Fractional Imputation for Longitudinal Data with Nonignorable Missing

Shu Yang

Department of Statistics, Iowa State University

February 17, 2012
Outline

- Introduction
  - EM algorithm
  - MCEM algorithm
- Proposed method: Parametric Fractional Imputation
- Approximation: Calibration Fractional Imputation
- Simulation study and Result
Longitudinal studies are designed to investigate changes over time in a characteristic measured repeatedly for each individual.

- E.g. In medical studies, blood pressure obtained from each individual, at different time points.

Mixed model could be used since individual effects can be modeled by the inclusion of random effects:

\[ y_{ij} = \beta x_{ij} + b_i + e_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m \]

- \( i \) indexes individual, and \( j \) indexes the repeated measurement on individual.
- \( b_i \ iid \sim N(0, \tau^2) \) specifying individual effects.
- \( e_{ij} \ iid \sim N(0, \sigma^2) \) specifying measurement error.
Introduction

- Missing data is frequently occurred, and destroys the representativeness of the remaining sample.

- Assumptions about the missing mechanism:
  - Missing at random (MAR) or ignorable: missingness is unrelated to the unobserved values;
  - Not missing at random (NMAR) or nonignorable: missingness depends on missing values;
  - A special case of the non-ignorable missing model

\[
Pr(r_{ij} = 1|x_{ij}, y_{ij}) = p(x_{ij}, b_i; \phi)
\]

- Solution: imputation assigning values for the missing responses to produce a complete data set.

- Pros:
  - Facilitate analyses using complete data analysis methods.
  - Ensure different analyses are consistent with one another.
  - Reduce nonresponse bias.
Basic Setup

- Data model: \( y_{ij} = \beta x_{ij} + b_i + e_{ij}, i = 1, \ldots, n, \ j = 1, \ldots, m \)
- \( x_{ij} \): always observed; \( y_{ij} \): subject to missing (scalar), \( b_i \): unobserved.

- \( r_{ij} = \begin{cases} 
1, & \text{if } y_{ij} \text{ is observed} \\
0, & \text{if } y_{ij} \text{ is missing} 
\end{cases} \)
- Response mechanism
  \[ r_{ji} | (x_{ij}, y_{ij}) \text{ indept} \sim \text{Bernoulli}(p_{ij}) \]
  \[ p_{ij} = p(x_{ij}, b_i; \phi) \text{ known up to } \phi \]

- Two types of parameter interested:
  - 1. \( \gamma = (\beta, \tau^2, \sigma^2, \phi) \): the parameters in the model
  - 2. \( \eta \): solution to \( E\{U(x, Y; \eta)\} = 0 \)
    - e.g.1: population mean \( U(x, Y; \eta) = \frac{1}{nm} \sum_{i,j=1}^{n,m} y_{ij} - \eta \).
    - e.g.2: proportion \( U(x, Y; \eta) = \frac{1}{nm} \sum_{i,j=1}^{n,m} I(y_{ij} \leq 2) - \eta \).
For the "complete" data for individual $i$: $(y_i, b_i, r_i)$.

$$p(y_i, b_i, r_i | \beta, \sigma^2, \tau^2, \phi) = p(y_i | \beta, \sigma^2, b_i)p(r_i | b_i, \phi)p(b_i | \tau^2) = \prod_{j=1}^{m} \left\{ p(y_{ij} | \beta, \sigma^2, b_i)p(r_{ij} | b_i, \phi) \right\} p(b_i | \tau^2)$$

$$l(\gamma) = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} \log p(y_{ij} | \beta, \sigma^2, b_i) + \sum_{j=1}^{m} \log p(r_{ij} | b_i, \phi) + \log p(b_i | \tau^2) \right\}$$

$$= l_1(\beta, \sigma^2) + l_2(\phi) + l_3(\tau^2)$$

MLE of $(\beta, \sigma^2)$, $\phi$ and $\tau^2$ can be obtained separately by solving the corresponding score equations:

- $S_1(\beta, \sigma^2) = \frac{\partial}{\partial(\beta, \sigma^2)} l_1(\beta, \sigma^2) = 0$
- $S_2(\phi) = l_2'(\phi) = 0$
- $S_3(\tau^2) = l_3'(\tau^2) = 0$
Introduction (EM algorithm)

- Expectation-maximization (EM) algorithm is an iterative method for finding mles of parameters.
- Denote \((X, Z)\): \(X\) is the observed part of the data, \(Z\) is the missing part of the data:

