Parametric fractional imputation for missing data analysis

Jae Kwang Kim *   Wayne Fuller *

Abstract

Under a parametric model for missing data, the EM algorithm is a popular tool for finding the maximum likelihood estimates (MLE) of the parameters of the model. Imputation, when carefully done, can be used to facilitate the parameter estimation by applying the complete-sample estimators to the imputed dataset. The basic idea is to generate the imputed values from the conditional distribution of the missing data given the observed data. Multiple imputation is a Bayesian approach to generate the imputed values from the conditional distribution.

In this article, parametric fractional imputation is proposed as a parametric approach for generating imputed values. Using fractional weights, the E-step of the EM algorithm can be approximated by the weighted mean of the imputed data likelihood where the fractional weights are computed from the current value of the parameter estimates. Some computational efficiency can be achieved using the idea of importance sampling in the Monte Carlo approximation of the conditional expectation. The resulting estimator of the specified parameters will be identical to the MLE under missing data if the fractional weights are adjusted using a calibration step.

The proposed imputation method provides efficient parameter estimates for the model parameters specified and also provides reasonable estimates for parameters that are not part of the imputation model, for example domain means. Thus, the proposed imputation method is a useful tool for general-purpose data analysis. Variance estimation is covered and results from a limited simulation study are presented.

Key Words: EM algorithm, Importance sampling, Monte Carlo EM, Multiple imputation, Observed likelihood, Observed information.

1. INTRODUCTION

Suppose that $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n$ are independent observations of a p-dimensional random variable $\mathbf{y}$ from a parametric distribution with density $f(\mathbf{y}; \boldsymbol{\theta}_0)$ with $\boldsymbol{\theta}_0 \in \Omega$. The MLE of $\boldsymbol{\theta}_0$ can be obtained as a solution to the following score equation:

$$\mathbf{S}_n(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} \mathbf{s}_i(\boldsymbol{\theta}) = \mathbf{0}, \tag{1}$$

where $\mathbf{s}_i(\boldsymbol{\theta}) = \partial \ln f(\mathbf{y}_i; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ and $\mathbf{S}_n(\boldsymbol{\theta})$ is the score function.

Given missing data, let $(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis})$ denote the observed part and missing part of $\mathbf{y}_i$, respectively. To simplify the presentation, we assume the response mechanism is Missing-At-Random (MAR) in the sense of Rubin (1976). Under MAR, the likelihood function is a marginal likelihood obtained by integrating out over the missing part. Thus, we can write the observed likelihood as

$$L_{obs}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{obs(i)}(\mathbf{y}_{obs,i}; \boldsymbol{\theta}), \tag{2}$$

where $f_{obs(i)}(\mathbf{y}_{obs,i}; \boldsymbol{\theta}) = \int f(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}; \boldsymbol{\theta})\, d\mathbf{y}_{i,mis}$ is the marginal density of $\mathbf{y}_{i,obs}$ and the subscript i is used in $f_{obs(i)}(\cdot)$ because the missing pattern can differ from observation to observation.

To compute the MLE that maximizes the observed likelihood (2), we need to solve the observed score equation for $\boldsymbol{\theta}$, where the observed score equation is

$$\mathbf{S}_{obs}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} \mathbf{s}_{i,obs}(\boldsymbol{\theta}) \equiv \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^{n} \ln\left\{ f_{obs(i)}(\mathbf{y}_{obs,i}; \boldsymbol{\theta}) \right\} = \mathbf{0}. \tag{3}$$

Instead of solving (3), the MLE of $\boldsymbol{\theta}_0$ can be obtained by solving

$$\bar{\mathbf{S}}(\boldsymbol{\theta}) \equiv E\{\mathbf{S}_n(\boldsymbol{\theta}) \mid \mathbf{Y}_{obs}\} \equiv \sum_{i=1}^{n} E\{\mathbf{s}_i(\boldsymbol{\theta}) \mid \mathbf{y}_{i,obs}\} = \mathbf{0}, \tag{4}$$

where $\mathbf{Y}_{obs} = (\mathbf{y}_{1,obs}, \mathbf{y}_{2,obs}, \cdots, \mathbf{y}_{n,obs})$, and $\bar{\mathbf{S}}(\boldsymbol{\theta})$ is called the mean score function. The equivalence of the observed score function and the mean score function was first proved by Fisher (1925).

*Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

Strictly speaking, the conditional expectation in (4) is evaluated at $\boldsymbol{\theta}$ and we should write the mean score equation as

$$\bar{\mathbf{S}}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} E\left\{\mathbf{s}_i(\boldsymbol{\theta}) \mid \mathbf{y}_{i,obs}, \boldsymbol{\theta}\right\} = \mathbf{0}. \tag{5}$$

The EM algorithm, proposed by Dempster el al (1977), computes the solution iteratively by defining $\hat{\boldsymbol{\theta}}_{(t+1)}$ to be the solution to

$$\sum_{i=1}^{n} E\left\{\mathbf{s}_i(\boldsymbol{\theta}) \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right\} = \mathbf{0}, \tag{6}$$

where $\hat{\boldsymbol{\theta}}_{(t)}$ is the estimate of $\boldsymbol{\theta}$ obtained at the t-th iteration. To compute the conditional expectation in (6), the Monte Carlo implementation of the EM (MCEM) algorithm of Wei and Tanner (1990) can be used. The MCEM method avoids the analytic computation of the conditional expectation (6) by using the Monte Carlo approximation based on the imputed data. Thus, one can interpret imputation as a Monte Carlo approximation of the conditional expectation given the observed data. The Monte Carlo methods of approximating the conditional expectation in (4) can be placed in two classes:

1. Bayesian approach: Generate the imputed values from the posterior predictive distribution of $\mathbf{y}_{i,mis}$ given $\mathbf{y}_{i,obs}$:

$$f(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}) = \int f(\mathbf{y}_{i,mis} \mid \boldsymbol{\theta}, \mathbf{y}_{i,obs}) f(\boldsymbol{\theta} \mid \mathbf{y}_{i,obs}) d\boldsymbol{\theta}. \tag{7}$$

   This is essentially the approach used in multiple imputation as proposed by Rubin (1987).

2. Frequentist approach: Generate the imputed values from the conditional distribution $f\left(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}\right)$ with an estimated value $\hat{\boldsymbol{\theta}}$.

The Bayesian approach to imputation has been proposed as a general method of handling missing data because of the feasibility of Bayesian computational methods and the simplicity of variance estimation. However, the convergence to a stable posterior predictive distribution (7) is difficult to check and often requires huge computation (Gelman et al, 1996). Also, the variance estimator used in multiple imputation is not always consistent. For examples, see Fay (1992), Wang and Robins (1998), and Kim et al (2006).

