6. Fractional Imputation in Survey Sampling

1 Introduction

• Consider a finite population of \( N \) units identified by a set of indices \( U = \{1, 2, \ldots, N\} \) with \( N \) known. Associated with each unit \( i \) in the population are study variables, \( z_i = (x_i, y_i) \) with \( y_i = (y_{i1}, \ldots, y_{ip}) \) being the vector of study variables that are subject to missingness and \( x_i \) being the vector of auxiliary variables that are always observed.

• We are interested in estimating \( \eta \), defined as a (unique) solution to the population estimating equation \( \sum_{i=1}^{N} U(\eta; z_i) = 0 \). Examples of \( \eta \) includes

  1. Population mean: \( U(\eta; x, y) = y - \eta \)
  2. Population proportion of \( Y \) less than \( q \): \( U(\eta; x, y) = I(y < q) - \eta \)
  3. Population \( p \)-th quantile: \( U(\eta; x, y) = I(y < \eta) - p \)
  4. Population regression coefficient: \( U(\eta; x, y) = (y - x\eta)x' \)
  5. Domain mean: \( U(\eta; x, y) = (y - \eta)D(x) \)

• Let \( A \) denote the set of indices for the units in a sample selected by a probability sampling mechanism with sample size \( n \). Under complete response, a consistent estimator of \( \eta \) is obtained by solving

\[
\sum_{i \in A} w_i U(\eta; z_i) = 0. \tag{1}
\]

where \( w_i = 1/\pi_i \) is the inverse of the first-order inclusion probability. Under some regularity conditions, we can establish that the solution \( \hat{\eta}_n \) to (1) converges in probability to \( \eta \) and is asymptotically normally distributed.
• Things to consider
  – Complex sampling: unequal probability of selection, multi-stage sampling
  – General purpose estimation: We do not know which parameter $\eta$ will be used by data analyst at the time of imputation.
  – Multivariate missingness with arbitrary missing pattern.
  – We cannot use large $M$. (We do not want to create a huge data file.)

• Two concepts of missing at random (MAR): (For simplicity, assume $p = 1$.)
  – Population missing at random (PMAR): MAR holds in the population level
    $$f(y \mid x) = f(y \mid x, \delta)$$
    That is, $Y \perp \delta \mid x$.
  – Sample missing at random (SMAR): MAR holds in the sample level
    $$f(y \mid x, I = 1) = f(y \mid x, I = 1, \delta)$$
    That is, $Y \perp \delta \mid (x, I = 1)$, where $I_i = 1$ if unit $i \in A$ and $I_i = 0$ otherwise.

• If the sampling design is such that $P(I = 1 \mid x, y) = P(I = 1 \mid x)$, which is often called noninformative sampling design, then PMAR implies SMAR. For informative sampling design, PMAR does not necessarily imply SMAR.

• Under PMAR, an imputed value $y_i^*$ of missing $y_i$ satisfies
  $$E\{y_i^* \mid x_i, \delta_i = 0\} = E\{y_i \mid x_i, \delta_i = 0\}$$ (2)
  while, under SMAR, it satisfies
  $$E\{y_i^* \mid x_i, I_i = 1, \delta_i = 0\} = E\{y_i \mid x_i, \delta_i = 0\}.$$ (3)

• Roughly speaking, fractional imputation is based on PMAR assumption, while multiple imputation (which will be covered next week) is based on SMAR assumption.
2 Parametric fractional imputation

- We assume that the finite population at hand is a realization from an infinite population, called superpopulation. In the superpopulation model, we often postulate a parametric distribution, \( f(y \mid x; \theta) \), which is known up to the parameter \( \theta \) with parameter space \( \Omega \). The parametric model has a joint density

\[
 f(y \mid x; \theta) = f_1(y_1 \mid x; \theta_1)f_2(y_2 \mid x, y_1; \theta_2) \cdots f_p(y_p \mid x, y_1, \cdots, y_{p-1}; \theta_p) \tag{4}
\]

where \( \theta_k \) is the parameter in the conditional distribution of \( y_k \) given \( x \) and \( y_1, \cdots, y_{k-1} \).

- For each \( y_i = (y_{i1}, \cdots, y_{ip}) \), we have \( \delta_i = (\delta_{i1}, \cdots, \delta_{ip}) \) where \( \delta_{ki} = 1 \) if \( y_{ki} \) is observed and \( \delta_{ki} = 0 \) otherwise.