**Expectation step (E step):**

- Calculate the expected value of the log likelihood function, with respect to the conditional distribution of \(Z\) given \(X\) under the current estimate of the parameters \(\gamma^{(t)}\):

\[
Q(\gamma | \gamma^{(t)}) = E_{Z|X, \gamma^{(t)}}[\log L(\gamma; X, Z)]
\]

or

\[
\bar{S}(\gamma | \gamma^{(t)}) = E_{Z|X, \gamma^{(t)}}[S(\gamma; X, Z)],
S(\gamma; X, Z) = \frac{\partial}{\partial \gamma} \log L(\gamma; X, Z)
\]

**Maximization step (M step):**

- Find the parameter that maximizes this quantity:

\[
\gamma^{(t+1)} = \arg \max_{\gamma} Q(\gamma | \gamma^{(t)})
\]

or

\[
\gamma^{(t+1)} \text{ sol:} \bar{S}(\gamma | \gamma^{(t)}) = 0
\]
Apply the EM algorithm:

- Denote $y_i = (y_{\text{mis},i}, y_{\text{obs},i})$.
- $b_i$ is unobserved, we view it as missing data as well.
- $X = (y_{\text{obs},i})$, $Z = (y_{\text{mis},i}, b_i)$

**E-step:** Obtain the conditional mean score equation under current parameter value: (Intractable integral, high dim integral)

$$\bar{S}(\gamma|\gamma^{(t)}) = E_{y_{\text{mis}},b}[S_1(\beta, \sigma^2) + S_2(\phi) + S_3(\tau^2)|y_{\text{obs}}, r, \gamma^{(t)}]$$

**M-step:** Find the solution to the conditional mean score equations. (No problem)
Computing the conditional expectation can be a challenging problem.

Monte Carlo approximation of the conditional expectation using imputation

strong law of large numbers:

\[
E \{ S_{1ij}(\beta, \sigma^2) | y_{\text{obs}, i}, r_i, \hat{\gamma} \} \approx \frac{1}{M} \sum_{k=1}^{M} S^*_1(k)(\beta, \sigma^2)
\]

\[
S^*_1(k)(\beta, \sigma^2) = \frac{\partial \log p(y_{ij}^*(k) | \beta, \sigma^2, b_{i}^{(*)})}{\partial (\beta, \sigma^2)}
\]

where \((y_{ij}^*(k), b_{i}^{(*)}) \sim p(y_{ij}^*(k), b_{i}^{(*)}) | y_{\text{obs}, i}, r_i, \hat{\gamma})\)

\(\hat{\gamma}\) is a consistent estimator of \(\gamma\).
Introduction (MCEM)

- Monte Carlo EM (MCEM) algorithms:

  step 1. Using rejection sampling or Importance sampling generate $M$ values for each missing value:
  
  $$(y_{ij}^{*(k)}, b_{i}^{*(k)}) \sim p(y_{ij}^{*(k)}, b_{i}^{*(k)} | y_{obs, i}, r_{i}, \gamma^{(t)})$$

  step 2. (E-step) compute the imputed score function
  
  $$\frac{1}{M} \sum_{k=1}^{M} S_{1ij}^{*(k)}(\beta, \sigma^2) = 0,$$
  
  $$\frac{1}{M} \sum_{k=1}^{M} S_{2ij}^{*(k)}(\phi) = 0, \quad \frac{1}{M} \sum_{k=1}^{M} S_{3ij}^{*(k)}(\tau^2) = 0$$

  (M-step) solve to update $\gamma^{(t+1)}$

- MCEM regenerate the imputed values for each EM iteration and the computation is quite heavy.

- The convergence of the MCEM sequence is not guaranteed for fixed $M$. 
**Parametric Fractional Imputation (PFI)**

- **Main idea:**
  - **MCEM** assign equal weights $\frac{1}{M}$ to each imputed values, adjust imputed values at each $t$.
  - **PFI** generate imputed values once using importance sampling, adjusted weights at each $t$.

- $M$ imputed values for two missing: $b_i^{*(k)} \sim h_1(\cdot)$ and $y_{ij}^{*(k)} \sim h_2(\cdot | b_i^{*(k)})$.

