Chapter 6: Fractional Imputation

Jae-Kwang Kim

Department of Statistics, Iowa State University

Recall: Monte Carlo EM

Remark

- Monte Carlo EM can be used as a frequentist approach to imputation.
- Convergence is not guaranteed (for fixed m).
- E-step can be computationally heavy. (May use MCMC method).

Parametric Fractional Imputation (Kim, 2011)

Parametric fractional imputation

1. More than one (say m) imputed values of $\mathbf{y}_{mis,i}$: $\mathbf{y}_{mis,i}^{*(1)}, \cdots, \mathbf{y}_{mis,i}^{*(m)}$ from some (initial) density $h(\mathbf{y}_{mis,i})$.
2. Create weighted data set

$$\left\{ \left(w_{ij}^*, \mathbf{y}_{ij}^* \right); j = 1, 2, \cdots, m; i = 1, 2 \cdots, n \right\}$$

where $\sum_{j=1}^m w_{ij}^* = 1$, $\mathbf{y}_{ij}^* = (\mathbf{y}_{obs,i}, \mathbf{y}_{mis,i}^{*(j)})$

$$w_{ij}^* \propto f(\mathbf{y}_{ij}^*, \boldsymbol{\delta}_i; \hat{\eta}) / h(\mathbf{y}_{mis,i}^{*(j)}),$$

$\hat{\eta}$ is the maximum likelihood estimator of η, and $f(\mathbf{y}, \boldsymbol{\delta}; \eta)$ is the joint density of $(\mathbf{y}, \boldsymbol{\delta})$.

3. The weight w_{ij}^* are the normalized importance weights and can be called fractional weights.

Remark

- Importance sampling idea: For sufficiently large m,

$$\sum_{j=1}^{m} w_{ij}^{*} g\left(y_{ij}^{*}\right) \cong \frac{\int g(y_i) \frac{f(y_i, \delta_i; \hat{\eta})}{h(y_{mis,i})} h(y_{mis,i}) dy_{mis,i}}{\int \frac{f(y_i, \delta_i; \hat{\eta})}{h(y_{mis,i})} h(y_{mis,i}) dy_{mis,i}} = E\left\{g\left(y_i\right) \mid y_{obs,i}, \delta_i; \hat{\eta}\right\}$$

  for any g such that the expectation exists.

- In the importance sampling literature, $h(\cdot)$ is called proposal distribution and $f(\cdot)$ is called target distribution.
- Do not need to compute the conditional distribution $f(y_{mis,i} \mid y_{obs,i}, \delta_i; \eta)$. Only the joint distribution $f(y_{obs,i}, y_{mis,i}, \delta_i; \eta)$ is needed because

$$\frac{f(y_{obs,i}, y_{mis,i}^{*(j)}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(j)})}{\sum_{k=1}^{m} f(y_{obs,i}, y_{mis,i}^{*(k)}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(k)})} = \frac{f(y_{mis,i}^{*(j)} \mid y_{obs,i}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(j)})}{\sum_{k=1}^{m} f(y_{mis,i}^{*(k)} \mid y_{obs,i}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(k)})}.$$

EM algorithm by fractional imputation

1. Imputation-step: generate $y_{i,\text{mis}}^{*(j)} \sim h\left(y_{i,mis}\right)$.
2. Weighting-step: compute

$$w_{ij(t)}^{*} \propto f\left(y_{ij}^{*}, \delta_i; \hat{\eta}_{(t)}\right) / h(y_{i,mis}^{*(j)})$$

   where $\sum_{j=1}^{m} w_{ij(t)}^{*} = 1$.
3. M-step: update

$$\hat{\eta}^{(t+1)} = \arg\max \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij(t)}^{*} \log f\left(\eta; y_{ij}^{*}, \delta_i\right).$$

4. (Optional) Check if $w_{ij(t)}^{*}$ is too large for some j. If so, set $h(y_{i,mis}) = f\left(y_{i,\text{mis}} \mid y_{i,\text{obs}}; \hat{\eta}_t\right)$ and goto Step 1.
5. Repeat Step 2 - Step 4 until convergence.

