Parametric Fractional Imputation using Adjusted Profile Likelihood for Linear Mixed Models with Nonignorable Missing Data

Shu Yang* Jae-Kwang Kim† Zhengyuan Zhu‡

Abstract

Inference in the presence of missing data is a widely encountered and difficult problem in statistics. Imputation is often used to facilitate parameter estimation, which uses the complete sample estimators to the imputed data set. We consider the problem of parameter estimation for linear mixed models with non-ignorable missing values, which assumes the missingness depends on the missing values only through the random effects, leading to shared parameter models (Follmann and Wu,1995). We develop a parametric fractional imputation (PFI) method proposed by Kim (2011) under this non-ignorable response model, which simplifies the computation associated with the EM algorithm for maximum likelihood estimation with missing data. In the M-step, the restricted or adjusted profiled maximum likelihood method is used to reduce the bias of maximum likelihood estimation of the variance components. Results from a simulation study are presented to compare the proposed method with the existing methods, which demonstrates that imputation can significantly reduce the non-response bias and the idea of adjusted profiled maximum likelihood works nicely in PFI for the bias correction in estimating the variance components.

Key Words: EM algorithm, Random effect, Restricted maximum likelihood, Longitudinal data.

1. Introduction

Mixed models are the statistical models containing both fixed effects and random effects. These models are useful in a wide variety of disciplines in the physical, biological and social sciences. They are particularly useful in settings where repeated measurements are made on the same statistical units, or where measurements are made on clusters of related statistical units. For instance, mixed models are useful in longitudinal studies which are designed to investigate changes over time in a characteristic measured repeatedly for each individual.

However, missing data frequently occurs and destroys the representativeness of the remaining sample. There are several assumptions about the missing mechanism. If the missing probability is unrelated to the missing value after adjusting for the observed auxiliary information, the missing mechanism is called missing at random (MAR) or ignorable; whereas if the missing probability is related to the missing value even after adjusting for the auxiliary information, the missing mechanism is called missing not at random (MNAR) or non-ignorable. To model the missing mechanism one can consider either the selection model (Diggle and Kenward, 1994) or the pattern-mixture model (Little, 1995). In this paper we consider a special case of the selection model, where we assume that missingness only depends on the random effects, which yields the so-called shared parameter models, considered by Follmann and Wu (1995).

To carry out likelihood-based inference, we need to obtain the marginal density of the observed data, which involves integrating out the missing part of the data. Except for a few special cases this is analytically infeasible and thus requires numerical integration. Usually,

*Department of Statistics, Iowa State University, 1219 Snedecor Hall, Ames, IA 50011
†Department of Statistics, Iowa State University, 1208 Snedecor Hall, Ames, IA 50011
‡Department of Statistics, Iowa State University, 1216 Snedecor Hall, Ames, IA 50011

the marginal likelihood involves a high dimensional integral and numerical integration may not be feasible or reliable. One solution to this problem is imputation. By imputation, one can construct a complete data set by assigning reasonable values for the missing data. It has several advantages. Firstly, it facilitates the parameter estimation by simply applying the complete-sample estimators to the imputed data set. Secondly, it ensures different analysis are consistent with one another. Thirdly, it reduces the non-response bias.

Integration approximated by imputation under non-ignorable missing was considered by Greenlees et al. (1982). Ibrahim et al. (1999) considered continuous y variable using a Monte Carlo EM method of Wei and Tanner (1990) to compute the E-step of the EM algorithm in a generalized linear mixed model. Booth and Hobert (1999) used an automated Monte Carlo EM algorithm to compute the E-step of the EM algorithm to speed up the convergence rate. Chan and Kuk (1997) applied Gibbs sampling in the E-step to obtain maximum likelihood estimates for the probit normal model for binary data. McCulloch (1997) proposed a Monte Carlo Newton-Raphson algorithm used of importance sampling idea in maximum likelihood algorithms for generalized linear mixed models. For Monte Carlo EM algorithm, in each E-setp, the imputed values are regenerated and thus the computation quite heavy. Also the convergence of Monte Carlo sequence of the estimators is not guaranteed for fixed Monte Carlo sample size (Booth & Hobert, 1999).

In this paper, we develop a parametric fractional imputation (PFI) method proposed by Kim (2011) which can be used to simplify the Monte Carlo implementation of the EM algorithm, for linear mixed models with the shared parameter response model. The main idea in PFI is to produce a complete data set by imputation and each imputed value is associated with fractional weights, by which the observed likelihood can be approximated by the weighted mean of the imputed data likelihood. The resulting estimator is close to the maximum likelihood estimator and thus has very nice asymptotic properties, such as efficiency and asymptotic normality.