- For example, \( p = 3 \), there are \( 8 = 2^3 \) possible missing patterns: \( A = A_{111} \cup A_{110} \cup \cdots \cup A_{000} \), where, for example, \( A_{100} \) is the set of sample indices with \( \delta_{i1} = 1, \delta_{i2} = 0, \) and \( \delta_{i3} = 0 \). For \( i \in A_{100} \), we need to create imputed values for \( y_{2i} \) and \( y_{3i} \) from \( f(y_{2i}, y_{3i} \mid x_i, y_{1i}) \).

- Without loss of generality, we can express \( y_i = (y_{obs,i}, y_{mis,i}) \), where \( y_{obs,i} \) and \( y_{mis,i} \) are the observed and the missing part of \( y_i \), respectively.

- Three steps for PFI under complex sampling
  
  1. Compute the pseudo maximum likelihood estimator of \( \theta \) using EM by PFI method with sufficiently large imputation size \( M \) (say, \( M = 1,000 \)).
  
  2. Select \( m \) (say, \( m = 10 \)) imputed values from the set of \( M \) imputed values.
  
  3. Construct the final fractional weights for the \( m \) imputed data.

- The first step is called Fully Efficient Fractional Imputation (FEFI) step. The second step is Sampling Step, The third step is called Weighting Step.

- Step 1 (FEFI step): The pseudo maximum likelihood estimator of \( \theta \) is computed by the following EM algorithm.
1. [I-step]: Set $t = 0$. Obtain $M$ imputed values of $y_{mis,i}$ generated from a proposal distribution $h(y_{mis,i} \mid x_i, y_{obs,i})$. One simple choice of $h(\cdot)$ is

$$h(y_{mis,i} \mid x_i, y_{obs,i}) = f(y_{mis,i} \mid x_i, y_{obs,i}; \hat{\theta}(0))$$ (5)

where $\hat{\theta}(0)$ is the initial estimator of $\theta$ obtained from the available respondents. Generating samples from (5) may require MCMC method or SIR (Sampling Importance Resampling) method. See Appendix B for an illustration of SIR method. Let $w_{ij}^{*(0)} = 1/M$ be the initial fractional weights for $y_{mis,i}^{*(j)}$.

2. [M-step]: Update the parameter $\hat{\theta}(t+1)$ by solving the following imputed score equation,

$$\hat{\theta}(t+1) : \text{solution to} \sum_{i \in A} w_i \sum_{j=1}^M w_{ij}^* S(\theta; x_i, y_{ij}^*) = 0,$$

where $y_{ij}^* = (y_{obs,i}, y_{mis,i}^{*(j)})$ and $S(\theta; x, y) = \partial \log f(y \mid x; \theta)/\partial \theta$ is the score function of $\theta$.

3. [W-step]: Set $t = t + 1$. Using the current value of the parameter estimates $\hat{\theta}(t)$, compute the fractional weights

$$w_{ij}^* \propto \frac{f(y_{obs,i}, y_{mis,i}^{*(j)} \mid x_i; \hat{\theta}(t))}{h(y_{mis,i}^{*(j)} \mid x_i, y_{obs,i})}$$

with $\sum_{j=1}^M w_{ij}^* = 1$. For $t = 0$, we set $w_{ij}^* = 1/M$.

4. Check if $w_{ij}^* > 1/m$ for some $j = 1, \cdots, M$. If yes, update the proposal distribution with $\hat{\theta}_0$ replaced by $\hat{\theta}(t)$ and goto [I-step]. If no, goto [M-step]. Stop if $\hat{\theta}(t)$ meets the convergence criterion.

[I-step] is the imputation step, [W-step] is the weighting step, and [M-step] is the maximization step. Note that the imputed values are not changed in the EM iteration. Only the fractional weights are updated.