Remark

- "Imputation Step" + "Weighting Step" = E-step.
- The imputed values are not changed for each EM iteration. Only the fractional weights are changed.
  1. Computationally efficient (because we use importance sampling only once).
  2. Convergence is achieved (because the imputed values are not changed). See Theorem 6.1.
- For sufficiently large t, $\hat{\eta}^{(t)} \longrightarrow \hat{\eta}^*$. Also, for sufficiently large m, $\hat{\eta}^* \longrightarrow \hat{\eta}_{MLE}$.
- For estimation of ψ in $E\{U(\psi; Y)\} = 0$, simply use

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^* U(\psi; \mathbf{y}_{ij}^*) = 0$$

where $w_{ij}^* = w_{ij}^*(\hat{\eta})$ and $\hat{\eta}$ is obtained from the above EM algorithm.

Theorem 6.1 (Theorem 1 of Kim (2011))

Theorem

Let

$$Q^*(\eta \mid \hat{\eta}_{(t)}) = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij(t)}^* \log f\left(\eta; y_{ij}^*, \delta_i\right).$$

If

$$Q^*(\hat{\eta}_{(t+1)} \mid \hat{\eta}_{(t)}) \geq Q^*(\hat{\eta}_{(t)} \mid \hat{\eta}_{(t)}) \tag{1}$$

then

$$l_{\rm obs}^*(\hat{\eta}_{(t+1)}) \geq l_{\rm obs}^*(\hat{\eta}_{(t)}), \tag{2}$$

where $l_{\rm obs}^*(\eta) = \sum_{i=1}^{n} \ln\{f_{obs(i)}^*(\mathbf{y}_{i,obs}, \boldsymbol{\delta}_i; \eta)\}$ *and*

$$f_{obs(i)}^*(\mathbf{y}_{i,obs}, \boldsymbol{\delta}_i; \eta) = \frac{\sum_{j=1}^{m} f(\mathbf{y}_{ij}^*, \boldsymbol{\delta}_i; \eta)/h_m(\mathbf{y}_{i,\rm mis}^{*(j)})}{\sum_{j=1}^{m} 1/h_m(\mathbf{y}_{i,\rm mis}^{*(j)})}.$$

Proof

By using Jensen's inequality,

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

Therefore, (1) implies (2).

Example 6.3: Return to Example 3.15

- Fractional imputation
  1. Imputation Step: Generate $y_i^{*(1)}, \cdots, y_i^{*(m)}$ from $f\left(y_i \mid x_i; \hat{\theta}_{(0)}\right)$.
  2. Weighting Step: Using the m imputed values generated from Step 1, compute the fractional weights by

$$w_{ij(t)}^* \propto \frac{f\left(y_i^{*(j)} \mid x_i; \hat{\theta}_{(t)}\right)}{f\left(y_i^{*(j)} \mid x_i; \hat{\theta}_{(0)}\right)} \left\{1 - \pi(x_i, y_i^{*(j)}; \hat{\phi}_{(t)})\right\}$$

where

$$\pi(x_i, y_i; \hat{\phi}) = \frac{\exp\left(\hat{\phi}_0 + \hat{\phi}_1 x_i + \hat{\phi}_2 y_i\right)}{1 + \exp\left(\hat{\phi}_0 + \hat{\phi}_1 x_i + \hat{\phi}_2 y_i\right)}.$$

Example 6.3

- Using the imputed data and the fractional weights, the M-step can be implemented by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij(t)}^{*} S\left(\theta; x_i, y_i^{*(j)}\right) = 0$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij(t)}^{*} \left\{\delta_i - \pi(\phi; x_i, y_i^{*(j)})\right\}\left(1, x_i, y_i^{*(j)}\right) = 0, \qquad (3)$$

where $S(\theta; x_i, y_i) = \partial \log f(y_i \mid x_i; \theta)/\partial\theta$.

