University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2012 *Paper* 96

MISSING AT RANDOM AND IGNORABILITY FOR INFERENCES ABOUT INDIVIDUAL PARAMETERS WITH MISSING DATA

Roderick J. Little* Sahar Zanganeh†

*University of Michigan, rlittle@umich.edu

†University of Michigan - Ann Arbor, saharzz@umich.edu

MISSING AT RANDOM AND IGNORABILITY FOR INFERENCES ABOUT INDIVIDUAL PARAMETERS WITH MISSING DATA

Roderick J. Little and Sahar Zanganeh

Abstract

In a landmark paper, Rubin (1976 Biometrika) showed that the missing data mechanism can be ignored for likelihood-based inference about parameters when (a) the missing data are missing at random (MAR), in the sense that missingness does not depend on the missing values after conditioning on the observed data, and (b) distinctness of the parameters of the data model and the missing-data mechanism, that is, there are no a priori ties, via parameter space restrictions or prior distributions, between the parameters of the data model and the parameters of the model for the mechanism. Rubin (1976) described (a) and (b) as the “weakest simple and general conditions under which it is always appropriate to ignore the process that causes missing data”. However, it is important to note that these conditions are not necessary for ignoring the mechanism in all situations. We propose conditions for ignoring the missing-data mechanism for 2 likelihood inferences about subsets of the parameters of the data model. We present examples where the missing data are ignorable for some parameters, but the missing data mechanism is missing not at random (MNAR), thus extending the range of circumstances where the missing data mechanism can be ignored.

Draft November16, 2012

MISSING AT RANDOM AND IGNORABILITY FOR INFERENCES ABOUT INDIVIDUAL PARAMETERS WITH MISSING DATA

Roderick J.A. Little and Sahar Z Zangeneh

Departments of Biostatistics and Statistics, University of Michigan, Ann Arbor, MI,

ABSTRACT

In a landmark paper, Rubin (1976 Biometrika) showed that the missing data mechanism can be ignored for likelihood-based inference about parameters when (a) the missing data are missing at random (MAR), in the sense that missingness does not depend on the missing values after conditioning on the observed data, and (b) distinctness of the parameters of the data model and the missing-data mechanism, that is, there are no a priori ties, via parameter space restrictions or prior distributions, between the parameters of the data model and the parameters of the model for the mechanism. Rubin (1976) described (a) and (b) as the "weakest simple and general conditions under which it is always appropriate to ignore the process that causes missing data". However, it is important to note that these conditions are not necessary for ignoring the mechanism in all situations. We propose conditions for ignoring the missing-data mechanism for likelihood inferences about subsets of the parameters of the data model. We present examples where the missing data are ignorable for some parameters, but the missing data mechanism is missing not at random (MNAR), thus extending the range of circumstances where the missing data mechanism can be ignored.

Key words: Incomplete data, Missing data, missing at random, likelihood inference, Bayes inference.

1. Introduction

We consider the analysis of data with missing values. In a landmark paper, Rubin (1976) gave sufficient conditions under which the missing data mechanism can be ignored for frequentist and likelihood-based inference about parameters. For likelihood-based inference, these conditions are that (a) the missing data are missing at random (MAR), in the sense (formalized below) that missingness does not depend on the missing values after conditioning on the observed data, and (b) distinctness of the parameters of the data model and the missing-data mechanism, that is, there are no a priori ties, via parameter space restrictions or prior distributions, between the parameters of the data model and the parameters of the model for the mechanism.

Since that paper, ignorability of the mechanism has often been defined by these two conditions (see example Little and Rubin, 2002). However, Rubin (1976) described them as the "weakest simple and general conditions under which it is always appropriate to ignore the process that causes missing data", and it is important to note that these conditions are not necessary for ignoring the mechanism in all situations. In particular, we describe below examples where the mechanism can be ignored when interest concerns all, or a subset, of the model parameters.