• Step 2 (Sampling Step): For each $i$, we have $M$ possible imputed values $z_{i}^{*(j)} = (x_i, y_{obs,i}, y_{mis,i}^{*(j)})$ with their fractional weights $w_{ij}^*$, where $w_{ij}^*$ is computed from
the EM algorithm after convergence. For each $i$, we treat $z_i^* = \{z_i^*(j); j = 1, 2, \cdots, M\}$ as as a weighted finite population (with weight $w_{ij}^*$) and use an unequal probability sampling method to select a sample of size $m$ from $z_i^*$ using $w_{ij}^*$ as the selection probability. (We can use a PPS sampling or systematic πPS sampling to obtain an imputed data of size $m$.) Let $\tilde{z}_i^{*(1)}, \cdots, \tilde{z}_i^{*(m)}$ be the $m$ elements sampled from $z_i^*$ by the PPS sampling. That is,

$$
P(\tilde{z}_i^{*(k)} = z_i^*(j)) = w_{ij}^*, \quad \forall j = 1, \cdots, M; k = 1, \cdots, m.
$$

The fractional weights for the final $m$ imputed values are given by $\tilde{w}_{ij}^* = 1/m$.

- **Step 3 (Weighting Step):** Modify the initial fractional weights $\tilde{w}_{ij}^* = 1/m$ to satisfy the calibration constraints. The constraint is

$$
\sum_{i \in A} w_i \sum_{j=1}^{m} \tilde{w}_{ij}^* S(\hat{\theta}; x_i, \tilde{y}_{ij}^*) = 0 \tag{6}
$$

with $\sum_{j=1}^{m} \tilde{w}_{ij}^* = 1$, where $\hat{\theta}$ is the pseudo MLE of $\theta$ computed from the FEFI step and $\tilde{y}_{ij}^* = (y_{obs,i}; \tilde{y}_{mis,i}^{*(j)})$. That is, $\tilde{y}_{ij}^*$ is the $j$-th imputed element of $y_i$ selected from the PPS sampling in Step 2. A solution to this calibration problem is

$$
\tilde{w}_{ij}^* = \tilde{w}_{ij0}^* - \left\{ \sum_{i \in A} w_i S_{i0}^* \right\} \hat{T}^{-1} \tilde{w}_{ij0}^* (S_{ij}^* - \tilde{S}_{i0}^*)
$$

where

$$
S_{ij}^* = S(\hat{\theta}; x_i, \tilde{y}_{ij}^*)
$$

$$
\tilde{S}_{i0}^* = \sum_{j=1}^{m} \tilde{w}_{ij0}^* S_{ij}^*
$$

$$
\hat{T} = \sum_{i \in A} w_i \sum_{j=1}^{m} \tilde{w}_{ij}^* (S_{ij}^* - \tilde{S}_{i0}^*) (S_{ij}^* - \tilde{S}_{i0}^*)'.
$$

- Once the final fractional weights are computed, then the PFI estimator of $\eta$ is obtained by solving

$$
\sum_{i \in A} w_i \sum_{j=1}^{m} \tilde{w}_{ij}^* U(\eta; x_i, \tilde{y}_{ij}^*) = 0. \tag{7}
$$
Note that the above fractionally imputed estimating equation is an approximation to the following expected estimating equation

\[
\sum_{i \in A} w_i E\{ U(\eta; x_i, y_{obs,i}, Y_{mis,i}) \mid x_i, y_{obs,i}; \hat{\theta} \} = 0.
\]

- For variance estimation, we can use replication methods (such as jackknife or bootstrap). Details are given in Appendix C.

3 Nonparametric approach: Fractional Hot deck Imputation for multivariate continuous variable

- We do not want to make a parametric model assumptions about \( f(y_1, \cdots, y_p \mid x) \). However, some assumption of joint distribution of \((y_1, \cdots, y_p)\) is needed in order to preserve the correlation structure between the items.

- Easy if the data were categorical.

- Example (SRS of size \( n = 10 \))

<table>
<thead>
<tr>
<th>ID</th>
<th>Weight</th>
<th>( x )</th>
<th>( y_1 )</th>
<th>( y_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
<td>1</td>
<td>( y_{1,1} )</td>
<td>( y_{2,1} )</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>1</td>
<td>( y_{1,2} )</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
<td>1</td>
<td>M</td>
<td>( y_{2,3} )</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>1</td>
<td>( y_{1,4} )</td>
<td>( y_{2,4} )</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>1</td>
<td>( y_{1,5} )</td>
<td>( y_{2,5} )</td>
</tr>
<tr>
<td>6</td>
<td>0.10</td>
<td>2</td>
<td>( y_{1,6} )</td>
<td>( y_{2,6} )</td>
</tr>
<tr>
<td>7</td>
<td>0.10</td>
<td>2</td>
<td>M</td>
<td>( y_{2,7} )</td>
</tr>
<tr>
<td>8</td>
<td>0.10</td>
<td>2</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>9</td>
<td>0.10</td>
<td>2</td>
<td>( y_{1,9} )</td>
<td>( y_{2,9} )</td>
</tr>
<tr>
<td>10</td>
<td>0.10</td>
<td>2</td>
<td>( y_{1,10} )</td>
<td>( y_{2,10} )</td>
</tr>
</tbody>
</table>

M: Missing

- Fractional Imputation Idea: If both \( y_1 \) and \( y_2 \) are categorical, then fractional imputation is easy to apply.
  