In the frequentist approach to imputation, the imputed values are generated from the conditional distribution $f\left(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}\right)$ with a particular value $\hat{\boldsymbol{\theta}}$, often the MLE of $\boldsymbol{\theta}$. However, the frequentist approach for imputation has received less attention than Bayesian imputation. One notable exception is Wang and Robins (1998) who studied the asymptotic properties of multiple imputation and a (parametric) frequentist imputation procedure. Wang and Robins (1998) considered the estimated parameter $\hat{\boldsymbol{\theta}}$ to be given, and did not discuss parameter estimation.

We consider a frequentist imputation given a parametric model for the original distribution. We propose an alternative implementation of the MCEM method using parametric fractional imputation that does not require re-generation of the imputed values at each iteration. Only the fractional weights are re-computed for each iteration and we propose a simple method of computing the fractional weights without increasing the size of Monte Carlo samples. The proposed method uses the calibration technique to obtain the MLE and is computationally very attractive in many cases.

In Section 2, the parametric fractional imputation method is proposed. Variance estimation is discussed in Section 3 and the proposed method is extended for general purpose estimation in Section 4. Calibration fractional imputation is derived in Section 5. Results from a limited simulation study are presented in Section 6.

2. Proposed method

As discussed in Section 1, solving the mean score equation (4) requires an iterative method because the conditional distribution of $\mathbf{y}_{i,mis}$ given $\mathbf{y}_{i,obs}$, denoted by $f(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}, \boldsymbol{\theta})$, is a function of $\boldsymbol{\theta}$. Thus, since we cannot generate imputed values from the conditional distribution with unknown $\boldsymbol{\theta}$, the iterative procedure generates imputed values from the conditional distribution with the current value of $\boldsymbol{\theta}$ and then updates $\boldsymbol{\theta}$ based on the imputed score equation.

To avoid re-generating values from the conditional distribution at each step, we first generate M imputed values from some known distribution $q(\mathbf{y}_{i,mis})$ whose support includes that of $f(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}, \boldsymbol{\theta})$. Let the generated values be $\mathbf{y}_{i,mis}^{*(1)}, \cdots, \mathbf{y}_{i,mis}^{*(M)}$. Because

$$E\left\{\mathbf{s}_i(\boldsymbol{\theta}) \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right\} = \int \mathbf{s}_i(\boldsymbol{\theta}) \frac{f\left(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right)}{q(\mathbf{y}_{i,mis})} q(\mathbf{y}_{i,mis}) d\mathbf{y}_{i,mis}, \tag{8}$$

we can approximate the conditional expectation by

$$E\left\{\mathbf{s}_i(\boldsymbol{\theta}) \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right\} \doteq \frac{1}{M}\sum_{j=1}^{M} \mathbf{s}_i^{*(j)}(\boldsymbol{\theta}) \frac{f\left(\mathbf{y}_{i,mis}^{*(j)} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right)}{q\left(\mathbf{y}_{i,mis}^{*(j)}\right)}.$$

Thus, we propose the following algorithm for the parametric fractional imputation using importance sampling:

[Step 1] Obtain an initial estimator $\hat{\boldsymbol{\theta}}_{(0)}$ of $\boldsymbol{\theta}$. Also, generate M imputed values, $\mathbf{y}_{i,mis}^{*(1)}, \cdots, \mathbf{y}_{i,mis}^{*(M)}$, from some density $q(\mathbf{y}_{i,mis})$. Often, $q(\mathbf{y}_{i,mis}) = f\left(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(0)}\right)$.

[Step 2] With the current estimate of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}_{(t)}$, compute the fractional weights as

$$w_{ij(t)}^{*} = C_{i(t)} \frac{f\left(\mathbf{y}_{i,mis}^{*(j)} \mid \mathbf{y}_{i,obs}; \hat{\boldsymbol{\theta}}_{(t)}\right)}{q\left(\mathbf{y}_{i,mis}^{*(j)}\right)}, \quad (9)$$

where $C_{i(t)}$ is chosen to satisfy $\sum_{j=1}^{M} w_{ij(t)}^{*} = 1$.

[Step 3] Using the fractional weight obtained from Step 2, solve the weighted score equation

$$\hat{\boldsymbol{\theta}}_{(t+1)} \leftarrow \text{ solution to } \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij(t)}^{*} \mathbf{s}_i^{*(j)}(\boldsymbol{\theta}) = \mathbf{0}. \quad (10)$$

[Step 4] Go to Step 2. Stop if $\hat{\boldsymbol{\theta}}_{(t)}$ meets the convergence criterion.

The proposed method is computationally attractive because we use a weighted score equation to compute the parameter estimates. Unlike the MCEM method, the imputed values are not changed for each iteration, only the fractional weights are changed.

Remark 1 *In Step 2, fractional weights can be computed by using the joint density with the current parameter estimate* $\hat{\boldsymbol{\theta}}_{(t)}$*. Note that*

$$\frac{f\left(\mathbf{y}_{i,mis}^{*(j)} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right) / q\left(\mathbf{y}_{i,mis}^{*(j)}\right)}{\sum_{j=1}^{M} f\left(\mathbf{y}_{i,mis}^{*(j)} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right) / q\left(\mathbf{y}_{i,mis}^{*(j)}\right)} = \frac{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t)}\right) / q\left(\mathbf{y}_{i,mis}^{*(j)}\right)}{\sum_{j=1}^{M} f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t)}\right) / q\left(\mathbf{y}_{i,mis}^{*(k)}\right)}.$$

Thus, the fractional weights (9) can be computed as

$$w_{ij(t)}^{*} = C_{i(t)} \frac{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t)}\right)}{q\left(\mathbf{y}_{i,mis}^{*(j)}\right)},$$

which does not require the density of the conditional distribution. Only the joint density is needed.

Remark 2 *The choice of the initial density* $q(\mathbf{y}_{i,mis})$ *is somewhat arbitrary. If we choose* $q(\mathbf{y}_{i,mis}) = f\left(\mathbf{y}_{i,mis} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(0)}\right)$ *where* $\hat{\boldsymbol{\theta}}_{(0)}$ *is an initial parameter estimate of* $\boldsymbol{\theta}$*, the fractional weight with current parameter estimate* $\hat{\boldsymbol{\theta}}_{(t)}$ *is of the form*

$$w_{ij(t)}^{*} = C_{i(t)} \frac{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t)}\right)}{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(0)}\right)}, \quad (11)$$

where $C_{i(t)}$ *is a normalizing constant. The initial estimate* $\hat{\boldsymbol{\theta}}_{(0)}$ *is not necessarily* $\sqrt{n}$*-consistent.*

Given the M imputed values, $\mathbf{y}_{i,mis}^{*(1)}, \cdots, \mathbf{y}_{i,mis}^{*(M)}$, generated from $q(\mathbf{y}_{i,mis})$, the sequence of estimators $\left\{\hat{\boldsymbol{\theta}}_{(0)}, \hat{\boldsymbol{\theta}}_{(1)}, \cdots\right\}$ can be constructed from the parametric fractional imputation using importance sampling. The following theorem presents some convergence properties of the sequence of the estimators.