- Create a weighted data set for the missing data:

  $$\{(w_{ij1}^{*}(k), b_i^{*(k)}) : k = 1, 2, ..., M; i = 1, \ldots, n\}$$
  $$\{(w_{ij2}^{*}(k), y_{ij}^{*(k)}) : k = 1, 2, ..., M; j \in NR\}$$

  $$\sum_{k=1}^{M} w_{ij1}^{*}(k) = 1, \quad w_{ij1}^{*}(k) \propto \frac{f(b_i^{*(k)} | y_{obs,i}, \hat{\gamma})}{h_1(b_i^{*(k)})}$$

  $$\sum_{k=1}^{M} w_{ij2}^{*}(k) = 1, \quad w_{ij2}^{*}(k) \propto \frac{f(y_{ij}^{*(k)} | b_i^{*(k)}, y_{obs,i}, \hat{\gamma})}{h_2(y_{ij}^{*(k)})h_1(b_i^{*(k)})}, \text{ where } \hat{\gamma} \text{ is the MLE of } \gamma.$$  

- The weights $w_{ij1}^{*}(k), w_{ij2}^{*}(k)$ are the normalized importance weights, can be called fractional weights.
The target dsn: \( p(y_{mis,ij}, b_i | y_{obs,i}, r_i, \gamma^{(t)}) \)

Need to simulate the missing data from the below densities:

\[
p(b_i | y_{obs,i}, r_i, \gamma^{(t)}) \propto p(b_i, y_{obs,i}, r_i | \gamma^{(t)}) \\
\propto p(y_{obs,i} | b, \gamma^{(t)}) p(r_i | b_i, \gamma^{(t)}) p(b_i | \gamma^{(t)}) \\
= \prod_{j \in A_{Ri}} p(y_{ij} | b_i, \gamma^{(t)}) \prod_{j=1}^{m} p(r_{ij} | b_i, \gamma^{(t)}) p(b_i | \gamma^{(t)})
\]

\[
p(y_{mis,ij} | y_{obs,i}, r_i, b_i, \gamma^{(t)}) \propto p(y_{mis,ij}, y_{obs,i}, r_i | b_i, \gamma^{(t)}) \\
\propto p(y_{mis,ij} | b_i, \gamma^{(t)}) p(y_{obs,i} | b, \gamma^{(t)}) p(r_i | b_i, \gamma^{(t)}) \\
= p(y_{mis,ij} | b_i, \gamma^{(t)}) \prod_{j \in A_{Ri}} p(y_{ij} | b_i, \gamma^{(t)}) \prod_{j=1}^{m} p(r_{ij} | b_i, \gamma^{(t)})
\]
EM algorithm by fractional imputation

**Step1** (Importance sampling) for $k = 1, \ldots, M$, $M$ is the Monte Carlo size.

1. For $i = 1, \ldots, n$, $b_i^{*(k)} \sim h_1(\cdot)$
2. For $j \in A_{Mi}$, $y_{mis,ij}^{*(k)} \sim h_2(\cdot)$

**Step2** (E-step) Calculate weights and update parameter value

$$w_{ij}^{*(k)}(t) \propto p(y_{obs,i} \mid b_i^{*(k)}, \gamma(t))p(r_i \mid b_i^{*(k)}, \gamma(t))p(b_i^{*(k)} \mid \gamma(t)) / h_1(b_i^{*(k)})$$

$$\sum_{k=1}^{M} w_{ij}^{*(k)}(t) = 1, \forall i, j$$

(M-step) solve

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{M} w_{ij}^{*(k)} S_2(\phi \mid r_{ij}, b_i^{*(k)}, \gamma(t)) = 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{M} w_{ij}^{*(k)} S_3(\tau^2 \mid b_i^{*(k)}, \gamma(t)) = 0$$
EM algorithm by fractional imputation

Continue (E-step) Calculate weights and update parameter value

\[ w_{ij2(t)}^{(k)} \propto p(y_{obs,i}|b_i^{(k)}, \gamma^{(t)})p(r_i|b_i^{(k)}, \gamma^{(t)})p(b_i^{(k)}|\gamma^{(t)}) \]
\[ \times p(y_{mis,ij}^{(k)}|b_i^{(k)}, \gamma^{(t)})p(y_{obs,i}|b_i^{(k)}, \gamma^{(t)}) \]
\[ \times p(r_i|b_i^{(k)}, \gamma^{(t)})/h_1(b_i^{(k)})h_2(y_{mis,ij}^{(k)}) \]
\[ \sum_{k=1}^{M} w_{ij2(t)}^{(k)} = 1, \forall i, j \]

(M-step) solve

\[ \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{M} w_{ij2(t)}^{(k)} S_1(\beta, \sigma^2|y_{mis,ij}^{(k)}, b_i^{(k)}, \gamma^{(t)}) = 0 \]

repeat until convergent. then \( \hat{\gamma} \).
Fractional Imputation

- Property: The imputed values are generated only once, only the weights are changed for each EM iteration. Computationally efficient.

- The EM sequence \( \{\hat{\gamma}(t); t = 1, 2, \ldots\} \) converges to a stationary point \( \hat{\gamma}^* \).