For mixed models, it is well known that the maximum likelihood estimation of variance components in mixed model is biased downwards. A method related to maximum likelihood is restricted maximum likelihood (REML), (Patterson & Thompson, 1971), where the effect of estimating fixed effect is taken into account for ML estimation. Thus the variance components are estimated without being affected by the fixed effects.

Another way of taking into account the bias in estimating variance components is using the adjusted profile likelihood. The simplest approach is to maximize out the fixed effects for the variance components and to construct the profile likelihood. The profile likelihood is then treated as an ordinary likelihood function for estimation and inference about the variance components. Unfortunately, with large numbers of nuisance parameters, this procedure can produce inefficient or even inconsistent estimates. A number of authors proposed the modified profile likelihood (Barndorff-Nielsen, 1986) and the closely related conditional profile likelihood (Cox and Reid, 1987), in which they correct the inconsistency of the profile likelihood in some problems and automatically make “degrees of freedom” adjustments in normal theory cases where accepted solutions are available for evaluation of their approaches. In this paper, we develop a novel PFI uses adjusted profile likelihood idea to correct the bias in estimating variance components.

In section 2, we introduce the basic setup. Section 3 develops the parametric fractional imputation for the non-ignorable missing data mechanism and discusses the incorporation of adjusted profile likelihood estimation in the parametric fractional imputation to reduce the bias in estimating variance components. Section 4 develops variance estimation based on Taylor linearization. Section 5 presents a simulation study and we conclude with discussion in Section 6.

2. Basic Setup

In this section we introduce the data model and the missing mechanism model considered in the paper. We consider the linear mixed model,

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + b_i + e_{ij,} i = 1, \dots, n,\ j = 1, \dots, m. \tag{1}$$

where i indexes individual, and j indexes the repeated measurement within each individual, $b_i\ iid \sim N(0, \tau^2)$ specifying the unobserved individual effects and $e_{ij}\ iid \sim N(0, \sigma^2)$ is the measurement error within individual.

Let $\mathbf{y_i} = (y_{i1}, \dots, y_{im})'$ be the complete measurements on the i^{th} individual if they are fully observed. The observed and missing components are denoted as $\mathbf{y_{mis,i}}$, $\mathbf{y_{obs,i}}$ respectively, so $\mathbf{y_i} = (\mathbf{y_{mis,i}}, \mathbf{y_{obs,i}})$. Let $\mathbf{r_i} = (r_{i1}, \dots, r_{im})'$ be vector of indicators of missing data, so $r_{ij} = 1$ if y_{ij} is observed, otherwise, $r_{ij} = 0$.

To model the missing mechanism, we consider the selection model

$$f(\mathbf{y_i}, \mathbf{r_i}, b_i) = f(\mathbf{y_i}|b_i) f(\mathbf{r_i}|\mathbf{y_i}, b_i) f(b_i). \tag{2}$$

Furthermore, in the selection model, we assume that missingness depends on the missing values only through the random effects $f(\mathbf{r_i}|\mathbf{y_i}, b_i) = f(\mathbf{r_i}|b_i)$, which leads to non-ignorable missingness.

Under this assumption, the joint density becomes

$$f(\mathbf{y_i}, \mathbf{r_i}, b_i) = f(\mathbf{y_i}|b_i) f(\mathbf{r_i}|b_i) f(b_i).$$

which is called the shared parameter models (Follmann and Wu,1995).

We further assume that conditional on b_i, $\{r_{ij}\}_{j=1}^m$ are independent. Then we have

$$\begin{aligned} f(\mathbf{r_i}|b_i, \phi) &= \prod_{j=1}^m f(r_{ij}|b_i, \phi) \\ &= \prod_{j=1}^m \{f(r_{ij} = 1|b_i, \phi)\}^{r_{ij}} \{1 - f(r_{ij} = 0|b_i, \phi)\}^{1-r_{ij}}. \end{aligned}$$

for some unknown parameter ϕ.

For the i^{th} individual, the complete data density of ($\mathbf{y_i}$,b_i, $\mathbf{r_i}$) is given by

$$\begin{aligned} f(\mathbf{y_i}, b_i, \mathbf{r_i}|\gamma) &= f(\mathbf{y_i}|\beta, \sigma^2, b_i) f(\mathbf{r_i}|b_i, \phi) f(b_i|\tau^2) \\ &= \prod_{j=1}^m \Big\{ f(y_{ij}|\beta, \sigma^2, b_i) f(r_{ij}|b_i, \phi) \Big\} f(b_i|\tau^2). \end{aligned}$$

where $\gamma = (\beta, \sigma^2, \tau^2, \phi)$. The complete log likelihood function of γ is thus given by