  - We have only finite number of possible values.
  - Imputed values = possible values
The fractional weights are the conditional probabilities of the possible values given the observations.

Can use “EM by weighting” method of Ibrahim (1990) to compute the fractional weights.

- Example \((y_1, y_2; \text{dichotomous, taking 0 or 1})\)

<table>
<thead>
<tr>
<th>ID</th>
<th>Weight</th>
<th>(x)</th>
<th>(y_1)</th>
<th>(y_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.10</td>
<td>1</td>
<td>(y_{1,1})</td>
<td>(y_{2,1})</td>
</tr>
<tr>
<td>2</td>
<td>0.10(w_{2,1}^*)</td>
<td>1</td>
<td>(y_{1,2})</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.10(w_{2,2}^*)</td>
<td>1</td>
<td>(y_{1,2})</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.10(w_{3,1}^*)</td>
<td>1</td>
<td>0</td>
<td>(y_{2,3})</td>
</tr>
<tr>
<td></td>
<td>0.10(w_{3,2}^*)</td>
<td>1</td>
<td>1</td>
<td>(y_{2,3})</td>
</tr>
<tr>
<td>4</td>
<td>0.10</td>
<td>1</td>
<td>(y_{1,4})</td>
<td>(y_{2,4})</td>
</tr>
<tr>
<td>5</td>
<td>0.10</td>
<td>1</td>
<td>(y_{1,5})</td>
<td>(y_{2,5})</td>
</tr>
<tr>
<td>6</td>
<td>0.10</td>
<td>2</td>
<td>(y_{1,6})</td>
<td>(y_{2,6})</td>
</tr>
<tr>
<td>7</td>
<td>0.10(w_{7,1}^*)</td>
<td>2</td>
<td>0</td>
<td>(y_{2,7})</td>
</tr>
<tr>
<td></td>
<td>0.10(w_{7,2}^*)</td>
<td>2</td>
<td>1</td>
<td>(y_{2,7})</td>
</tr>
<tr>
<td>8</td>
<td>0.10(w_{8,1}^*)</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.10(w_{8,2}^*)</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.10(w_{8,3}^*)</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.10(w_{8,4}^*)</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>0.10</td>
<td>2</td>
<td>(y_{1,9})</td>
<td>(y_{2,9})</td>
</tr>
<tr>
<td>10</td>
<td>0.10</td>
<td>2</td>
<td>(y_{1,10})</td>
<td>(y_{2,10})</td>
</tr>
</tbody>
</table>

- Fractional weights are the conditional probabilities of the imputed values given the observations. For example,

\[
\begin{align*}
 w_{2,1}^* &= \hat{P}(y_2 = 0 \mid x = x_2, y_1 = y_{1,2}) \\
 w_{3,1}^* &= \hat{P}(y_1 = 0 \mid x = x_3, y_2 = y_{3,2}) \\
 w_{7,1}^* &= \hat{P}(y_1 = 0 \mid x = x_7, y_2 = y_{2,7})
\end{align*}
\]

and

\[
 w_{8,1}^* = \hat{P}(y_1 = 0, y_2 = 0 \mid x = x_8).
\]

The conditional probabilities are computed from the joint probabilities.

- M-step: Update the joint probability \(\pi_{be|a} = P(y_1 = b, y_c = b \mid x = a)\) by solving

\[
\hat{\pi}_{be|a} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{M_i} w_i w_{i,j}^* I(x_i = a, y_{1i}^*(j) = b, y_{2i}^*(j) = c)}{\sum_{i=1}^{n} w_i I(x_i = a)}.
\]
• For continuous $y$, let’s consider an approximation using categorical transformation.

For simplicity, let $Y = (Y_1, Y_2, Y_3)$ be the study variables that have missingness.

1. Preliminary Step

For each item $k$, create a transformation of $Y_k$ into $\tilde{Y}_k$, a discrete version of $Y_k$.