Theorem 1 *Assume that the M imputed values are generated from $q(\mathbf{y}_{i,mis})$. Let*

$$Q^*\left(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_{(t)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij(t)}^* \ln f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \boldsymbol{\theta}\right), \tag{12}$$

where $w_{ij(t)}^* = w_{ij}^*\left(\hat{\boldsymbol{\theta}}_{(t)}\right)$. If

$$Q^*\left(\hat{\boldsymbol{\theta}}_{(t+1)} \mid \hat{\boldsymbol{\theta}}_{(t)}\right) \geq Q^*\left(\hat{\boldsymbol{\theta}}_{(t)} \mid \hat{\boldsymbol{\theta}}_{(t)}\right) \tag{13}$$

then

$$L_{obs}^*\left(\hat{\boldsymbol{\theta}}_{(t+1)}\right) \geq L_{obs}^*\left(\hat{\boldsymbol{\theta}}_{(t)}\right), \tag{14}$$

where $L_{obs}^*(\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{obs(i)}^*(\mathbf{y}_{i,obs}; \boldsymbol{\theta})$ with

$$f_{obs(i)}^*(\mathbf{y}_{i,obs}; \boldsymbol{\theta}) = \frac{\sum_{j=1}^{M} f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \boldsymbol{\theta}\right)/q\left(\mathbf{y}_{i,mis}^{*(j)}\right)}{\sum_{j=1}^{M} 1/q\left(\mathbf{y}_{i,mis}^{*(j)}\right)}.$$

Proof. By the Jensen's inequality,

$$\begin{aligned}
\ln L_{obs}^*\left(\hat{\boldsymbol{\theta}}_{(t+1)}\right) - \ln L_{obs}^*\left(\hat{\boldsymbol{\theta}}_{(t)}\right) &= \sum_{i=1}^{n} \ln \sum_{j=1}^{M} w_{ij(t)}^* \frac{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t+1)}\right)}{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t)}\right)} \\
&\geq \sum_{i=1}^{n}\sum_{j=1}^{M} \ln w_{ij(t)}^* \frac{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t+1)}\right)}{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \hat{\boldsymbol{\theta}}_{(t)}\right)} \\
&= Q^*\left(\hat{\boldsymbol{\theta}}_{(t+1)} \mid \hat{\boldsymbol{\theta}}_{(t)}\right) - Q^*\left(\hat{\boldsymbol{\theta}}_{(t)} \mid \hat{\boldsymbol{\theta}}_{(t)}\right).
\end{aligned}$$

Therefore, (13) implies (14). ∎

Note that $L_{obs}^*(\boldsymbol{\theta})$ is an imputed version of the observed likelihood based on the the M imputed values, $\mathbf{y}_{i,mis}^{*(1)}, \cdots, \mathbf{y}_{i,mis}^{*(M)}$, generated from $q(\mathbf{y}_{i,mis})$. Under fairly general conditions, the solution to the imputed score equation (10) satisfies (13). Thus, by Theorem 1, the sequence $L_{obs}^*\left(\hat{\boldsymbol{\theta}}_{(t)}\right)$ is monotonically increasing. Also, under the fairly general conditions stated in Wu (1983), the convergence of $\hat{\boldsymbol{\theta}}_{(t)}$ follows for fixed M. Theorem 1 does not hold for the sequence obtained from the MCEM method for fixed M.

3. Variance estimation

To discuss variance estimation, note that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \bar{S}(\boldsymbol{\theta}) = -I_{obs}(\boldsymbol{\theta}), \tag{15}$$

where

$$I_{obs}(\boldsymbol{\theta}) = E\left\{-\frac{\partial}{\partial \boldsymbol{\theta}} S_n(\boldsymbol{\theta}) \mid Y_{obs}, \boldsymbol{\theta}\right\} + \bar{S}(\boldsymbol{\theta})^{\otimes 2} - E\left\{S_n(\boldsymbol{\theta})^{\otimes 2} \mid Y_{obs}, \boldsymbol{\theta}\right\} \tag{16}$$

with $S_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{s}_i(\boldsymbol{\theta})$ and $S^{\otimes 2} = SS'$. Louis (1982) first proved (15) to estimate the variance of the MLE obtained by the EM algorithm.

Let $\hat{\boldsymbol{\theta}}^*$ be the solution to the approximate mean score equation

$$\bar{S}^*(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*(\boldsymbol{\theta}) \mathbf{s}_i^{*(j)}(\boldsymbol{\theta}) = \mathbf{0}, \tag{17}$$

where $\mathbf{s}_i^{*(j)}(\boldsymbol{\theta}) = \mathbf{s}\left(\boldsymbol{\theta}; \mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}\right)$ and

$$w_{ij}^*(\boldsymbol{\theta}) = \frac{f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}; \boldsymbol{\theta}\right)/q\left(\mathbf{y}_{i,mis}^{*(j)}\right)}{\sum_{k=1}^{M} f\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(k)}; \boldsymbol{\theta}\right)/q\left(\mathbf{y}_{i,mis}^{*(k)}\right)}. \tag{18}$$

Note that

$$E\left\{\bar{S}^*(\boldsymbol{\theta}) \mid Y_{obs}\right\} = \bar{S}(\boldsymbol{\theta}) \tag{19}$$

where $\bar{S}(\boldsymbol{\theta})$ is defined in (5) and the expectation in (19) is over the imputation mechanism. Here, superscript $*$ is used in $\hat{\boldsymbol{\theta}}^*$ to emphasize that the solution is obtained from the approximate mean score equation (17), not from the exact mean score equation (5). An EM-type algorithm such as (10) can be used to find a solution $\hat{\boldsymbol{\theta}}^*$ to (17). Using the Taylor linearization,

$$\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}_0 \cong -\left[E\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\bar{S}(\boldsymbol{\theta}_0)\right\}\right]^{-1}\bar{S}^*(\boldsymbol{\theta}_0).$$

Thus, we can use the sandwich formula to compute the variance of $\hat{\boldsymbol{\theta}}^*$ that is the solution to (4). Note that, by (19),

$$Var\left\{\bar{S}^*(\boldsymbol{\theta}_0)\right\} = Var\left\{\bar{S}(\boldsymbol{\theta}_0)\right\} + Var\left\{\bar{S}^*(\boldsymbol{\theta}_0) - \bar{S}(\boldsymbol{\theta}_0)\right\}. \tag{20}$$