$$\begin{aligned} l_{com}(\gamma) &= \log f(\mathbf{y_i}, b_i, \mathbf{r_i}|\gamma) \\ &= \sum_{i=1}^n \log \Big\{ \Big(\prod_{j=1}^m f(y_{ij}|\beta, \sigma^2, b_i) f(r_{ij}|\phi, b_i) \Big) f(b_i|\tau^2) \Big\} \\ &= \sum_{i=1}^n \sum_{j=1}^m \log f(y_{ij}|\beta, \sigma^2, b_i) + \sum_{i=1}^n \sum_{j=1}^m \log f(r_{ij}|\phi, b_i) + \sum_{i=1}^n \log f(b_i|\tau^2) \\ &= l_1(\beta, \sigma^2) + l_2(\phi) + l_3(\tau^2). \end{aligned}$$

Under the complete response and assuming that b_i's are fully observed, the maximum likelihood estimator of γ can be obtained by maximizing $l_1(\beta, \sigma^2)$, $l_2(\phi)$, and $l_3(\tau^2)$, respectively.

When we only observe $(\mathbf{ry}, \mathbf{r})$, the observed density can be obtained by integrating out the unobserved random effects and missing values of the joint complete density,

$$f_{obs}(y_{obs};\gamma) = \prod_{i=1}^{n} \int\int \{(\prod_{j=1}^{m} p(y_{ij}|\beta,\sigma^2,b_i)p(r_{ij}|\phi,b_i))p(b_i|\tau^2)dy_{mis,ij}\}db_i.$$

Then the observed log likelihood function of γ is specified by,

$$\begin{aligned} l_{obs}(\gamma) &= \log f_{obs}(y_{obs},\gamma) \\ &= \sum_{i=1}^{n} \log\{\int\int(\prod_{j=1}^{m} f(y_{ij}|\beta,\sigma^2,b_i)f(r_{ij}|\phi,b_i))f(b_i|\tau^2)dy_{mis,ij}db_i\} \\ &= \sum_{i=1}^{n} \log f_{obs,i}(y_{i,obs};\gamma). \end{aligned}$$

where $f_{obs,i}(y_{i,obs};\gamma) = \int\int(\prod_{j=1}^{m} f(y_{ij}|\beta,\sigma^2,b_i)f(r_{ij}|\phi,b_i))f(b_i|\tau^2)dy_{mis,ij}db_i$. As we can see, since y_{ij} depends on b_i and r_{ij} depends on b_i as well, (β,σ^2), ϕ, τ^2 cannot be separated in $l_{obs}(\gamma)$ as we do in $l_{com}(\gamma)$. Thus parameters γ need to be estimated simultaneously.

Maximum likelihood estimator $\hat{\gamma}$ can be obtained by maximizing $l_{obs}(\gamma)$. Louis (1982) showed an alternative way to find the maximum likelihood estimator $\hat{\gamma}$ by maximizing

$$Q(\gamma) = E\{l_{com}(\gamma; y_{obs}, Y_{mis})|\ \mathbf{y_{obs}}, \mathbf{r}\}. \tag{3}$$

The conditional expectation usually has no analytical expression and thus requires numerical approximation. Instead of using the Monte Carlo EM method, Kim (2011) proposed a parametric fractional imputation method which modifies the importance sampling idea which largely reduces the computation burden and also guarantees convergence of estimation under fixed Monte Carlo size. We will apply the parametric fractional imputation idea to linear mixed model and also incorporate the adjusted profile likelihood idea in the next section.

3. Parametric fractional imputation

3.1 Fractional imputation in EM algorithm for Maximum likelihood estimation

To apply EM algorithm, write function (3) as

$$Q(\gamma|\gamma) = [Q_1(\beta,\sigma^2|\gamma), Q_2(\phi|\gamma)', Q_3(\tau^2|\gamma)].$$

where

$$\begin{aligned} Q_1(\beta,\sigma^2|\gamma) &= E\{l_1(\beta,\sigma^2)\mid \mathbf{y_{obs}}, \mathbf{r};\gamma\}. \\ Q_2(\phi|\gamma) &= E\{l_2(\phi)\mid \mathbf{y_{obs}}, \mathbf{r};\gamma\}. \\ Q_3(\tau^2|\gamma) &= E\{l_3(\tau^2)\mid \mathbf{y_{obs}}, \mathbf{r},\gamma\}. \end{aligned}$$

The MLE can be obtained by the EM-type algorithm,

$$\hat{\gamma}_{(t+1)} \leftarrow argmax Q(\gamma|\hat{\gamma}_{(t)})$$

The Monte Carlo EM method (MCEM) computes $Q(\gamma|\hat{\gamma}_{(t)})$ by regenerating the imputed values for each EM iteration and assigning equal weights $1/M$ to each imputed value.