In this article, we formalize this idea by proposing conditions for ignoring the missing-data mechanism for likelihood inferences about subsets of the parameters of the data model, consistent with those originally defined by Rubin (1976) for all the parameters. We then present examples where the missing data are ignorable for some parameters, but the missing data mechanism is missing not at random (MNAR), thus extending the range of circumstances where the missing data mechanism can be ignored. Our examples include regression with missing covariates, and inference for survey data with missing observation and external post-stratification information.

2. Inference from Data with Missing Values, Ignoring the Missing Data Mechanism

Let *D* denote the set of complete data if there were no missing values, and let *M* denote a set of binary variables indicating whether individual components of *D* are observed (0) or missing (1). We initially model the density of the joint distribution of *D* and *M* using the "selection model" factorization (Little and Rubin, 2002):

$$f_{D,M}(D,M \mid \theta,\phi) = f_D(D \mid \theta)\, f_M(M \mid D,\phi), \qquad (1)$$

where θ is the parameter of the data model, and ϕ is the parameter of the model for the missing data mechanism. Let $D = (D_{\text{obs}}, D_{\text{mis}})$, where D_{obs} is the observed part of *D* and D_{mis} is the missing part of *D*. Then the full likelihood based on the observed data is

$$L(\theta,\phi \mid D_{\text{obs}}, M) = \text{const.} \times \int f_D(D \mid \theta)\, f_M(M \mid D,\phi)\, dD_{\text{mis}}, \qquad (2)$$

treated as a function of the parameters (θ, ϕ). The likelihood of θ *ignoring the missing-data mechanism* is

$$L(\theta \mid D_{\text{obs}}) = \text{const.} \times \int f_D(D \mid \theta) dD_{\text{mis}}, \qquad (3)$$

which does not involve the model for *M*. Rubin's (1976) sufficient conditions for likelihood inference about θ ignoring the missing data mechanism are that (a) the missing data are missing at random (MAR), defined as

$$f_{M|D}(M \mid D_{\text{obs}}, D_{\text{mis}}, \phi) = f_{M|D}(M \mid D_{\text{obs}}, \phi) \text{ for all } D_{\text{mis}}, \phi, \qquad (4)$$

and (b) θ and ϕ are distinct, that is, their joint parameter space is the product of the parameter space of θ and ϕ. For Bayesian inference (Little and Rubin, 2002), distinctness involves the additional assumption that θ and ϕ are a priori independent, that is the prior distribution of θ and ϕ has the form

$$\pi(\theta, \phi) = \pi_1(\theta) \times \pi_2(\phi).$$

If the data are MAR as in (4), it is easy to see that the full likelihood (2) factorizes as

$$L(\theta, \phi \mid D_{\text{obs}}, M) = \text{const.} \times f_D(D_{\text{obs}} \mid \theta) \times f_{M|D}(M \mid D, \phi), \qquad (5)$$

and hence, likelihood inference based on (3) is valid, and fully efficient if θ and ϕ are distinct.

We now propose definitions of missing at random and ignorability for likelihood inference for a subset θ_1 of the parameters θ.

Definition 1: Write $\theta = (\theta_1, \theta_2)$, where θ_1 and θ_2 are subsets of parameters. The missing data mechanism is MAR for inference about θ_1, denoted $\text{MAR}(\theta_1)$, if the likelihood (1) can be factorized as

$$L(\theta_1, \theta_2, \phi \mid D_{\text{obs}}, M) = \text{const.} \times L_1(\theta_1 \mid D_{\text{obs}}) \times L(\theta_2, \phi \mid D_{\text{obs}}, M) \text{ for all } \theta_1, \theta_2, \phi . \quad (6)$$