The value of $\tilde{Y}_k$ will serve the role of imputation cell for $Y_k$. If $Y_k$ is missing, then $\tilde{Y}_k$ is also missing. Let $M_k$ be the number of cells for item $Y_k$. The maximum number of cells for $p = 3$ is then $G = M_1 \times M_2 \times M_3$.

2. FEFI Step: Two-stage imputation is used. For each $i$ in the sample, $\tilde{y}_i$ is decomposed into $\tilde{y}_i = (\tilde{y}_{obs,i}, \tilde{y}_{mis,i})$. In the stage 1 imputation, we impute the imputation cells. In the stage 2 imputation, we impute for missing observations within imputation cells. To perform two-stage imputation, we first compute the estimated joint probability $\tilde{\pi}_{ijk} = \Pr(\tilde{y}_1 = i, \tilde{y}_2 = j, \tilde{y}_3 = k)$ using the EM algorithm (or other estimation methods).

(a) Stage 1 imputation: For each $i$, identify all possible values of $\tilde{y}_{mis,i}$. Let $G_i$ be the number of possible values of $\tilde{y}_{mis,i}$. In the stage 1 FEFI method, we create $G_i$ imputed values of $\tilde{y}_{mis,i}$ with the fractional weights corresponding to $\tilde{y}^*_{mis,i(g)}$ is

$$w^*_{ig(1)} = \frac{\tilde{\pi}(\tilde{y}_{i,obs}, \tilde{y}^*_{i,mis(g)})}{\sum_g \tilde{\pi}(\tilde{y}_{i,obs}, \tilde{y}^*_{i,mis(g)})}$$

where $\tilde{y}^*_{i,mis(g)}$ is the $g$-th realization of the $\tilde{y}_{i,mis}$, the missing part of $\tilde{Y}$ for unit $i$.

(b) Stage 2 imputation: For each $g$-th imputed cell in Stage 1 imputation, we identify the donor set from the respondents of $Y_{mis(i)}$ to impute all the observed values of $Y_k$ in the same cell. For example, if we observe $y_{1i}$ and $y_{3i}$ but $y_{2i}$ is not observed. In this case, the donor set for unit $i$ is $D_i = \{ j \in A; \delta_{1j} = \delta_{3j} = 1, \tilde{y}_{1j} = \tilde{y}_{1i}, \tilde{y}_{3j} = \tilde{y}_{3i} \}$. The within-cell fractional
weight for donor $j$ is then

$$w_{ij}(2) = \frac{w_j}{\sum_{j \in D_i} w_j}.$$ 

The final fractional weight for donor $j$ in unit $i$ is

$$w_{ij}^* = w_{ij}^*(1)w_{ij}^*(2).$$

Note that $\sum_j w_{ij}^* = 1$. Note that, for the fractional weight $w_{ij}^*$, the imputed value for $y_i = (y_{obs,i}, y_{mis,i})$ is $y_{ij}^* = (y_{obs,i}, y_{mis(i),j})$, where $y_{mis(i),j}$ is the value corresponding to variable $y_{mis(i)}$ for unit $j$.

3. Sampling Step

From the two stage FEFI data, we use a PPS sampling (with rejective sampling) and calibration weighting to obtain an approximation. That is, for each $i$, from the set $\{ (w_{ij}^*, y_{ij}^*); j \in A \}$, we perform a systematic PPS sampling of size $m$ using $w_{ij}^*$ in (9) as the size measure. Let $y_{i1}^*, \ldots, y_{im}^*$ be the $m$ imputed values from the PPS selection. The initial fractional weight assigned to $y_{ij}^*$ is given by $\tilde{w}_{ij0} = 1/m$.

4. Weighting Step

The fractional weights are further adjusted to match to the marginal probabilities $\tilde{\pi}_{i+i}$, $\tilde{\pi}_{i+j}$, and $\tilde{\pi}_{i+k}$. That is, we may use

$$\sum_{i \in A} w_i \left\{ \delta_i I(\tilde{y}_i = g) + (1 - \delta_i) \sum_{j=1}^m \tilde{w}_{ij}^* I(\tilde{y}_{ij}^* = g) \right\} = \hat{\pi}_g, \ g = 1, \ldots, G.$$ 

In this case, a raking ratio estimation method can be used to achieve these calibration constraints.