The first term in the right side of (20) can be estimated by $I_{obs}\left(\hat{\boldsymbol{\theta}}^*\right)^{-1}$, as suggested by Louis (1982). The observed information (16) can be easily computed from fractional imputation. That is, we use $\hat{I}_{obs}(\hat{\boldsymbol{\theta}}^*)$ an an estimator of $I_{obs}(\boldsymbol{\theta}_0)$, where

$$\begin{aligned}\hat{I}_{obs}(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*\left\{-\partial\mathbf{s}_i\left(\boldsymbol{\theta};\mathbf{y}_{i,obs},\mathbf{y}_{i,mis}^{*(j)}\right)/\partial\boldsymbol{\theta}\right\} \\ &+ \sum_{i=1}^{n}\left\{\bar{\mathbf{s}}_i^*(\boldsymbol{\theta})\right\}^{\otimes 2} - \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*\left\{\mathbf{s}_i^{*(j)}(\boldsymbol{\theta})\right\}^{\otimes 2}\end{aligned} \tag{21}$$

where $\bar{\mathbf{s}}_i^*(\boldsymbol{\theta}) = \sum_{j=1}^{M} w_{ij}^*\mathbf{s}_i^{*(j)}(\boldsymbol{\theta})$ and $w_{ij}^* = w_{ij}^*(\hat{\boldsymbol{\theta}})$. Thus, the estimator in (21) is based on the Monte Carlo approximation of the conditional expectation (16) using fractional imputation where the fractional weight corresponds to the importance weight of importance sampling. Because the Monte Carlo expectation not only approximates the mean score equation (5) but also approximates the observed information (16), the fractional imputation (FI) method provides consistent variance estimation for sufficiently large M.

To estimate the second term of (20), we consider the case when $\mathbf{y}_{i,mis}^{*(1)},\cdots,\mathbf{y}_{i,mis}^{*(M)}$ are independent samples from $q(\mathbf{y}_{i,mis})$. In this case, we can express

$$\bar{S}^*(\boldsymbol{\theta}) = \frac{1}{M}\sum_{j=1}^{M}\bar{S}^{*(j)}(\boldsymbol{\theta})$$

where $\bar{S}^{*(j)}(\boldsymbol{\theta}) = M\sum_{i=1}^{n} w_{ij}^*\mathbf{s}_i^{*(j)}(\boldsymbol{\theta})$ and we have

$$\frac{1}{M}B(\boldsymbol{\theta}) = \frac{1}{M}\frac{1}{M-1}\sum_{j=1}^{M}\left\{\bar{S}^{*(j)}(\boldsymbol{\theta}) - \bar{S}^*(\boldsymbol{\theta})\right\}^{\otimes 2}$$

to be unbiased for second term of (20). Therefore, the proposed variance estimator is

$$\hat{V}(\hat{\boldsymbol{\theta}}^*) = \left[I_{obs}\left(\hat{\boldsymbol{\theta}}^*\right)\right]^{-1} + \left[I_{obs}\left(\hat{\boldsymbol{\theta}}^*\right)\right]^{-1}\left\{\frac{1}{M}B(\hat{\boldsymbol{\theta}}^*)\right\}\left[I_{obs}\left(\hat{\boldsymbol{\theta}}^*\right)'\right]^{-1}. \tag{22}$$

Often, the second term in (20) is very small for large M or for an efficient imputation method. In this case, the second term (22) can be safely omitted in the variance estimation.

4. Extensions

So far, we have considered the case where the parameter of interest is estimated by the maximum likelihood method. We consider an extension where the parameter of interest is not necessarily estimated from the maximum likelihood method, but is estimated by solving an estimating equation. Suppose that, under complete response, a parameter of interest, denoted by $\boldsymbol{\eta}$, is estimated as the unique solution to the estimating equation

$$\mathbf{U}(\boldsymbol{\eta}) \equiv \sum_{i=1}^{n}\mathbf{u}(\boldsymbol{\eta};\mathbf{y}_i) = \mathbf{0}, \tag{23}$$

for some function $\mathbf{u}(\boldsymbol{\eta}; \mathbf{y}_i)$ of $\boldsymbol{\eta}$ with continuous partial derivatives. Let $\hat{\boldsymbol{\eta}}$ be the solution to (23). Under some regularity conditions,

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0\right) \sim N\left[\mathbf{0}, \left\{\mathbf{g}\left(\boldsymbol{\eta}_0\right)\right\}^{-1} V\left\{\mathbf{u}\left(\boldsymbol{\eta}_0; \mathbf{y}\right)\right\}\left\{\mathbf{g}\left(\boldsymbol{\eta}_0\right)\right\}^{-1}\right]$$

where $\mathbf{g}(\boldsymbol{\eta}) = E\left\{\partial \mathbf{u}(\boldsymbol{\eta}; \mathbf{y}) / \partial \boldsymbol{\eta}\right\}$ and $\boldsymbol{\eta}_0$ is a unique solution to $E\{U(\boldsymbol{\eta})\} = \mathbf{0}$.

Under nonresponse, a consistent estimator of $\boldsymbol{\eta}_0$ can be obtained as a solution to the following estimating equation

$$\bar{\mathbf{U}}\left(\boldsymbol{\eta} \mid \hat{\boldsymbol{\theta}}\right) \equiv \sum_{i=1}^{n} E\left\{\mathbf{u}\left(\boldsymbol{\eta}; \mathbf{y}_i\right) \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}\right\} = \mathbf{0}, \tag{24}$$

where $\hat{\boldsymbol{\theta}}$ is the solution to (5). The estimating equation (24) is called the expected estimating equation. The use of an expected estimating equation has been discussed by, among others, Wang and Pepe (2000) and Robins and Wang (2000).

Using the fractional imputation approach discussed in Section 2, we can construct a Monte Carlo approximation to the estimating equation

$$\hat{\boldsymbol{\eta}}^* \leftarrow \text{ solution to } \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*(\hat{\boldsymbol{\theta}}^*)\ \mathbf{u}_i^{*(j)}\left(\boldsymbol{\eta}\right) = \mathbf{0}, \tag{25}$$

where $\mathbf{u}_i^{*(j)}\left(\boldsymbol{\eta}\right) = \mathbf{u}(\boldsymbol{\eta}; \mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)})$, $w_{ij}^*\left(\boldsymbol{\theta}\right)$ is defined in (18), and $\hat{\boldsymbol{\theta}}^*$ is the solution to (17). Note that we do not have to update the solution $\hat{\boldsymbol{\theta}}^*$ iteratively in (25) and only the final estimate $\hat{\boldsymbol{\theta}}^*$ is needed.

The following theorem presents some asymptotic properties of the estimator that is the solution to (24), or the solution to (25).