Definition 2. The missing data mechanism is ignorable for likelihood inference about θ_1, denoted $\text{LIGN}(\theta_1)$, if (a) the missing data mechanism is $\text{MAR}(\theta_1)$, and (b) θ_1 and (θ_2, ϕ) are distinct sets of parameters. Under (a) likelihood inference about θ_1 can be based on $L_1(\theta_1 \mid D_{\text{obs}})$, which does not involve the model for the mechanism *M*, and under (a) and (b), inference based on $L_1(\theta_1 \mid D_{\text{obs}})$ is fully efficient. For Bayesian inference, the mechanism can be ignored if, in addition, θ_1 and (θ_2, ϕ) are a-priori independent. The posterior distribution of θ_1 is then

$$p(\theta_1 \mid D, M) = \text{const} \times \pi_1(\theta_1) \times L(\theta_1 \mid D_{\text{obs}}), \quad (7)$$

where $\pi_1(\theta_1)$ is the prior distribution of θ_1. Note that (7) does not involve the model for the mechanism. If the mechanism is $\text{MAR}(\theta_1)$ but θ_1 and (θ_2, ϕ) are not distinct sets of parameters, likelihood inference based on $L_1(\theta_1 \mid D_{\text{obs}})$ is valid but not fully efficient, and might still be entertained to avoid the additional assumptions involved in modeling the mechanism.

When $\theta_1 = \theta$, these definitions deviate slightly from Rubin's (1976) original conditions. The distinctness condition reduces to distinctness between θ and ϕ, as defined by Rubin; the $\text{MAR}(\theta_1)$ condition (6) is however not equivalent to Rubin's

(1976) definition of MAR, Eq. (4), but it does lead to Eq. (5), which is the key condition for valid inferences about θ based on the likelihood Eq. (3), which ignores the mechanism.

3. Examples

We now illustrate the definitions in Section 2 with some examples, some simple and others more complex. We start with an example of ignorability as originally discussed by Rubin, for a simple missing data pattern.

Example 1. Monotone Bivariate Data

Let $D = \{(y_{i1}, y_{i2}), i = 1,...,n\}$ denote an independent sample from two variables Y_1, Y_2 which have probability density $f(y_{i1}, y_{i2} \mid \theta)$ indexed by unknown parameters θ. Suppose $D_{\text{obs}} = \{(y_{i1}, y_{i2}), i = 1,...,r\}$ and $\{y_{i1}, i = r+1,...,n\}$, so that Y_1 is fully observed and Y_2 has missing values (Figure 1A). Let $M = \{m_i\}, i = 1,...,n$ where $m_i = 1$ if y_{i2} is missing and $m_i = 0$ if y_{i2} is observed. The missing data mechanism is assumed to depend only on Y_1, and is modeled as:

$$\Pr(m_i = 1 \mid y_{i1}, y_{i2}, \phi) = g(y_{i1}, \phi), \tag{7}$$

where g is a known function with support between 0 and 1. This mechanism meets Rubin's (1976) definition of MAR; likelihood inferences for θ are then ignorable if the parameters θ and ϕ are distinct, and Bayesian inferences are ignorable if θ and ϕ are a priori independent.

Example 2. Ignorability for parameters of distributions of fully-observed variables. It is reasonable to expect that a nonignorable mechanism might often be MAR or ignorable for the parameters of distributions of variables that are not missing; our definition allows for this possibility. In Example 1, suppose we assume the more general MNAR mechanism where missingness of Y_2 depends on both Y_1 and Y_2:

$$\Pr(m_i = 1 \mid y_{i1}, y_{i2}, \phi) = g(y_{i1}, y_{i2}, \phi), \tag{8}$$

where as before g is a known function with support between 0 and 1.