- For sufficiently large \( M \), \( \hat{\gamma}^* \) is asymptotically equivalent to the mle.

\[
\sum_{k=1}^{M} w_{ij}^{*}(k) S(x_{ij}, y_{ij}^{*}(k); \hat{\gamma}) \approx \frac{\int f(y_{ij}|x_{ij}, r_{ij}=0; \hat{\psi}) q(y_{ij}|x_{ij}) S(x_{ij}, y_{ij}^{*}(k); \hat{\psi}) q(y_{ij}|x_{ij}) dy_{ij}}{\int f(y_{ij}|x_{ij}, r_{ij}=0; \hat{\gamma}) q(y_{ij}|x_{ij}) dy_{ij}} = E\{S(x_{ij}, y_{ij}; \gamma|x_{ij}, r_{ij} = 0; \hat{\gamma})\}
\]
Using the final fractional weights after convergence, we can estimate $\eta$ by solving a fractionally imputed estimating equation

$$\bar{U}^*(\eta) = \frac{1}{nm} \sum_{i,j} \sum_{k=1}^{M} \bar{w}_{ij}^{*(k)} U(x_{ij}, y_{ij}^{*(k)}; \eta) = 0$$

- $\bar{U}^*(\eta)$ converges $\bar{U}(\eta|\hat{\psi}) = E\{U(X, Y; \eta)|X, y_{obs}, r; \hat{\psi}\}$ for sufficiently large $M$ almost everywhere.
- $\hat{\eta}^*$ is consistent and can obtain central limit theorem for it.
Calibration

- The proposed method is based on the large sample theory for Monte Carlo sampling, thus it requires $M$ to be large.

- For public use, small or moderate size $M$ is preferred.

- To reduce the Monte Carlo sampling error, consider a calibration weighting technique in survey sampling.

Suppose we have the initial imputation set of size $M_1 = 100$: \( \{y_{ij}^*(k), w_{ij0}^*(k)\} \) (population).

Sample a small subset from it \( \{y_{ij}^*(k), w_{ij}^*(k)\}_{i,j,k=1,...,M=10} \), adjust the weights such that

1. \( \sum_{k=1}^{M} w_{ij}^*(k) = 1, \forall i,j \)

2. \( \sum_{ijk} w_{ij}^*(k) S_1(\beta, \sigma^2 | y_{ij}^*(k), b_i^*(k), \hat{\gamma}) = 0 \)

- Using regression weighting technique, the weights can be calculated as

\[
w_{ij}^*(k) = w_{ij0}^*(k) - \left( \sum_{i,j} s_{ij}^* \right)^T \left\{ \sum_{i,j,k} w_{ij0}^*(k) (s_{ij}^*(k) - \bar{s}_{ij}^*) \otimes 2 \right\}^{-1} w_{ij0}^*(k) (s_{ij}^*(k) - \bar{s}_{ij}^*)
\]

where \( s_{ij}^*(k) = S_1(\beta, \sigma^2 | y_{ij}^*(k), b_i^*(k), \hat{\gamma}) \) and \( \bar{s}_{ij}^* = \sum_{k=1}^{M} w_{ij0}^*(k) s_{ij}^*(k) \).
Simulation Study

- $B = 2000$ Monte Carlo samples of sizes $n \times m = 10 \times 15 = 150$ were generated from
  \[ y_{ij} = \beta_0 + \beta_1 x_{ij} + b_i + e_{ij} \]
  where

- $x_{ij} = j/m$, $b_i \sim N(0, \tau^2)$, $iid$, $e_{ij} \sim N(0, \sigma^2)$,
- with $\beta_0 = 2$, $\beta_1 = 1$, $\sigma^2 = 0.1$, $\tau^2 = 0.5$

- $r_{ij} \sim Bernoulli(\pi_{ij})$
  \[ \text{logit}(\pi_{ij}) = \phi_0 + \phi_1 b_i \]
  with $\phi_0 = 0$, $\phi_1 = 1$.

- Under this model setup, the average response rate is about 50%.
Simulation study: simulation setup

The following parameters are computed.

1. $\beta_1, \tau^2, \sigma^2$: slope, variance components in the linear mixed effect model
2. $\mu_y$: the marginal mean of $y$.
3. Proportion: $Pr(Y < 2)$.