Theorem 2 *Let $\hat{\boldsymbol{\theta}}^*$ be the Monte Carlo approximation of the MLE of $\boldsymbol{\theta}$ that is computed by solving the approximated mean score equation (17). Under some regularity conditions, the solution $\hat{\boldsymbol{\eta}}^*$ to (25) satisfies*

$$\sqrt{n}\left(\hat{\boldsymbol{\eta}}^* - \tilde{\boldsymbol{\eta}}^*\right) = o_p\left(1\right) \tag{26}$$

where

$$E\left(\tilde{\boldsymbol{\eta}}^*\right) = \boldsymbol{\eta}_0$$

and

$$Var\left(\tilde{\boldsymbol{\eta}}^*\right) = \left\{\mathbf{g}\left(\boldsymbol{\eta}_0\right)\right\}^{-1} Var\left\{\tilde{\mathbf{U}}^*\left(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0\right)\right\}\left\{\mathbf{g}\left(\boldsymbol{\eta}_0\right)'\right\}^{-1}. \tag{27}$$

Here, $\mathbf{g}(\boldsymbol{\eta}) = E\left\{\sum_{i=1}^{n} \partial \mathbf{u}(\boldsymbol{\eta}; \mathbf{y}_i) / \partial \boldsymbol{\eta}\right\}$ *and*

$$\tilde{\mathbf{U}}^*\left(\boldsymbol{\eta}, \boldsymbol{\theta}\right) = \bar{\mathbf{U}}^*\left(\boldsymbol{\eta}, \boldsymbol{\theta}\right) + K'\bar{\mathbf{S}}^*\left(\boldsymbol{\theta}\right), \tag{28}$$

where

$$\begin{aligned}
\bar{\mathbf{U}}^*\left(\boldsymbol{\eta}, \boldsymbol{\theta}\right) &= \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*\left(\boldsymbol{\theta}\right)\mathbf{u}_i^{*(j)}\left(\boldsymbol{\eta}\right) \\
\bar{\mathbf{S}}^*\left(\boldsymbol{\theta}\right) &= \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*\left(\boldsymbol{\theta}\right)\mathbf{s}_i^{*(j)}\left(\boldsymbol{\theta}\right)
\end{aligned}$$

and

$$K = \left[\mathcal{I}_{obs}\left(\boldsymbol{\theta}_0\right)\right]^{-1} E\left[\mathbf{S}_{mis}\left(\boldsymbol{\theta}_0\right)\mathbf{U}'\left(\boldsymbol{\eta}_0\right)\right]. \tag{29}$$

Here, $\mathcal{I}_{obs}\left(\boldsymbol{\theta}\right) = -E\left\{\partial S_{obs}\left(\boldsymbol{\theta}\right) / \partial \boldsymbol{\theta}\right\}$ *and* $\mathbf{S}_{mis}\left(\boldsymbol{\theta}\right) = S_n\left(\boldsymbol{\theta}\right) - S_{obs}\left(\boldsymbol{\theta}\right)$.

The result in Theorem 2 can be used to derive a variance estimator for $\hat{\boldsymbol{\eta}}$ that is a solution to (25). The crucial part is to estimate the variance of the linearized term (28). Note that we can write

$$Var\left\{\tilde{\mathbf{U}}^*\left(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0\right)\right\} = Var\left\{\tilde{\mathbf{U}}\left(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0\right)\right\} + Var\left\{\tilde{\mathbf{U}}^*\left(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0\right) - \tilde{\mathbf{U}}\left(\boldsymbol{\eta}_0, \boldsymbol{\theta}_0\right)\right\}, \tag{30}$$

where

$$\tilde{\mathbf{U}}\left(\boldsymbol{\eta}, \boldsymbol{\theta}\right) = p\lim_{M \to \infty} \tilde{\mathbf{U}}^*\left(\boldsymbol{\eta}, \boldsymbol{\theta}\right)$$

If we write

$$\tilde{\mathbf{U}}(\boldsymbol{\eta},\boldsymbol{\theta}) = \bar{U}(\boldsymbol{\eta},\boldsymbol{\theta}) - K'\bar{S}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left\{\bar{\mathbf{u}}_i(\boldsymbol{\eta},\boldsymbol{\theta}) - K'\bar{\mathbf{s}}_i(\boldsymbol{\theta})\right\} = \sum_{i=1}^{n}\tilde{\mathbf{u}}_i,$$

a plug-in estimator of $Var\left\{\tilde{\mathbf{U}}(\boldsymbol{\eta}_0,\boldsymbol{\theta}_0)\right\}$ is

$$\frac{n}{n-1}\sum_{i=1}^{n}\left(\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}}\right)\left(\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}}\right)'$$

where

$$\hat{\mathbf{u}}_i = \bar{\mathbf{u}}_i(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\theta}}) - \hat{K}'\bar{\mathbf{s}}_i\left(\hat{\boldsymbol{\theta}}\right).$$

The terms $\bar{\mathbf{u}}_i(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\theta}})$ and $\bar{\mathbf{s}}_i(\hat{\boldsymbol{\theta}})$ are easily computed from the fractional imputation with fractional weights. To estimate the second term of (30), write

$$\tilde{\mathbf{U}}^*(\boldsymbol{\eta},\boldsymbol{\theta}) = \frac{1}{M}\sum_{j=1}^{M}\tilde{U}^{*(j)}(\boldsymbol{\eta},\boldsymbol{\theta}),$$

where $\tilde{U}^{*(j)}(\boldsymbol{\eta},\boldsymbol{\theta}) = M\sum_{i=1}^{n} w_{ij}^*(\boldsymbol{\theta})\left\{\mathbf{u}_i^{*(j)}(\boldsymbol{\eta}) - K'\mathbf{s}_i^{*(j)}(\boldsymbol{\theta})\right\}$. The second term in (30) can be consistently estimated by

$$\frac{1}{M}\frac{1}{M-1}\sum_{j=1}^{M}\left\{\tilde{U}^{*(j)}\left(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\theta}}\right) - \tilde{\mathbf{U}}^*\left(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\theta}}\right)\right\}^{\otimes 2}.$$

To estimate K term in (29), we need to estimate the two terms in (29) separately. The first term, $\mathcal{I}_{obs}(\boldsymbol{\theta})$, can be computed using (21), the estimated observed information based on the Louis formula. Now, to estimate the second term in K, we use

$$\begin{aligned} & E\left\{U(\boldsymbol{\eta},\boldsymbol{\theta})S_{mis}(\boldsymbol{\theta}) \mid Y_{obs},\hat{\boldsymbol{\theta}}\right\} \\ = \;& E\left\{U(\boldsymbol{\eta},\boldsymbol{\theta})S_n'(\boldsymbol{\theta}) \mid Y_{obs},\hat{\boldsymbol{\theta}}\right\} - \bar{U}(\boldsymbol{\eta},\boldsymbol{\theta})\bar{S}(\boldsymbol{\theta})'. \end{aligned}$$

The first expectation can be estimated by the fractional imputation. That is, we can estimate $E\left\{U(\boldsymbol{\eta},\boldsymbol{\theta})S_n'(\boldsymbol{\theta}) \mid Y_{obs},\hat{\boldsymbol{\theta}}\right\}$ by

$$\sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*\mathbf{u}_i^{*(j)}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\theta}})\mathbf{s}_i^{*(j)}(\hat{\boldsymbol{\theta}})'$$

with $\mathbf{u}_i^{*(j)}(\boldsymbol{\eta},\boldsymbol{\theta}) = \mathbf{u}_i(\boldsymbol{\eta},\boldsymbol{\theta};\mathbf{y}_{i,obs},\mathbf{y}_{i,mis}^{*(j)})$ and $\mathbf{s}_i^{*(j)}(\boldsymbol{\theta}) = \mathbf{s}_i(\boldsymbol{\theta};\mathbf{y}_{i,obs},\mathbf{y}_{i,mis}^{*(j)})$.