Suppose we factorize the joint distribution of Y_1 and Y_2 as

$$f(y_{i1}, y_{i2} \mid \theta) = f_1(y_{i1} \mid \theta_1) \times f_2(y_{i2} \mid y_{i1}, \theta_2). \tag{9}$$

The likelihood then factors as

$$L(\theta_1, \theta_2, \phi \mid D_{\text{obs}}, M) = \text{const.} \times L_1(\theta_1 \mid D_{\text{obs}}) \times L_2(\theta_2, \phi \mid D_{\text{obs}}, M) \text{ for all } \theta_1, \theta_2, \phi, \text{ where}$$

$$L_1(\theta_1 \mid D_{\text{obs}}) = \prod_{i=1}^{n} f_1(y_{i1}, \theta_1)$$

$$L_2(\theta_2, \phi \mid D_{\text{obs}}, M) = \prod_{i=1}^{r} f_2(y_{i2} \mid y_{i1}, \theta_2)\big((1 - g(y_{i1}, y_{i2}, \phi)\big) \times \prod_{i=r+1}^{n} \int f_2(y_{i2} \mid y_{i1}, \theta_2) g(y_{i1}, y_{i2}, \phi) dy_{i2}.$$

Hence the mechanism is MNAR and not $\text{MAR}(\theta_2)$, but it is $\text{MAR}(\theta_1)$, and $\text{LIGN}(\theta_1)$ if θ_1 and (θ_2, ϕ) are distinct sets of parameters.

Example 3. Complete-case analysis where the mechanism can be ignored when missingness depends on covariates. In Example 1, suppose Y_1 as well as Y_2 is missing when $m_i = 0$, so the observed data consist only of the complete cases, $D_{\text{obs}} = \{(y_{i1}, y_{i2}), i = 1,...,r\}$ (Figure 1B). The models for *Y* and *M* are as in Example 1, and suppose we factorize the joint distribution of Y_1 and Y_2 as in Eq. (9). Then the mechanism is MNAR, since missingness depends on values of Y_1, which are missing for the incomplete cases. The likelihood is

$$L(\theta_1,\theta_2,\phi \mid D_{\text{obs}}, M) = \text{const.} \times L_1(\theta_2 \mid D_{\text{obs}}) \times L_2(\theta_1,\phi \mid D_{\text{obs}}, M) \text{ for all } \theta_1,\theta_2,\phi, \text{ where}$$

$$L_1(\theta_2 \mid D_{\text{obs}}) = \prod_{i=1}^{r} f_2(y_{i2} \mid y_{i1},\theta_2)$$

$$L_2(\theta_1,\phi \mid D_{\text{obs}}, M) = \prod_{i=1}^{r} f_1(y_{i1} \mid \theta_1)\big((1-g(y_{i1},\phi)\big) \times \prod_{i=r+1}^{n} \int f_1(y_{i1} \mid \theta_1) g(y_{i1},\phi) dy_{i1}$$

So this MNAR mechanism is not $\text{MAR}(\theta_1)$, but it is $\text{MAR}(\theta_2)$, and $\text{LIG}(\theta_2)$ if θ_2 and (θ_1,ϕ) are distinct sets of parameters. This example represents a simple case of regression based on the complete cases yielding valid inferences when missingness depends on the covariates. For further elaborations of this idea, see Zhang and Little (2011).

Example 4. A sample with auxiliary data where the mechanism is MNAR but ignorable. Suppose the data in Example 3, we also have the sample of values of the Y_1 recorded for all units in the sample, but these data are external to, and hence not linked with, the respondent sample (Figure 1c). The observed data are then $D_{\text{obs}} = (D_{\text{resp}}, D_{\text{aux}})$, where $D_{\text{resp}} = \{(y_{i1}, y_{i2}), i = 1,...,r\}$ and $D_{\text{aux}} = \{y^*_{j1}, j = 1,...,n\}$. The latter set includes the respondent values of Y_1, but we do not know which they are. Data of this form arise in sample surveys, where the external data are available for the whole sample or often the entire population from a census. The mechanism is technically MNAR, since the mechanism depends on Y_1, but we do not know the values of Y_1 for individual nonrespondents. However, intuitively the marginal distribution of Y_1 can be estimated from D_{aux}, and the conditional distribution of Y_2 given Y_1 can be estimated from D_{resp}. In fact, the mechanism is $\text{MAR}(\theta)$, and $\text{LIG}(\theta)$ if θ and ϕ are distinct, so this is an example where Rubin's conditions are not necessary. To verify this fact, let $\mathbb{S}$ denote the set of permutations of the external data $\pi(1,...,n) = (\pi(1),...,\pi(n))$ that map D_{resp} into the set of respondent values of Y_1, in the sense that $y^*_{\pi(i)1} = y_{i1}, i = 1,...,r$. Let $\|\mathbb{S}\|$ be the size of this set and let $\theta_2^{(mi)}$ denote the values of parameter θ_2 for m_i =0,1. The observed likelihood is then