The computation is cumbersome because it often requires an iterative algorithm such as Metropolis-Hastings algorithm for each EM iteration also it is not guaranteed for the convergence of the MCEM sequence of fixed Monte Carlo sample sizes. Alternatively, the parametric fractional imputation (PFI) modifies the idea of importance sampling to implement the Monte Carlo EM algorithm. In the PFI method, we generate the imputed values only in the beginning of the EM iteration and in each iteration keep the imputed values and only update the importance weights and the parameter estimates. Because the imputed values are not regenerated, it is much more computationally efficient and the convergence of the EM sequence is guaranteed.

We extend the PFI method to non-ignorable missing in linear mixed model setup. The M imputed values $b_i^{*(1)}, \ldots, b_i^{*(M)} \sim h_1(\cdot)$, $y_{ij}^{*(1)}, \ldots, y_{ij}^{*(M)} \sim h_2(\cdot)$ are generated from initial densities$h_1(b_i)$ and $h_2(y_{ij}|x_{ij})$ with the same support as $f(y_{ij})$. Given the current parameter estimates $\hat{\gamma}_{(t)}$ and the M imputed values $b_i^{*(1)}, \ldots, b_i^{*(M)}$ and $y_{ij}^{*(1)}, \ldots, y_{ij}^{*(M)}$ generated above, the joint density of $(\mathbf{y_{i,obs}}, \mathbf{y^{*(k)}_{i,mis}}, b_i^{*(k)})$ for each individual i, where $\mathbf{y^{*(k)}_{i,mis}} = (y^{*(k)}_{mis,ij})_{j \in Miss}$ is a vector of imputed values for missing, is

$$f_i^{*(k)}(\gamma) = \prod_{j=1}^{m} (f(y_{ij}^{*(k)}|\beta, \sigma^2, b_i^{*(k)}) f(r_{ij}|\phi, b_i^{*(k)})) f(b_i^{*(k)}|\tau^2). \tag{4}$$

For each individual i, assign the k^{th} imputed data vector $\mathbf{y_i^{*(k)}} = (\mathbf{y_{i,obs}}, \mathbf{y^{*(k)}_{i,mis}})$ a fractional weight as

$$w_i^{*(k)}(\gamma^{(t)}) = \frac{f_i^{*(k)}(\gamma^{(t)})/(\prod_{j\in M} h_2(y_{ij}^{*(k)}|b_i^{*(k)}))h_1(b_i^{*(k)})}{\sum_{l=1}^{M} f_i^{*(l)}(\gamma^{(t)})/(\prod_{j\in M} h_2(y_{ij}^{*(l)}|b_i^{*(l)}))h_1(b_i^{*(l)})}. \tag{5}$$

The Monte Carlo approximate of the observed likelihood function is

$$\begin{aligned} Q^*(\gamma|\gamma^{(t)}) &= \sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)}(\gamma^{(t)}) \log f_i^{*(k)}(\gamma) \qquad (6) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)}(\gamma^{(t)}) \big\{ \log f(y_i^{*(k)}|\beta, \sigma^2) + \log f(r_i|\phi) + \log f(b_i^{*(k)}|\tau^2) \big\} \\ &\equiv Q_1^*(\beta, \sigma^2|\gamma^{(t)}) + Q_2^*(\phi|\gamma^{(t)}) + Q_3^*(\tau^2|\gamma^{(t)}). \end{aligned}$$

where

$$\begin{aligned} Q_1^*(\beta, \sigma^2|\gamma^{(t)}) &= \sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)} \big(-\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_{ij}^{*(k)} - \beta_0 - \beta_1 x_{ij} - b_i^{*(k)})^2 \big). \\ Q_2^*(\phi|\gamma^{(t)}) &= \sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)} \big(\sum_{j=1}^{m}(r_{ij}(\phi_0 + \phi_1 b_i^{*(k)}) - \log(1 + \exp\{\phi_0 + \phi_1 b_i^{*(k)}\})) \big). \\ Q_3^*(\tau^2|\gamma^{(t)}) &= \sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)} \big(-\frac{1}{2}\log(2\pi\tau^2) - \frac{1}{2\tau^2}(b_i^{*(k)})^2 \big). \end{aligned}$$

Thus, the PFI method computes the E-step of the EM algorithm using fractional weights in (5). In the M-step, the updated parameters are computed by maximizing the imputed mean likelihood function. That is, we obtain $\hat{\gamma}_{(t+1)}$ by maximizing $Q_1^*(\beta, \sigma^2|\gamma^{(t)})$, $Q_2^*(\phi|\gamma^{(t)})$, $Q_3^*(\tau^2|\gamma^{(t)})$ for γ.