5. Calibration

The proposed estimation method can be viewed as a method of implementing a MCEM algorithm using importance sampling. The MCEM method is subject to sampling error when approximating the conditional expectation by a summation. In general, the size M of the Monte Carlo sample needs to be very large for satisfactory approximation. For moderate size M, there are two situations when the approximation is accurate. The first situation is when there are only finite number of possible values for $\mathbf{y}_{i,mis}$. In this case, we take the possible values as the imputed values and compute the conditional probability of $\mathbf{y}_{i,mis}^*$ by the following Bayes formula:

$$p\left(\mathbf{y}_{i,mis}^{*(j)} \mid \mathbf{y}_{i,obs},\hat{\boldsymbol{\theta}}\right) = \frac{f\left(\mathbf{y}_{i,obs},\mathbf{y}_{i,mis}^*;\hat{\boldsymbol{\theta}}\right)}{\sum_{j=1}^{M_i} f\left(\mathbf{y}_{i,obs},\mathbf{y}_{i,mis}^{*(j)};\hat{\boldsymbol{\theta}}\right)},$$

where M_i is the number of possible values of $\mathbf{y}_{i,mis}$ and $\hat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$. The conditional expectation in (6) can be written

$$E\left\{\mathbf{s}_i(\boldsymbol{\theta}) \mid \mathbf{y}_{i,obs},\hat{\boldsymbol{\theta}}_{(t)}\right\} = \sum_{j=1}^{M_i}\mathbf{s}_i^{*(j)}(\boldsymbol{\theta})\,p\left(\mathbf{y}_{i,mis}^{*(j)} \mid \mathbf{y}_{i,obs},\hat{\boldsymbol{\theta}}_{(t)}\right). \qquad (31)$$

Here, the estimated probability $p\left(\mathbf{y}_{i,mis}^{*(j)} \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right)$ takes the role of the fractional weight. Ibrahim (1990) proposed using (31) in the E-step of the EM algorithm for discrete data.

The approximation is exact when the distribution belongs to the exponential family of the form

$$f(\mathbf{y};\boldsymbol{\theta}) = \exp\left\{\mathbf{t}(\mathbf{y})'\boldsymbol{\theta} + \phi(\boldsymbol{\theta}) + A(\mathbf{y})\right\}. \tag{32}$$

Under the model (32), the score equation (1) under complete response is equal to

$$\sum_{i=1}^{n}\left\{\mathbf{t}(\mathbf{y}_i) + \frac{\partial\phi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\} = \mathbf{0}$$

and the mean score equation (4) can be written

$$\sum_{i=1}^{n}\left\{E\left[\mathbf{t}(\mathbf{y}_i) \mid \mathbf{y}_{i,obs}, \boldsymbol{\theta}\right] + \frac{\partial\phi(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\} = \mathbf{0}.$$

Thus, the integration problem in (6) reduces to the problem of computing the integration $E\{\mathbf{t}(\mathbf{y}_i) \mid \mathbf{y}_{i,obs}, \boldsymbol{\theta}\}$, which is often a known function of $\mathbf{y}_{i,obs}$ and $\boldsymbol{\theta}$. In this case, the implementation of the EM algorithm simplifies.

Define

$$\mathbf{g}(\mathbf{y}_{i,obs}, \boldsymbol{\theta}) = E\{\mathbf{t}(\mathbf{y}_i) \mid \mathbf{y}_{i,obs}, \boldsymbol{\theta}\}. \tag{33}$$

Recall that, in the fractional imputation approach, we can express the conditional expectation by a weighted summation

$$E\left\{\mathbf{t}(\mathbf{y}_i) \mid \mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right\} = \sum_{j=1}^{M_i} w_{ij(t)}^{*}\mathbf{t}\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}\right), \tag{34}$$

where $\mathbf{y}_{i,mis}^{*(j)}$ is the j-th imputed value of $\mathbf{y}_{i,mis}$ and $w_{ij(t)}^{*}$ is the fractional weight which is the conditional probability of $\mathbf{y}_{i,mis} = \mathbf{y}_{mis,i}^{(j)}$ given $\mathbf{y}_{obs,i}$ using the current parameter value $\hat{\boldsymbol{\theta}}_{(t)}$. Thus, it is proposed that

$$\sum_{j=1}^{M_i} w_{ij(t)}^{*}\mathbf{t}\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}\right) = \mathbf{g}\left(\mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right) \tag{35}$$

be used as as a constraint for finding the fractional weights. We can use the regression weighting technique or the empirical likelihood technique to find a solution to (35). Here, M_i need not be large.

Example 1 *Suppose that* $\mathbf{y}_i = (y_{i1}, y_{i2})'$ *has a bivariate normal distribution:*

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \overset{i.i.d.}{\sim} N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right].$$

Under the bivariate normal distribution, a set of sufficient statistics for the parameter $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_{11}, , \sigma_{12}, \sigma_{22})'$ *is* $\sum_{i=1}^{n}\left(y_{i1}, y_{i2}, y_{i1}^2, y_{i1}y_{i2}, y_{i2}^2\right)$. *Therefore, constraint (35) can be satisfies if*

$$\begin{aligned} &\sum_{j=1}^{M} w_{ij(t)}^{*}\left\{1, y_{i1}^{*(j)}, \left(y_{i1}^{*(j)}\right)^2\right\} \\ = &\left\{1, E\left(y_{1i} \mid y_{2i}, \hat{\boldsymbol{\theta}}_{(t)}\right), \left\{E\left(y_{1i} \mid y_{2i}, \hat{\boldsymbol{\theta}}_{(t)}\right)\right\}^2 + \hat{\sigma}_{11\cdot 2(t)}\right\}, \quad \textit{for } i \in A_{MR} \end{aligned}$$

and

$$\begin{aligned} &\sum_{j=1}^{M} w_{ij(t)}^{*}\left\{1, y_{i2}^{*(j)}, \left(y_{i2}^{*(j)}\right)^2\right\} \\ = &\left\{1, E\left(y_{2i} \mid y_{1i}, \hat{\boldsymbol{\theta}}_{(t)}\right), \left\{E\left(y_{2i} \mid y_{1i}, \hat{\boldsymbol{\theta}}_{(t)}\right)\right\}^2 + \hat{\sigma}_{22\cdot 1(t)}\right\}, \quad \textit{for } i \in A_{RM} \end{aligned}$$

where

$$\begin{aligned} E\left(y_{1i} \mid y_{2i}, \hat{\boldsymbol{\theta}}\right) &= \hat{\mu}_1 + \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{22}}(y_{i2} - \hat{\mu}_2) \\ E\left(y_{2i} \mid y_{1i}, \hat{\boldsymbol{\theta}}\right) &= \hat{\mu}_2 + \frac{\hat{\sigma}_{12}}{\hat{\sigma}_{11}}(y_{i1} - \hat{\mu}_1), \end{aligned}$$

$\sigma_{11\cdot 2} = \sigma_{11} - \sigma_{12}^2/\sigma_{22}$, *and* $\sigma_{22\cdot 1} = \sigma_{22} - \sigma_{12}^2/\sigma_{11}$.