Example 6.4: Back to Example 3.18 (GLMM)

- Level 1 model
$$y_{ij} \sim f_1(y_{ij} \mid x_{ij}, a_i; \theta_1)$$
for some fixed θ_1 and a_i random.
- Level 2 model
$$a_i \sim f_2(a_i; \theta_2)$$
- Latent variable: a_i
- We are interested in generating a_i from
$$p(a_i \mid \mathbf{x}_i, \mathbf{y}_i; \theta_1, \theta_2) \propto \left\{ \prod_{j=1}^{n_i} f_1(y_{ij} \mid x_{ij}, a_i; \theta_1) \right\} f_2(a_i; \theta_2)$$

Example 6.4 (Cont'd)

- E-step
  1. Imputation Step: Generate $a_i^{*(1)}, \cdots, a_i^{*(m)}$ from $f_2(a_i; \hat{\theta}_2^{(t)})$.
  2. Weighting Step: Using the m imputed values generated from Step 1, compute the fractional weights by

$$w_{ij(t)}^* \propto g_1(\mathbf{y}_i \mid \mathbf{x}_i, a_i^{*(j)}; \hat{\theta}_1^{(t)})$$

  where $g_1(\mathbf{y}_i \mid \mathbf{x}_i, a_i; \hat{\theta}_1) = \prod_{j=1}^{n_i} f_1(y_{ij} \mid x_{ij}, a_i; \theta_1)$.

- M-step: Update the parameters by solving

$$\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij(t)}^* S_1\left(\theta_1; \mathbf{x}_i, \mathbf{y}_i, a_i^{*(j)}\right) = 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij(t)}^* S_2\left(\theta_2; a_i^{*(j)}\right) = 0.$$

Example 6.5: Measurement error model

- Interested in estimating θ in $f(y \mid x; \theta)$.
- Instead of observing x, we observe z which can be highly correlated with x.
- Thus, z is an instrumental variable for x:

  $$f(y \mid x, z) = f(y \mid x)$$

  and

  $$f(y \mid z = a) \neq f(y \mid z = b)$$

  for $a \neq b$.
- In addition to original sample, we have a separate calibration sample that observes (x_i, z_i).

Example 6.5 (Cont'd)

Table: Data Structure

	Z	X	Y
Calibration Sample	o	o	
Original Sample	o		o

- The goal is to generate x in the original sample from

$$
\begin{aligned}
f(x_i \mid z_i, y_i) &\propto f(x_i \mid z_i)\, f(y_i \mid x_i, z_i) \\
&= f(x_i \mid z_i)\, f(y_i \mid x_i)
\end{aligned}
$$

- Obtain a consistent estimator $\hat{f}(x \mid z)$ from calibration sample.
- E-step
  1. Generate $x_i^{*(1)}, \cdots, x_i^{*(m)}$ from $\hat{f}(x_i \mid z_i)$.
  2. Compute the fractional weights associated with $x_i^{*(j)}$ by

$$
w_{ij}^* \propto f(y_i \mid x_i^{*(j)}; \hat{\theta})
$$

- M-step: Solve the weighted score equation for θ.

Remarks for Computation

- Recall that, writing

$$\bar{S}^*(\eta \mid \eta) = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}^*(\eta) S(\eta; \mathbf{y}_{ij}^*, \boldsymbol{\delta}_i)$$

where $w_{ij}^*(\eta)$ is the fractional weight associated with y_{ij}^*, denoted by

$$w_{ij}^*(\eta) = \frac{f(\mathbf{y}_{ij}^*, \boldsymbol{\delta}_i; \eta)/h_m(\mathbf{y}_{i,\mathrm{mis}}^{*(j)})}{\sum_{k=1}^{m} f(\mathbf{y}_{ik}^*, \boldsymbol{\delta}_i; \eta)/h_m(\mathbf{y}_{i,\mathrm{mis}}^{*(k)})}, \tag{4}$$

and $S(\eta; \mathbf{y}, \boldsymbol{\delta}) = \partial \log f(\mathbf{y}, \boldsymbol{\delta}; \eta)/\partial\eta$, the EM algorithm for fractional imputation can be expressed as

$$\hat{\eta}^{(t+1)} \leftarrow \text{ solve } \bar{S}^*(\eta \mid \hat{\eta}^{(t)}) = 0.$$