$$
\begin{aligned}
L(\theta_1,\theta_2,\phi \mid D_{\text{obs}}, M) &= \text{const.} \times \prod_{i=1}^{r} f_1(y_{i1} \mid \theta_1) f_2(y_{i2} \mid y_{i1}, \theta_2)\big((1-g(y_{i1},\phi)\big) \\
&\quad \times \sum_{\pi\in\mathbb{S}} \prod_{i=r+1}^{n} f_1(y^*_{\pi(i)1}) g(y_{\pi(i)1},\phi) / \|\mathbb{S}\| \\
&= \prod_{j=1}^{n} f_1(y^*_{j1} \mid \theta_1) \times \prod_{i=1}^{r} f_2(y_{i2} \mid y_{i1}, \theta_2) \times \prod_{i=1}^{r} \big((1-g(y_{i1},\phi)\big) \times \prod_{j=r+1}^{n} g(y_{j1},\phi),
\end{aligned}
$$

since each of the $\|\mathbb{S}\|$ permutations has the same probability, and the aggregate of the product from $r+1$ to n is the same for each permutation. According to Eq. (7), $\theta_2^{(1)} = \theta_2^{(0)} = \theta_2$, and so the mechanism is $\text{MAR}(\theta)$ and $\text{LIG}(\theta)$ if θ and ϕ are distinct.

Example 5. A nonignorable pattern-mixture model for which some parameters are MAR. So far we have considered the selection model factorization of the joint distribution of D and M, as in Eq. (1). This example models the joint distribution of D and M using the pattern-mixture factorization:

$$f(D,M \mid \gamma,\psi) = f_{D|M}(D \mid M,\gamma) f_M(M \mid \pi) \tag{10}$$

where γ is the parameter of the data model stratified by pattern *M*, and π is the parameter of the marginal model for the missing data patterns. As in Example 2, let $D = \{(y_{i1}, y_{i2}), i = 1,...,n\}$, an independent sample from two variables Y_1, Y_2. Also, let $D_{\text{obs}} = \{(y_{i1}, y_{i2}), i = 1,...,r\}$ and $\{y_{i1}, i = r+1,...,n\}$, so that Y_1 is fully observed and Y_2 has missing values (Figure 1A). We assume the normal pattern-mixture model

$$(y_{i1}, y_{i2} \mid m_i = j, \gamma) \sim_{\text{ind}} G(\mu^{(j)}, \Sigma^{(j)}),\ j = 0,1$$
$$m_i \sim_{\text{ind}} \text{Bern}(\pi)$$

where $\gamma = \{(\mu^{(j)}, \Sigma^{(j)}), j = 0,1\}$, $G(\mu^{(j)}, \Sigma^{(j)})$ denotes the bivariate normal (Gaussian) distribution with mean vector $\mu^{(j)}$ and covariance matrix $\Sigma^{(j)}$, and $\text{Bern}(\pi)$ denotes the Bernouilli distribution with $\Pr(m_i = 1) = \pi$. We assume that missingness of Y_2 depends on Y_2, and an MNAR mechanism which implies that the normal distribution of Y_1 given Y_2 and *M* is the same for complete and incomplete cases. Let $\beta_{12.0}$, $\beta_{12.1}$ and $\sigma_{11.2}$ denote the intercept, slope and residual variance of this distribution, and let $\mu_k^{(j)}, \sigma_{kk}^{(j)}$ denote the mean and variance of Y_k given $m_i = j$. Parameterizing the distribution of the observed data by $\theta_{1.2} = (\beta_{12.0}, \beta_{12.1}, \sigma_{11.2})^t$, $\psi = \left(\mu_2^{(0)}, \sigma_{22}^{(0)}, \mu_1^{(1)}, \sigma_{11}^{(1)}\right)^t$ and π, the likelihood of the observed data is