In practice, instead of (35), the fractional weights are computed from

$$\sum_{i \in A_c} \sum_{j=1}^{M} w_{ij(t)}^{*} \mathbf{t}\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}\right) = \sum_{i \in A_c} \mathbf{g}\left(\mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}_{(t)}\right), \tag{36}$$

where A_c is the set of sample indices in a cell c. Imposing fractional weighting constraints in each cell rather than for each unit reduces the chance of extreme weights.

Variance estimation with fractionally imputed data can be performed using linearization or replication. The plug-in method discussed in Section 3 is essentially the linearization method.

Assume that, under complete response, let $w_i^{[k]}$ be the k-th replication weight for unit i. Assume that the replication variance estimator

$$\hat{V}_n = \sum_{k=1}^{L} c_k \left(\hat{\theta}_n^{[k]} - \hat{\theta}_n\right)^2, \tag{37}$$

where $\hat{\theta}_n = \sum_{i=1}^{n} w_i y_i$ and $\hat{\theta}_n^{[k]} = \sum_{i=1}^{n} w_i^{[k]} y_i$, is consistent for the variance of $\hat{\theta}_n$.

For replication with the calibration fractional imputation method, we consider the following steps for creating replicated fractional weights. Here, we assume that the calibration fractional weights are computed from (36).

[Step 1] Compute $\hat{\boldsymbol{\theta}}^{[k]}$, the k-th replicate of $\hat{\boldsymbol{\theta}}$, using fractional weights.

[Step 2] Using the $\hat{\boldsymbol{\theta}}^{[k]}$ computed from Step 1, compute the replicated fractional weights by

$$\sum_{i \in A_c} w_i^{[k]} \sum_{j=1}^{M} w_{ij}^{*[k]} \mathbf{t}\left(\mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}\right) = \sum_{i \in A_c} w_i^{[k]} \mathbf{g}\left(\mathbf{y}_{i,obs}, \hat{\boldsymbol{\theta}}^{[k]}\right), \tag{38}$$

using the regression weighting technique.

Equation (38) is the calibration equation for the replicated fractional weights. In general, Step 1 can be computationally problematic since $\hat{\boldsymbol{\theta}}^{[k]}$ is computed from the iterative algorithm (10) for each replication. Thus, we consider an approximation for $\hat{\boldsymbol{\theta}}^{[k]}$ using Taylor linearization. Let

$$\bar{S}^{[k]}(\boldsymbol{\theta}) = \sum_{i=1}^{n} w_i^{[k]} \bar{\mathbf{s}}_i(\boldsymbol{\theta})$$

where $\mathbf{s}_i(\boldsymbol{\theta}) = E\left\{\mathbf{s}_i(\boldsymbol{\theta}) \mid \mathbf{y}_{i,obs}, \boldsymbol{\theta}\right\}$. Using (15) and (21), the approximation formula can be implemented as

$$\hat{\boldsymbol{\theta}}^{[k]} \cong \hat{\boldsymbol{\theta}} + \left[\hat{I}_{obs}^{[k]}\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1} \bar{S}^{[k]}\left(\hat{\boldsymbol{\theta}}\right), \tag{39}$$

where

$$\begin{aligned} \hat{I}_{obs}^{[k]}(\boldsymbol{\theta}) &= \sum_{i=1}^{n} w_i^{[k]} \sum_{j=1}^{M} w_{ij}^{*} \left\{-\partial \mathbf{s}_i\left(\boldsymbol{\theta}; \mathbf{y}_{i,obs}, \mathbf{y}_{i,mis}^{*(j)}\right) / \partial \boldsymbol{\theta}\right\} \\ &+ \sum_{i=1}^{n} w_i^{[k]} \left\{\bar{\mathbf{s}}_i^{*}(\boldsymbol{\theta})\right\}^{\otimes 2} - \sum_{i=1}^{n} w_i^{[k]} \sum_{j=1}^{M} w_{ij}^{*} \left\{\mathbf{s}_i^{*(j)}(\boldsymbol{\theta})\right\}^{\otimes 2} \end{aligned} \tag{40}$$

and

$$\bar{S}^{[k]}(\boldsymbol{\theta}) = \sum_{i=1}^{n} w_i^{[k]} \sum_{j=1}^{M} w_{ij}^{*} \mathbf{s}_i^{*(j)}(\boldsymbol{\theta}).$$

6. Simulation Study

In a limited simulation study, we generated $B = 5{,}000$ Monte Carlo samples of size $n = 200$ from a bivariate normal distribution with $\mu_1 = 0, \mu_2 = 2, \sigma_{11} = 1$, $\sigma_{12} = 1$, and $\sigma_{22} = 2$. The probability of both responding is 0.42, the probability of only y_1 responding 0.18, and the probability of only y_2 responding 0.28. We considered the following seven parameters:

1. Five parameters in the bivariate normal distribution:

$$\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22}$$

2. Proportion of y_1 less than 0.8.
3. Domain mean where the probability of being in the domain is 0.4. (The probability of being in the domain does not depend on y_1 or y_2.)

For each parameter, we have computed four estimators:

1. The MLE using the EM algorithm
2. The fractional imputation estimator proposed in Section 2 with $M = 100$ and $M = 10$.
3. The calibration fractional imputation estimator proposed in Section 5 with $M = 10$ using the regression weighting method.
4. Multiple imputation (MI) with $M = 10$ imputations.

In fractional imputation, imputed values are generated by a systematic sampling method described in Appendix B, with $M^* = 1,000$. The basic idea is to generate M^* initial imputed values and then use a version of systematic sampling to get the final M imputed values. In the calibration fractional imputation method, the regression fractional weights are computed by (35). In multiple imputation, the imputed values are generated from the posterior predictive distribution iteratively using Gibbs sampling.

For variance estimation, we considered the FI estimator (without calibration), the calibration FI estimator, and multiple imputation. For variance estimation of the fractional imputation, we used the plug-in estimator discussed in Section 3 and Section 4. For variance estimation of the calibration FI estimator, we used the one-step jackknife variance estimator discussed in Section 5. For variance estimation of the multiple imputation, we used the variance formula of Rubin (1987). Table 1 presents the Monte Carlo means and variances of the four estimators. Table 2 presents the Monte Carlo relative biases and t-statistics for the variance estimators. The t-statistic is the statistic for testing zero bias in the variance estimator.