Maximizing $Q_1^*,\ Q_2^*,\ Q_3^*$ can be easily implemented by incorporating the fractional weights in the existing software. The EM sequence $\{\hat{\gamma}_{(t)};\ t = 1, 2, \ldots\}$ converges to a stationary point $\hat{\gamma}^*$ since the imputed values are unchanged and only the weights are changed. Under some regularity conditions, specified in Kim (2011), $\hat{\gamma}^*$ is asymptotically equivalent to the maximum likelihood estimator for large M.

3.2 Adjusted profile likelihood for bias correction

We now take into account the bias in estimating variance components by using the adjusted profile likelihood. The simplest approach is to maximize out the fixed effects for the variance components and to construct the profile likelihood. The profile likelihood is then treated as an ordinary likelihood function for estimation and inference about the variance components. Unfortunately, with large numbers of nuisance parameters, this procedure can produce inefficient or even inconsistent estimates. A number of authors proposed the modified profile likelihood (Barndorff-Nielsen, 1986) and the closely related conditional profile likelihood (Cox and Reid, 1987), in which they correct the inconsistency of the profile likelihood which automatically make "degrees of freedom" adjustments in normal theory cases where accepted solutions are available for evaluation of their approaches. The adjustment can be interpreted as the information concerning the variance components carried by the fixed effects in the ordinary profile likelihood.

In the normal case we shall see the adjusted profile likelihood matches exactly the restricted maximum likelihood (REML) (Patterson and Thompson, 1971) using the marginal distribution of the error term $\mathbf{y} - X\hat{\beta}_\theta$. The data can be divided into two independent parts, the error term $\mathbf{y} - X\hat{\beta}_\theta = S\mathbf{y}$ and $Q\mathbf{y}$, $S = I - X(X^t\Sigma^{-1}X)^{-}X^t\Sigma^{-1}$ and $Q = X^t\Sigma^{-1}$. The likelihood l_1 is separated into l_1' and l_1'',

$$\begin{aligned} l_1(\beta,\theta) &= l_1'(\theta) + l_1''(\beta,\theta) \\ &= P_\beta(l_1;\theta) + l_1''(\beta,\theta). \end{aligned}$$

where

$$\begin{aligned} P_\beta(l_1;\theta) &= l_p(\theta) - \frac{1}{2}\log|X^t\Sigma^{-1}X/(2\pi)| \\ &= -\frac{1}{2}\log|2\pi\Sigma| - \frac{1}{2}(\mathbf{y} - X\hat{\beta}_\theta)^t\Sigma^{-1}(\mathbf{y} - X\hat{\beta}_\theta) \\ &\quad -\frac{1}{2}\log|X^t\Sigma^{-1}X/(2\pi)|. \end{aligned} \tag{7}$$

and

$$\begin{aligned} l_1''(\beta,\theta) &= -\frac{1}{2}\log|X^t\Sigma^{-1}X| \\ &\quad -\frac{1}{2}(\mathbf{y} - X\beta)^t\Sigma^{-1}X(X^t\Sigma^{-1}X)^{-1}X\Sigma^{-1}(\mathbf{y} - X\beta). \end{aligned} \tag{8}$$

The REML estimate of θ is obtained by maximizing $P_\beta(l_1;\theta)$. And the estimate of β is obtained by maximizing l_1'', which is given by

$$\hat{\beta} = (X^t\hat{\Sigma}^{-1}X)^{-1}X^t\hat{\Sigma}^{-1}\mathbf{y}.$$

with fixed $\hat{\theta}$.

In order to obtain REML estimate under missingness, we can re-write function (3) as

$$\begin{aligned} Q(\gamma) &= E\{l_1(\beta,\theta) + l_2(\phi)|\,\mathbf{y_{obs}},\mathbf{r}\} \\ &= E\{P_\beta(l_1;\theta) + l_1''(\beta,\theta) + l_2(\phi)|\,\mathbf{y_{obs}},\mathbf{r}\}. \end{aligned} \tag{9}$$

and further write function (9) as

$$Q(\gamma|\gamma) = Q_1'(\beta,\theta|\gamma) + Q_1''(\beta,\theta|\gamma)' + Q_2(\phi|\gamma).$$

where

$$\begin{aligned} Q'_1(\beta,\theta|\gamma) &= E\{P_\beta(l_1;\theta) \mid \mathbf{y_{obs}},\mathbf{r};\gamma\}. \\ Q''_1(\beta,\theta|\gamma) &= E\{l''_1(\beta,\theta) \mid \mathbf{y_{obs}},\mathbf{r},\gamma\}. \\ Q_2(\phi|\gamma) &= E\{l_2(\phi) \mid \mathbf{y_{obs}},\mathbf{r};\gamma\}. \end{aligned}$$