$$L(\psi, \pi \mid D_{\text{obs}}, M) = \text{const.} \times L_1(\theta_{1\cdot 2} \mid D_{\text{obs}}) \times L_2(\pi, \psi \mid D_{\text{obs}}, M),\ \text{where}$$
$$L_1(\theta_1 \mid D_{\text{obs}}) = \prod_{i=1}^{r} f_{1\cdot 2}(y_{i1} \mid y_{i2}, \theta_{1\cdot 2}),$$
$$L_2(\pi, \psi \mid D_{\text{obs}}, M) = (1-\pi)^r \pi^{n-r} \prod_{i=1}^{r} f_2(y_{i2} \mid m_i = 0, \mu_2^{(0)}, \sigma_{22}^{(0)}) \times \prod_{i=r+1}^{n} f_1(y_{i1} \mid m_i = 1, \mu_1^{(1)}, \sigma_{11}^{(1)})$$

Hence the missing-data mechanism is MNAR, but it is $\text{MAR}(\theta_{1\cdot 2})$. However the mechanism is not $\text{LIG}(\theta_{1\cdot 2})$, since $\theta_{1\cdot 2}$ and (ψ, π) are not distinct sets of parameters. Specifically, $\sigma_{11\cdot 2} \le \sigma_{11}^{(1)}$ since the residual variance of Y_1 in pattern $M=1$ cannot be larger than the marginal variance. The complete-case estimate of $\theta_{1\cdot 2}$ based on $L_1(\theta_1 \mid D_{\text{obs}})$ is in fact maximum likelihood provided that $\hat{\sigma}_{11\cdot 2} \le \hat{\sigma}_{11}^{(1)}$, where $\hat{\sigma}_{11\cdot 2}$ is the sample residual variance of the regression of Y_1 on Y_2 from the complete cases, and $\hat{\sigma}_{11}^{(1)}$ is the sample variance of Y_1 from the incomplete cases.

4. CONCLUSION

We have proposed definitions of MAR and ignorability for subsets of model parameters. This is useful since in many problems the primary focus is on a particular parameter or subset of parameters, and weaker conditions suffice for a subset. Example 2 is a rather obvious example, but the other examples are more substantive. Our definitions differ slightly from Rubin (1976) when applied to all the model parameters, in that cases like Example 4 can be formulated where the mechanism is MAR and ignorable for all the parameters, but the mechanism is not MAR according to Rubin's definition. The case of auxiliary information in Example 4 is important in survey settings, where auxiliary data is available from external data sources; in the future we plan to extend this example to situations with item nonresponse, and more extensive auxiliary information.

We have focused on simple examples to bring out the key ideas, but it is clearly of interest to apply our definitions to more general situations, like a general pattern of missing data, or the block conditional models described by Zhou, Kalbfleisch, and Little

(2010).

REFERENCES

Little, R. J. A., and Rubin, D. B. (2002). *Statistical Analysis with Missing Data (2nd ed)*. New Jersey: John Wiley.

Little, R. J. and Zhang, N. (2011). Subsample ignorable likelihood for regression analysis with missing data. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 60, 4, 591–605. DOI: 10.1111/j.1467-9876.2011.00763.x

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

Zhou, Y., Kalbfleisch, J.D. & Little, R.J. (2010). Block-Conditional MAR Models for Missing Data. *Statistical Science*, 25, 4, 517-532. DOI: 10.1214/10-STS344.

Figure 1. Patterns of Missing Data in the Examples

Figure 1A Figure 1B Figure 1C