For point estimation, the calibration FI estimator and the the EM method give the same values for the parameters specified in the model. The (uncalibrated) fractional imputation estimator shows fairly good efficiency for many parameters, which suggests that the systematic sampling method used in the fractional imputation is already quite efficient. Multiple imputation shows less efficiency than the FI estimators for all parameters.

For estimation of the proportion and the domain mean, it is possible for the FI estimator with $M = 100$ to be more efficient than the calibration FI estimator with $M = 10$ because these parameters are not directly considered in the calibration step. The differences in efficiencies for these two parameters are less than one percent.

For variance estimation of the FI estimators, both linearization and replication methods provide consistent estimates for the variance of the parameter estimates. Variance estimation for domain estimation is biased under multiple imputation, as was identified by Kim and Fuller (2004).

REFERENCES

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

Fay, R. E. (1992), "When are inferences from multiple imputation valid?" In *Proceedings in Survey Research Method Section,* Washington, DC: American Statistical Association, pp. 227–232.

Fisher, R.A. (1925), "Theory of statistical estimation," *Proceedings of the Cambridge Philosophical Society*, 22, 700–725.

Gelman, A., Meng, X.-L., and Stern, H. (1996), "Posterior predictive assessment of model fitness via realized discrepancies (with discussion)," *Statistica Sinica*, 6, 733–807.

Ibrahim, J. G. (1990), "Incomplete data in generalized linear models," *Journal of the American Statistical Association*, 85, 765–769.

Kim, J.K., Brick, M.J., Fuller, W.A., and Kalton, G. (2006), "On the bias of the multiple imputation variance estimator in survey sampling," *Journal of the Royal Statistical Society*, Ser. B, 68, 509–521.

Kim, J.K. and Fuller, W.A. (2004). "Fractional hot deck imputation," *Biometrika*, 91, 559–578.

Louis, T. A. (1982), "Finding the observed information matrix when using the EM algorithm," *Journal of the Royal Statistical Society*, Ser. B, 44, 226–233.

Robins, J.M. and Wang, N. (2000), "Inference for imputation estimators," *Biometrika*, 87, 113–124.

Rubin, D. B. (1976), "Inference and missing data," *Biometrika,* 63, 581–590.

Rubin, D.B. (1987), *Multiple imputation for nonresponse in surveys,* New York: Wiley.

Wang, C.-Y. and Pepe, M. S. (2000), "Expected estimating equations to accommodate covariate measurement error," *Journal of the Royal Statistical Society*, Ser. B, 62, 509–524.

Wang, N. and Robins, J.M. (1998), "Large-sample theory for parametric multiple imputation procedures," *Biometrika*, 85, 935–948.

Wei, G.C.G. and Tanner, M.A. (1990), "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm," *Journal of the American Statistical Association*, 85, 699–704.

Wu, C.F.J. (1983), "On the convergence properties of the EM algorithm," *The Annals of Statistics*, 11, 95–103.

Table 1: Monte Carlo means and variances of the imputed estimators, based on 5,000 samples

Parameter	Method	Mean	Variance
μ_1	EM	0.00	0.007247
	FI (M=100)	0.00	0.007247
	FI (M=10)	0.00	0.007250
	Calib. FI (M=10)	0.00	0.007247
	MI (M=10)	0.00	0.007538
μ_2	EM	2.00	0.01294
	FI (M=100)	2.00	0.01294
	FI (M=10)	2.00	0.01294
	Calib. FI (M=10)	2.00	0.01294
	MI (M=10)	2.00	0.01322
σ_{11}	EM	1.00	0.01564
	FI (M=100)	1.00	0.01565
	FI (M=10)	1.00	0.01575
	Calib. FI (M=10)	1.00	0.01564
	MI (M=10)	1.00	0.01664
σ_{12}	EM	1.00	0.02196
	FI (M=100)	1.00	0.02197
	FI (M=10)	1.00	0.02204
	Calib. FI (M=10)	1.00	0.02196
	MI (M=10)	1.00	0.02291
σ_{22}	EM	1.99	0.05635
	FI (M=100)	1.99	0.05636
	FI (M=10)	1.99	0.05671
	Calib. FI (M=10)	1.99	0.05635
	MI (M=10)	2.00	0.05892
Proportion	FI (M=100)	0.79	0.001008
	FI (M=10)	0.79	0.001012
	Calib. FI (M=10)	0.79	0.001011
	MI (M=10)	0.79	0.001044
Domain Mean	FI (M=100)	2.00	0.02375
	FI (M=10)	2.00	0.02375
	Calib. FI (M=10)	2.00	0.02375
	MI (M=10)	2.00	0.02430

Table 2: Monte Carlo relative biases and t-statistics of the variance estimators, based on 5,000 samples

Parameter	Method	Rel. Bias (%)	t-statistics
$Var(\hat{\mu}_1)$	Linearize (for FI with $M = 100$)	-2.27	-1.16
	Linearize (for FI with $M = 10$)	-2.02	-1.03
	One-step JK (for calibration FI)	0.79	0.40
	MI (M=10)	4.71	2.39
$Var(\hat{\mu}_2)$	Linearize (for FI with $M = 100$)	-0.49	-0.24
	Linearize (for FI with $M = 10$)	-0.33	-0.16
	One-step JK (for calibration FI)	1.67	0.83
	MI (M=10)	0.76	0.37
$Var(\hat{\sigma}_{11})$	Linearize (for FI with $M = 100$)	-0.05	-0.02
	Linearize (for FI with $M = 10$)	-0.85	-0.43
	One-step JK (for calibration FI)	7.19	3.55
	MI (M=10)	3.45	1.71
$Var(\hat{\sigma}_{12})$	Linearize (for FI with $M = 100$)	-1.45	-0.73
	Linearize (for FI with $M = 10$)	-2.37	-1.19
	One-step JK (for calibration FI)	4.31	2.15
	MI (M=10)	2.15	1.06
$Var(\hat{\sigma}_{22})$	Linearize (for FI with $M = 100$)	-2.85	-1.36
	Linearize (for FI with $M = 10$)	-5.82	-2.76
	One-step JK (for calibration FI)	-0.56	-0.26
	MI (M=10)	-2.17	-1.03
$Var(\hat{p})$	Linearize (for FI with $M = 100$)	-2.83	-1.43
	Linearize (for FI with $M = 10$)	3.11	1.57
	One-step JK (for calibration FI)	-0.46	-0.23
	MI (M=10)	22.98	11.14
$Var(\hat{\mu}_d)$	Linearize (for FI with $M = 100$)	4.95	2.44
	Linearize (for FI with $M = 10$)	8.48	4.18
	One-step JK (for calibration FI)	7.20	3.55
	MI (M=10)	33.64	16.50