The imputed Q functions is given by

$$\begin{aligned} Q'^*_1(\theta|\gamma) &= -\frac{1}{2}\log|2\pi\Sigma| - \frac{1}{2}\log|X^t\Sigma^{-1}X/(2\pi)| \\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)}(\gamma)\big((\mathbf{y_i^{*(k)}} - X_i\hat{\beta}_\theta)^t V_i^{-1}(\mathbf{y_i^{*(k)}} - X_i\hat{\beta}_\theta)\big). \end{aligned} \tag{10}$$

$$\begin{aligned} Q''^*_1(\beta,\theta|\gamma) &= -\frac{1}{2}\log|X^t\Sigma^{-1}X| - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)}(\gamma) \\ & (\mathbf{y_i^{*(k)}} - X_i\beta)^t V_i^{-1}X_i(X_i^tV_i^{-1}X_i)^{-1}X_iV_i^{-1}(\mathbf{y_i^{*(k)}} - X_i\beta). \end{aligned} \tag{11}$$

where the weights $w_i^{*(k)}(\gamma)$ are given by (5). The REML can be obtained by the EM-type algorithm:

$$\hat{\theta}^{(t+1)} \leftarrow argmax Q'^*_1(\theta|\gamma^{(t)})$$

$$\hat{\beta}^{(t)} \leftarrow argmax Q''^*_1(\beta,\hat{\theta}^{(t+1)}|\gamma^{(t)})$$

i.e. $\hat{\beta}^{(t)} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)}(\gamma^{(t)})(X_i^t(\hat{V}_i^{(t+1)})^{-1}X_i)^{-1}X_i^t(\hat{V}_i^{(t+1)})^{-1}\mathbf{y_i^{*(k)}}$.

$$\hat{\phi}^{(t)} \leftarrow argmax Q_2^*(\phi|\hat{\gamma}^{(t)})$$

3.3 Estimation of parameters not in the model:

In survey sampling, we are often interested in estimating parameters other than the parameters in the model, say η, which can be written as a solution to

$$E\{U(Y,\mathbf{b};\eta)\} = 0. \tag{12}$$

For example, if we are interested in the population mean, then $U(Y,\mathbf{b};\eta) = n^{-1}\sum_{i=1}^{n} y_i - \eta$.

Under complete response, a consistent estimator of η can be obtained by solving

$$\hat{U}(\eta) \equiv n^{-1}\sum_{i=1}^{n} U(\mathbf{y_i}, b_i;\eta) = 0. \tag{13}$$

for η. Under non-response, we can obtain a fractionally imputed estimating equation

$$\bar{U}^*(\eta) \equiv n^{-1}\sum_{i=1}^{n}\sum_{k=1}^{M}\{w_i^{*(k)}U(\mathbf{y_i^{*(k)}}, b_i;\eta)\} = 0. \tag{14}$$

where $w_i^{*(k)} = \lim_{t\to\infty} w_{i(t)}^{*(k)}$ and $w_{i(t)}^{*(k)}$ is defined in (5). Thus, the final fractional weights $w_i^{*(k)}$ are computed by the MLE (or REML) of γ, denoted by $\hat{\gamma}$, instead of the t^{th} EM estimate of γ in (5). By the law of large numbers

$$p\lim_{M\to\infty}\sum_{k=1}^{M} w_i^{*(k)}U(\mathbf{y_i^{*(k)}},, b_i;\eta) = E\{U(\mathbf{Y_i}, b_i;\eta)|\mathbf{r_i},\hat{\gamma}\}.$$

and $\bar{U}^*(\eta)$ converges to $\bar{U}(\eta|\hat{\gamma}) = E\{U(\mathbf{Y}, \mathbf{b}; \eta)|\mathbf{y_{obs}}, \mathbf{r}; \hat{\gamma}\}$ for sufficiently large M almost surely. The resulting estimator $\hat{\eta}^*$ obtained from (14) is asymptotically consistent and efficient.