For variance estimation, we can use the replication method by repeating [Step 2]-[Step 4] to obtain replicated fractional weights.
Appendix

A. Lemma 6.1

Lemma 6.1: If either (2) or (3) holds, the imputed estimator of \( Y = \sum_{i=1}^{N} y_i \) of the form \( \hat{Y}_I = \sum_{i \in A} w_i \{ \delta_i y_i + (1 - \delta_i) y_i^* \} \) is unbiased for \( Y \) in the sense that \( E(\hat{Y}_I - Y) = 0 \).

Proof. We first introduce an extended definition of \( \delta_i \), where \( \delta_i = 1 \) if unit \( i \) responds when sampled and \( \delta_i = 0 \) otherwise. With this extended definition, \( \delta_i \) is defined throughout the finite population. Fay (1992), Shao and Steel (1999) and Kim and Rao (2009) also used this extended definition.

Now, we will first show that (3) implies unbiased. Let \( \hat{Y}_n = \sum_{i \in A} w_i y_i \) be an unbiased estimator of \( Y \). Note that

\[
\hat{Y}_I - \hat{Y}_n = \sum_{i=1}^{N} I_i w_i (1 - \delta_i)(y_i^* - y_i). \tag{10}
\]

Thus, writing \( I = (I_1, I_2, \ldots, I_N) \) and \( \delta = (\delta_1, \ldots, \delta_N) \),

\[
E\{\hat{Y}_I - \hat{Y}_n \mid I, \delta\} = \sum_{i=1}^{N} I_i w_i (1 - \delta_i) E\{y_i^* - y_i \mid I_i = 1, \delta_i = 0\} = 0. \tag{11}
\]

Since we can write

\[
E\{\hat{Y}_I - Y\} = E\{\hat{Y}_I - \hat{Y}_n\} + E\{\hat{Y}_n - Y\}, \tag{12}
\]

the first term is zero by (11) and the second term is zero by the design-unbiasedness of \( \hat{Y}_n \).

Finally, we will show that condition (2) also implies unbiasedness. From (10), by taking expectation with respect to the sampling design, we have

\[
E\{\hat{Y}_I - \hat{Y}_n \mid \delta, Y\} = \sum_{i=1}^{N} (1 - \delta_i) (y_i^* - y_i)
\]

and so

\[
E\{\hat{Y}_I - \hat{Y}_n \mid \delta\} = \sum_{i=1}^{N} (1 - \delta_i) E\{y_i^* - y_i \mid \delta_i = 0\} = 0.
\]

Therefore, the unbiasedness of the imputed estimator also follows as the first term of (12) can be shown to be zero from (2). \( \square \)
B. SIR algorithm in the I-step

To discuss I-step in Step 1, we give an illustration for \( p = 2 \). Extension to \( p > 2 \) case is straightforward. Note that the joint density can be written

\[
f(y_1, y_2 | x) = f_1(y_1 | x; \theta_1) f_2(y_2 | x, y_1; \theta_2).
\]

The sample is partitioned into four sets, \( A_{11}, A_{10}, A_{01}, A_{00} \), according to the missing patterns. We first obtain the initial parameter estimate \( \hat{\theta}(0) = (\hat{\theta}_1(0), \hat{\theta}_2(0)) \) using available respondents. That is, we use the observations in \( A_{11} \cup A_{10} \) to estimate \( \theta_1 \) and use the observations in \( A_{11} \) to estimate \( \theta_2 \).

Now, we want to generate \( M \) imputed values from (5). In the case of \( p = 2 \), we have the proposal distribution can be written as

\[
h(y_{\text{mis}, i} | x_i, y_{\text{obs}, i}) = \begin{cases} 
  f_2(y_{2i} | x_i, y_{1i}; \hat{\theta}_2(0)) & \text{if } i \in A_{10} \\
  f(y_{1i} | x_i, y_{2i}; \hat{\theta}(0)) & \text{if } i \in A_{01} \\
  f(y_{1i}, y_{2i} | x_i; \hat{\theta}(0)) & \text{if } i \in A_{00}
\end{cases}
\]

where

\[
f(y_{1i} | x_i, y_{2i}; \hat{\theta}(0)) = \frac{f_1(y_{1i} | x_i; \hat{\theta}_1(0)) f_2(y_{2i} | x_i, y_{1i}; \hat{\theta}_2(0))}{\int f_1(y_{1i} | x_i; \hat{\theta}_1(0)) f_2(y_{2i} | x_i, y_{1i}; \hat{\theta}_2(0)) dy_{1i}}. \tag{13}
\]

