;63:581–592.
- Schafer, JL. Analysis of incomplete multivariate data. Chapman & Hall; London: 1997.
- White H. Maximum likelihood estimation of misspecified models. *Econometrica* 1982;50:1–25.
- Yuan K-H. Normal distribution based pseudo ML for missing data: With applications to mean and covariance structure analysis. *Journal of Multivariate Analysis* 2009;100:1900–1918.
- Yuan K-H, Lu L. SEM with missing data and unknown population distributions using two-stage ML: Theory and its application. *Multivariate Behavioral Research* 2008;62:621–652.