Remarks for Computation

- Instead of EM algorithm, Newton-type algorithm can also be used. The Newton-type algorithm for computing the MLE from the fractionally imputed data is given by

$$\hat{\eta}^{(t+1)} = \hat{\eta}^{(t)} + \left\{ I_{obs}^{*}(\hat{\eta}^{(t)}) \right\}^{-1} \bar{S}^{*}(\hat{\eta}^{(t)} \mid \hat{\eta}^{(t)})$$

where

$$\begin{aligned} I_{obs}^{*}(\eta) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}(\eta)\dot{S}(\eta; \mathbf{y}_{ij}^{*}, \boldsymbol{\delta}_i) \\ &\quad -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}(\eta)\left\{ S(\eta; \mathbf{y}_{ij}^{*}, \boldsymbol{\delta}_i) - \bar{S}_i^{*}(\eta) \right\}^{\otimes 2}, \end{aligned}$$

$\dot{S}(\eta; \mathbf{y}, \boldsymbol{\delta}) = \partial S(\eta; \mathbf{y}, \boldsymbol{\delta})/\partial \eta$ and $\bar{S}_i^{*}(\eta) = \sum_{j=1}^{M} w_{ij}^{*}(\eta)\dot{S}(\eta; \mathbf{y}_{ij}^{*}, \boldsymbol{\delta}_i)$.

Estimation of general parameter

- Parameter Ψ is defined through $E\{U(\Psi; Y)\} = 0$.
- The FI estimator of Ψ is computed by solving

$$\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}^{*}(\hat{\eta}) U(\Psi; \mathbf{y}_{ij}^{*}) = 0. \tag{5}$$

Note that $\hat{\eta}$ is the solution to

$$\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}^{*}\left(\hat{\eta}\right) S\left(\hat{\eta}; \mathbf{y}_{ij}^{*}\right) = 0.$$

Estimation of general parameter

- We can use either linearization method or replication method for variance estimation. For linearization method, using Theorem 4.2, we can use sandwich formula

$$\hat{V}\left(\hat{\Psi}\right) = \hat{\tau}^{-1}\hat{\Omega}_q\hat{\tau}^{-1'} \tag{6}$$

where

$$\begin{aligned}\hat{\tau} &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}\dot{U}\left(\hat{\Psi};\mathbf{y}_{ij}^{*}\right)\\ \hat{\Omega}_q &= n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\left(\hat{q}_i^{*}-\bar{q}_n^{*}\right)^{\otimes 2},\end{aligned}$$

with $\hat{q}_i^* = \bar{U}_i^* + \hat{\kappa}\bar{S}_i^*$, where $(\bar{U}_i^*, \bar{S}_i^*) = \sum_{j=1}^{m} w_{ij}^*(U_{ij}^*, S_{ij}^*)$, $U_{ij}^* = U(\hat{\Psi}; \mathbf{y}_{ij}^*)$, $S_{ij}^* = S(\hat{\eta}; \mathbf{y}_{ij}^*)$, and

$$\hat{\kappa} = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}(\hat{\eta})\left(U_{ij}^{*}-\bar{U}_{i}^{*}\right)S_{ij}^{*}\left\{I_{\rm obs}^{*}(\hat{\eta})\right\}^{-1}$$

Estimation of general parameter

- For replication method, we first obtain the k-th replicate $\hat{\eta}^{(k)}$ of $\hat{\eta}$ by solving
$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_i^{(k)} w_{ij}^{*}\left(\eta\right) S\left(\eta; \mathbf{y}_{ij}^{*}\right) = 0.$$
- Once $\hat{\eta}^{(k)}$ is obtained, then the k-th replicate $\hat{\Psi}^{(k)}$ of $\hat{\Psi}$ is obtained by solving
$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_i^{(k)} w_{ij}^{*}(\hat{\eta}^{(k)}) U(\Psi; \mathbf{y}_{ij}^{*}) = 0$$
for ψ.
- The replication variance estimator of $\hat{\Psi}$ from (5) is obtained by
$$\hat{V}_{rep}(\hat{\Psi}) = \sum_{k=1}^{L} c_k \left(\hat{\Psi}^{(k)} - \hat{\Psi}\right)^2.$$

Fractional hot deck imputation (Yang and Kim, 2014)