4. Variance estimation

Since β and θ are information orthogonal, we can use Louis's formula to construct the confidence intervals for β.

$$I_{obs}(\beta) = -\sum_{i=1}^{n} E\{\dot{S}(\beta; y_i)|y_{i,obs}\} - \sum_{i=1}^{n} V\{S(\beta; y_i)|y_{i,obs}\} = [V(\beta)]^{-1}. \tag{15}$$

which can be approximated by

$$-\sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)} \dot{S}(\hat{\beta}; y_i^{*(k)}) - \sum_{i=1}^{n}\sum_{k=1}^{M} w_i^{*(k)} \{S(\hat{\beta}; y_i^{*(k)}) - \bar{S}_i(\hat{\beta})\}^{\otimes 2}. \tag{16}$$

where $S(\beta; y) = \partial \log f(y; \beta)/\partial\beta = X^t V^{-1}(y - X\beta)$, $\dot{S}(\beta; y) = \partial S(\beta; y)/\partial\beta = -X^t V^{-1} X$ and $\bar{S}_i(\beta) = \sum_{k=1}^{M} w_i^{*(k)} S(\beta; y_i^{*(k)})$.

For variance estimation of $\hat{\eta}$, based on Taylor linearization, we can write $\bar{U}(\eta|\hat{\gamma}) \approx \bar{U}(\eta_0|\gamma_0) + K'\bar{S}(\gamma_0)$, where K is defined as

$$K = -[E\{\partial\bar{S}(\gamma_0)/\partial\gamma\}]^{-1} E\{S_{mis}(\gamma_0)U(\eta_0)\}.$$

If we write

$$\bar{U}(\eta|\gamma) + K'\bar{S}(\gamma) = n^{-1}\sum_{i=1}^{n}\{\bar{\mathbf{u}}_i(\eta|\gamma) + K'\bar{\mathbf{s}}_i(\gamma)\} = n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{u}}_i.$$

the plug-in estimator of $Var(\sum_{i=1}^{n}\tilde{\mathbf{u}}_i)$ is $\sum_{i=1}^{n}(\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}})(\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}})'$, where $\hat{\mathbf{u}}_i = \bar{\mathbf{u}}_i(\hat{\eta}; \hat{\gamma}) + \hat{K}'\bar{\mathbf{s}}_i(\hat{\gamma})$. The terms $\bar{\mathbf{u}}_i(\hat{\eta}; \hat{\gamma})$ and $\bar{\mathbf{s}}_i(\hat{\gamma})$ can be computed from fractional imputation with fractional weights.

5. Simulation study

To test our theory, we performed a limited simulation study. In the simulation study, $B = 2000$ Monte Carlo samples of sizes $n \times m = 10 \times 15 = 150$ were generated from $b_i \sim N(0, \tau^2)$, $e_{ij} \sim N(0, \sigma^2)$, $x_{ij} = j/m$ and $y_{ij} = \beta_0 + \beta_1 x_{ij} + b_i + e_{ij}$ with $\beta_0 = 2, \beta_1 = 1, \sigma^2 = 0.1, \tau^2 = 0.5$ and the response indicator variable r_{ij} for missing is distributed as Bernoulli(π_{ij}) where $logit(\pi_{ij}) = \phi_0 + \phi_1 b_i$ with $\phi_0 = 0, \phi_1 = 1$. Under this model setup, the average response rate is about 50%. The following parameters are computed.

1. β_1, τ^2, σ^2: slope, variance components in the linear mixed effect model
2. μ_y: the marginal mean of y.
3. Proportion: $Pr(Y < 2)$.

For each parameter, compute the following estimators:

1. Complete sample estimator,
2. Incomplete sample estimator,

3. Parametric fractional imputation (PFI) for ML estimation with imputed sample size of M=100,

4. PFI with adjusted profile likelihood estimation with imputed sample size of M=100.

Table (1) presents Monte Carlo mean, variance and standardized variances of the point estimators. The incomplete sample estimators are biased for the mean type of the parameters, as expected. From the response model, individuals with large b_i values are likely to response; whereas individuals with small b_i values are likely to not response. Thus the estimate of the population mean will tend to be larger than the true mean (in the simulation study, we know the true mean is 2.54) and the proportion of $y < 2$ will tend to be smaller than the true probability (the true probability is 0.26). On the other hand, the proposed estimators are essentially unbiased in estimating the mean type of parameters. Imputation can largely reduce non-response bias. For estimating variance component τ^2, the imputed ML estimator is biased downward; however the imputed APL estimator can correct the bias and thus is essentially unbiased for estimating the variance component. The imputation method works well for estimating the variance parameters after incorporating the adjusted profile likelihood idea. PFI (either MLE or APL) is efficient, which can be seen from the Std Var column in Table (1) , for μ_y, $Pr(y < 2)$, and τ^2, the variance for PFI is even smaller.