Except for some special cases such as normal \( f_1 \) and normal \( f_2 \), the conditional distribution in (13) is not of known form. Thus, some computational tools (such as Metropolis-Hastings algorithm) to generate samples from (13) for \( i \in A_{01} \). We introduce SIR (Sampling Importance Resampling) algorithm as an alternative computational tool for generating imputed values from (13). The SIR consists of the following steps:

1. Generate \( B \) (say \( B = 100 \)) samples from \( f_1(y_{1i} | x_i; \hat{\theta}_1(0)) \).

2. Select a PPS sample of size one from the \( B \) elements of \( y_{1i}^* \) with size measure \( f_2(y_{2i} | x_i, y_{1i}^*; \hat{\theta}_2(0)) \).

3. Repeat Step 1 and Step 2 independently \( M \) times to obtain \( M \) imputed values.

Once we obtain \( M \) imputed values of \( y_{1i} \), we can use

\[
\hat{h}(y_{\text{mis}, i} | x_i, y_{\text{obs}, i}) \propto f_1(y_{1i} | x_i; \hat{\theta}_1(0)) f_2(y_{2i} | x_i, y_{1i}; \hat{\theta}_2(0))
\]
as an estimator for the proposal density in (5). Since \( \sum_{j=1}^{M} w_{ij}^* = 1 \), we do not need to compute the density for the conditional density in (13).

C. Replication variance estimation

In the variance estimation, we use a replication variance method. Let \( \hat{Y}_n = \sum_{i \in A} w_i y_i \) be the complete sample estimator of \( Y \) under complete response and let

\[
\hat{V}_{\text{rep}} = \sum_{k=1}^{n} c_k \left( \hat{Y}^{(k)}_n - \hat{Y}_n \right)^2
\]

be a replication variance estimator with \( \hat{Y}^{(k)}_n = \sum_{i \in A} w_i^{(k)} y_i \).

To discuss variance estimation of the PFI method presented in Section 2, recall that the PFI method consists of three steps: (1) FEFI step, (2) Sampling Step, (3) Weighting Step. We mimic the procedures for each replication but want to avoid regenerating the imputed values. The proposed variance estimation employs similar steps but uses the replication weights \( w_i^{(k)} \) instead of the original weight \( w_i \).

1. FEFI step: Compute the replicate \( \hat{\theta}^{(k)} \) of \( \hat{\theta} \) by applying the same EM algorithm using \( w_i \) replaced by \( w_i^{(k)} \).

2. Sampling Step: We will use the same imputed data for each replication. The replicates for the fractional weights for the final \( m \) imputed values are given by

\[
\tilde{w}_{ij0}^{* (k)} \propto \frac{f(y_{obs,i}, \tilde{y}^{*(j)}_{mis,i}; \hat{\theta}^{(k)})}{f(y_{obs,i}, \tilde{y}^{*(j)}_{mis,i}; \hat{\theta})}
\]

with \( \sum_{j=1}^{m} \tilde{w}_{ij0}^{* (k)} = 1 \).

3. Weighting step: Modify the initial fractional weights \( \tilde{w}_{ij0}^{* (k)} \) in (14) to satisfy the following calibration constraint

\[
\sum_{i \in A} w_i^{(k)} \sum_{j=1}^{m} \tilde{w}_{ij}^{* (k)} S(\hat{\theta}^{(k)}; x_i, \tilde{y}_{ij}^*) = 0
\]

with \( \sum_{j=1}^{m} \tilde{w}_{ij}^{* (k)} = 1 \), where \( \hat{\theta}^{(k)} \) is the pseudo MLE of \( \theta \) computed from the FEFI step using the replication weights.
Once the final replicated fractional weights are computed, then the variance estimation of $\hat{\eta}_{PFI}$ obtained from (7) is given by

$$\hat{V}_{PFI} = \sum_{k=1}^{n} c_k \left( \hat{\eta}^{(k)}_{PFI} - \hat{\eta}_{PFI} \right)^2$$

where $\hat{\eta}^{(k)}_{PFI}$ is computed from

$$\sum_{i \in A} w_i^{(k)} \sum_{j=1}^{m} \tilde{w}_{ij}^{(k)} U(\eta; x_i, \tilde{y}_{ij}^*) = 0.$$