For each parameter, compute the following estimators:

1. Complete sample estimator (for comparison)
2. Incomplete sample estimator (if using the respondents only)
3. Parametric fractional imputation (PFI) with $M=100$ and $M=10$
4. The CFI method, with $M=10$. 
## Results

**Table:** Mean, variance of the point estimators based on 2000 ”Monte Carlo samples”

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Method</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>glmer</td>
<td>1.00</td>
<td>0.0086</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>0.99</td>
<td>0.0182</td>
</tr>
<tr>
<td></td>
<td>Imputation100</td>
<td>1.00</td>
<td>0.0191</td>
</tr>
<tr>
<td></td>
<td>Imputation10</td>
<td>1.00</td>
<td>0.0202</td>
</tr>
<tr>
<td></td>
<td>Calibration10</td>
<td>1.00</td>
<td>0.0200</td>
</tr>
<tr>
<td>$\mu_y$</td>
<td>glmer</td>
<td>2.54</td>
<td>0.0498</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>2.74</td>
<td>0.0548</td>
</tr>
<tr>
<td></td>
<td>Imputation100</td>
<td>2.55</td>
<td>0.0514</td>
</tr>
<tr>
<td></td>
<td>Imputation10</td>
<td>2.56</td>
<td>0.0515</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
<td>2.55</td>
<td>0.0514</td>
</tr>
<tr>
<td>$Pr(y &lt; 2)$</td>
<td>glmer</td>
<td>0.25</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>0.17</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td>Imputation100</td>
<td>0.25</td>
<td>0.0098</td>
</tr>
<tr>
<td></td>
<td>Imputation10</td>
<td>0.24</td>
<td>0.0097</td>
</tr>
<tr>
<td></td>
<td>Calibration10</td>
<td>0.25</td>
<td>0.0097</td>
</tr>
</tbody>
</table>
## Results

**Table**: Mean, variance of the point estimators based on 2000 ”Monte Carlo samples”

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Method</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau^2$</td>
<td>glmer</td>
<td>0.49</td>
<td>0.0534</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>0.48</td>
<td>0.0557</td>
</tr>
<tr>
<td></td>
<td>Imputation100</td>
<td>0.43</td>
<td>0.0469</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>glmer</td>
<td>0.10</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>Incomplete</td>
<td>0.10</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>Imputation100</td>
<td>0.10</td>
<td>0.0013</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>Imputation100</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>Imputation100</td>
<td>0.97</td>
<td></td>
</tr>
</tbody>
</table>
Results

- The incomplete sample estimators are biased for the mean type of the parameters. From the response model, individuals with large $b_i$ are likely to respond.

- For point estimation for the fixed effects and the mean type of parameters, the proposed estimators are essentially unbiased.

- Compared to the imputed estimator with $M=10$, the CFI estimator is more efficient, which is expected.

- For point estimation for the variance components, the proposed estimator is biased downwards.

$$\bar{S}_2^*(\tau) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^M w_{ij}^*(k) \left\{ (b_i^*(k))^2 - \tau^2 \right\} = 0$$

- Thus for large $M$, we are expecting $\hat{\tau}^2 \approx E[(b_i^*(k))^2 | y_{obs,i}, r_i, \hat{\gamma}]$, which would essentially be unbiased!

$$E[\hat{\tau}^2] \approx E[E[(b_i^*(k))^2 | y_{obs,i}, r_i, \hat{\gamma}]] = E[(b_i^*(k))^2] = \tau^2$$
**Discussion**

- MLE of the variance component is biased downwards since it fails to account for the loss of degrees of freedom needed to estimate $\beta$.

- Experiment: Using true value of $\beta$ and only estimate the variance components and the proposed method is unbiased! The problem could be the loss of degrees of freedom because of $\hat{\beta}$.

- The restricted maximum likelihood estimation (REML) can be derived from a hierarchy where integrate out the fixed effects first out of the integral. However, since the fixed effects are still hidden in the weights, this approach is not applicable.

- We can apply parametric bootstrapping to do bias correction; however, the objective of fractional imputation is to gain computational efficiency. Parametric bootstrapping is really not the thing we’d like to go for.
Table: Mean, variance of the point estimators based on 2000 ”Monte Carlo samples”

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Method</th>
<th>Mean</th>
<th>Variance</th>
<th>asym 95%ci</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>glmer</td>
<td>5.35</td>
<td>2.26</td>
<td>(5.28, 5.42)</td>
</tr>
<tr>
<td></td>
<td>Imputation</td>
<td>5.29</td>
<td>1.83</td>
<td>(5.22, 5.35)</td>
</tr>
<tr>
<td>$\tau^2$</td>
<td>glmer</td>
<td>5.41</td>
<td>23.98</td>
<td>(5.19, 5.63)</td>
</tr>
<tr>
<td></td>
<td>Imputation</td>
<td>2.26</td>
<td>2.68</td>
<td>(2.18, 2.33)</td>
</tr>
</tbody>
</table>
Thank You!