Table (2) presents the Monte Carlo relative bias and the t-statistics of the variance estimators. Relative biases of the variance estimators were computed by dividing the Monte Carlo bias of the variance estimator by the Monte Carlo variance of the point estimator. The t-statistics are constructed to test the significance of the bias of the variance estimators. A justification of the t-statistics is given in Appendix D of Kim (2004). The variance estimators for PFI are nearly unbiased for the parameters that we are interested in.

6. Discussion Remark

Parametric fraction imputation is proposed as a general tool for estimation with missing data. If the parametric fractional imputation is used to construct the score function, the solution to the imputed score equation is very close to the maximum likelihood estimator for the parameters in the model. The imputation method is very flexible by easily incorporating the restricted maximum likelihood idea or adjusted profile likelihood idea. The variance estimator can be obtained from a Taylor linearization.

REFERENCES

Barndorff-Nielsen, O.E. (1986), Inference on full and partial parameters based on the standardized signed log likelihood ratio, *Biometrika* 73, 307-22

Booth, J. G. Hobert, J. P. (1999). Maximizing generalized linear models with an automated Monte Carlo EM algorithm, *Journal of the Royal Statistical Society, Series B*, 61, 625-85.

Chan, J. S. K., and Kuk, A. Y. C. (1997). Maximum Likelihood Estimation for Probit-linear Mixed Models with Correlated Random Effects, *Biometrics* 53, 86-7.

Cox,D.R. and Reid, N (1987). Parameter orthogonality and approximate conditional inference (with discussion), *Journal of the Royal Statistical Society, Series B*, 49, 1-39.

Diggle, P.J. and Kenward, M.G. (1994). Informative drop-out in longitudinal analysis, *Applied Statistics*, 43, 49-93.

Follmann, D.A. and Wu, M.C. (1995). "An Approximate Generalized Linear Model with Random Effects for Informative Missing Data," *Biometrics* 51, 151-168.

G.C.G. Wei, M.A. Tanner. (1990). A Monte Carlo implementation of the EM algorithm and the poor man data augmentation algorithm, *Journal of the American Statistical Association*, 85, 699-04

J.G. Ibrahim, S.R. Lipsitz, M. Chen. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable, *Journal of the Royal Statistical Society, Series B*, 61 , 173-90

Kim, J.K. (2011). Parametric fractional imputation for missing data analysis, *Biometrika*, 98, 119-132.

Little, R.J.A. (1995). Modeling the drop-out mechanism in repeated-measures studies, *Journal of the American Statistical Association*, 90, 1112-1121.

McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed Models, *Journal of the American Statistical Association*. 42 (437) 162-70.

Patterson, H. D. and Thoponson, R. (1971). Recovery of inter-Block Information when Block sizes are unequal, *Biometrka*, 58, 545-554.

Troxel, A.B., Harrington, D.P. and Lipsitz, S.R. (1998). Analysis of longitudinal measurements with nonignorable non-monotone missing values, *Applied Statistics*, 47, 425-438.

Greenlees, J.S., K.D. Zieschang. (1982). Imputation of missing values when the probability of response depends on the variable being imputed, *Journal of the American Statistical Association*, 77, 251-61.

Table 1: Mean, variance and standardized variance of the point estimators, based on 2000 Monte Carlo samples.

Parameter	Method	Mean	Var	Std Var
$\beta_1 = 1$	Complete	1.00	0.00832	100
	Incomplete	1.00	0.01822	219
	PFI(MLE)	1.00	0.01751	211
	PFI(APL)	1.00	0.01439	173
$\mu_y = 2.54$	Complete	2.54	0.04900	100
	Incomplete	2.74	0.05344	109
	PFI(MLE)	2.54	0.04696	96
	PFI(APL)	2.54	0.03526	72
$Pr(y < 2) = 0.26$	Complete	0.26	0.00969	100
	Incomplete	0.17	0.00651	67
	PFI(MLE)	0.26	0.00942	97
	PFI(APL)	0.26	0.00725	75
$\tau^2 = 0.5$	Complete	0.50	0.06000	100
	Incomplete	0.50	0.06049	101
	PFI(MLE)	0.48	0.04970	83
	PFI(APL)	0.50	0.05873	98
$\sigma^2 = 0.1$	Complete	0.10	0.00014	100
	Incomplete	0.10	0.00032	222
	PFI(MLE)	0.10	0.00034	241
	PFI(APL)	0.10	0.00027	191

Table 2: Monte Carlo relative biases and t-statistics of the variance estimator for the imputation, based on 1000 Monte Carlo samples.

Parameter	Method	R.B. (%)	t-statistics
β_1	PFI(APL)	0.0497	1.0687
μ_y	PFI(APL)	0.2497	0.0174
$Pr(y < 2)$	PFI(APL)	0.2101	0.7723