- For simplicity, we consider a bivariate data structure (x_i, y_i) with x_i fully observed. Let $\{y_1, \cdots, y_r\}$ be the set of respondents and we want to obtain m imputed values, $y_i^{*(1)}, \cdots, y_i^{*(m)}$, for $i = r+1, \cdots, n$ from the respondents. Let w_{ij}^* be the fractional weights assigned to $y_i^{*(j)}$ for $j = 1, 2, \cdots, m$.
- We consider the special case of $m = r$. In this case, the fractional weight represents the point mass assigned to each responding y_i. Thus, it is desirable to compute the fractional weights $w_{i1}^*, \cdots, w_{ir}^*$ such that $\sum_{j=1}^r w_{ij}^* = 1$ and

$$\sum_{j=1}^{r} w_{ij}^* I(y_j < y) \cong Pr\left(y_i < y \mid x_i, \delta_i = 0\right). \tag{7}$$

- Note that for the special case of $w_{ij}^* = 1/r$, the left side of the above equality estimates $Pr\left(y_i < y \mid \delta_i = 1\right)$.

FHDI

- If we can assume a parametric model $f(y \mid x;\theta)$ for the conditional distribution of y on x and the response probability model is given by $Pr\left(\delta_i = 1 \mid x_i, y_i\right) = \pi(x_i, y_i; \phi)$, then the fractional weights satisfying (7) are given by

$$
\begin{aligned}
w_{ij}^{*} &\propto f\left(y_j \mid x_i, \delta_i = 0; \hat{\theta}, \hat{\phi}\right) / f(y_j \mid \delta_j = 1) \\
&\propto f\left(y_j \mid x_i, \hat{\theta}\right)\left\{1 - \pi\left(x_i, y_j; \hat{\phi}\right)\right\} / f(y_j \mid \delta_j = 1).
\end{aligned}
$$

FHDI

- Since

$$
\begin{aligned}
f\left(y_j \mid \delta_j = 1\right) &\propto \int \pi\left(x, y_j\right) f\left(y_j \mid x\right) f(x) dx \\
&\cong \frac{1}{n} \sum_{i=1}^{n} \pi\left(x_i, y_j\right) f\left(y_j \mid x_i\right),
\end{aligned}
$$

we can express

$$
w_{ij}^{*} \propto \frac{f(y_j \mid x_i, \hat{\theta}) \left\{1 - \pi(x_i, y_j; \hat{\phi})\right\}}{\sum_{k=1}^{n} \pi(x_k, y_j; \hat{\phi}) f(y_j \mid x_k; \hat{\theta})}.
$$

FHDI

- Under MAR, $\pi(x, y) = \pi(x)$ and the fractional weight is

$$w_{ij}^* \propto \frac{f(y_j \mid x_i; \hat{\theta})}{\sum_{k;\delta_k=1} f(y_j \mid x_k; \hat{\theta})}$$

  with $\sum_{j;\delta_j=1} w_{ij}^* = 1$.
- Once the fractional imputation is created, the imputed estimating equation for ψ is computed by

$$\sum_{i=1}^{n} \left\{ \delta_i U(\psi; x_i, y_i) + (1 - \delta_i) \sum_{j;\delta_j=1} w_{ij}^* U(\psi; x_i, y_j) \right\} = 0 \qquad (8)$$

  where

$$w_{ij}^* = \frac{f(y_j \mid x_i; \hat{\theta}) / \{\sum_{k;\delta_k=1} f(y_j \mid x_k; \hat{\theta})\}}{\sum_{j;\delta_j=1} \left[f(y_j \mid x_i; \hat{\theta}) / \{\sum_{k;\delta_k=1} f(y_j \mid x_k; \hat{\theta})\} \right]}. \qquad (9)$$

Example 6.7

- Consider the setup of Example 5.10, except that $e_i = u_i - 1$ where u_i is the exponential distribution with mean 1. Suppose that the imputation model for the error term is $e_i \sim N(0, \sigma^2)$. Thus, the imputation model is not correct because the true sampling distribution is $e_i \sim \text{Exp}(1) - 1$.
- We are interested in estimating $\theta_1 = E(Y)$ and $\theta_2 = P(Y < 1)$.
- A simulation study was performed to compare the three imputation methods: multiple imputation, parametric fractional imputation of Kim (2011), and fractional hot deck imputation, with $m = 50$ for all methods.

Table: Simulation Results of the Point Estimators

Parameter	Method	Bias	Standard Error
θ_1	MI	0.00	0.084
	PFI	0.00	0.084
	FHDI	0.00	0.084
θ_2	MI	–0.014	0.026
	PFI	–0.014	0.026
	FHDI	–0.001	0.029