

Parametric fractional imputation for mixed models with nonignorable missing data

SHU YANG*, JAE-KWANG KIM AND ZHENGYUAN ZHU

Inference in the presence of non-ignorable missing data is a widely encountered and difficult problem in statistics. Imputation is often used to facilitate parameter estimation, which allows one to use the complete sample estimators on the imputed data set. We develop a parametric fractional imputation (PFI) method proposed by Kim (2011), which simplifies the computation associated with the EM algorithm for maximum likelihood estimation with missing data. We first consider the problem of parameter estimation for linear mixed models with non-ignorable missing values, which assumes that missingness depends on the missing values only through the random effects, leading to shared parameter models (Follmann and Wu, 1995). 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 limited 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. Variance estimation is also discussed. We next extend PFI to generalized linear mixed model and the flexibility of this method is illustrated by analyzing the infamous salamander mating data (McCullagh and Nelder, 1989).

KEYWORDS AND PHRASES: EM algorithm, Generalized linear mixed model, Longitudinal data, Nonignorable missingness, Random effect, Restricted maximum likelihood, Salamander data.

1. INTRODUCTION

Mixed models are the statistical models containing both fixed effects and random effects. These models are particularly useful in settings where repeated measurements are made on the same statistical units, or where measurements are made on the clustered elements.

However, missing data frequently occurs in mixed models 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 nonignorable. To model the non-ignorable missing mechanism we considered the selection model (Diggle and Kenward, 1994) approach. Specially, we consider a special case of the selection model, where missingness depends only on the random effects, which yields the so-called shared parameter models, considered by Wu and Carroll (1988), Follmann and Wu (1995) and Ten Have et al. (1998). Wu and Carroll (1988) considered a linear mixed effects model and a discrete-time survival model for the drop-out process that share a random effect structure. Follmann and Wu (1995) considered a conditional model to approximate the shared parameter binary response model conditional on missing data patterns. Ten Have et al. (1998) proposed mixed effects logistic regression models for longitudinal binary response data with informative drop-out.

To carry out the likelihood-based inference under the nonignorable missing, we may 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, 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. First, it facilitates the parameter estimation by simply applying the complete-sample estimators to the imputed data set. Second, it ensures different analyses are consistent with one another. Also, proper choice of imputation method often reduces the non-response bias.

Integration approximated by imputation under nonignorable missing was considered by many authors. Greenlees et al. (1982) considered the normal-theory linear regression model using a version of EM algorithm. Ibrahim et al. (1999) considered continuous 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

*Corresponding author.

likelihood estimates for the probit normal model for binary data. McCulloch (1997) proposed a Monte Carlo Newton-Raphson algorithm in the maximum likelihood algorithm for generalized linear mixed models. For Monte Carlo EM algorithm, in each E-step, the imputed values are regenerated and thus the computation can be quite heavy. Also the convergence of Monte Carlo sequence of the estimators is not guaranteed for fixed Monte Carlo sample size (Booth and 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 and for the generalized linear mixed model. The main idea in PFI is to produce a complete data set by imputation and each imputed value is associated with a fractional weight, by which the observed likelihood can be approximated by the weighted average of the imputed data likelihood. The resulting estimator is close to the maximum likelihood estimator and has very nice asymptotic properties, such as efficiency and asymptotic normality.

PFI can also be extended to generalized linear mixed model. The flexibility of this method is illustrated by analyzing the infamous salamander mating data (McCullagh and Nelder, 1989). The data is challenging since the response variable is binary and the experimental design is crossed which causes serious limitations due to the intensive computations to approximate the intractable joint distribution.

In Section 2, we introduce linear mixed model with nonignorable missing and develop the PFI method for this model. In Section 3, we discuss incorporation of adjusted profile likelihood estimation in PFI to reduce the bias in estimating variance components. In Section 4, results from a limited simulation study are presented. Section 5 demonstrates PFI in a generalized linear mixed model setting by analyzing the salamander mating data. We conclude with discussion in Section 6.

2. LINEAR MIXED MODEL WITH NONIGNORABLE MISSING VALUES

2.1 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

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

where i indexes individual, j indexes the repeated measurement within each individual, b_i 's are i.i.d. from $N(0, \tau^2)$ specifying the unobserved individual effects and e_{ij} are i.i.d. from $N(0, \sigma^2)$, which are the measurement errors within individuals.

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 $\mathbf{y}_{\text{obs},i}$, $\mathbf{y}_{\text{mis},i}$ respectively, so $\mathbf{y}_i = (\mathbf{y}_{\text{obs},i}, \mathbf{y}_{\text{mis},i})$. Let $\mathbf{r}_i = (r_{i1}, \dots, r_{im})'$ be vector of indicators of response status, so $r_{ij} = 1$ if y_{ij} is observed, otherwise, $r_{ij} = 0$. As a motivating example, consider a disease longitudinal study. When patients experience an increase in pain, they might decide not to show up at some of the scheduled visits for disease evaluation. In the above cases, if we simply ignore the missingness process and use the standard procedure to analyze the data, the inference will be seriously biased. To take care of such an informative nonresponse, joint modeling of disease measurement and missing process is necessary.

To model the missing mechanism, we consider the selection model

$$(2) \quad f(\mathbf{y}_i, \mathbf{r}_i, b_i) = f(\mathbf{y}_i|b_i)f(\mathbf{r}_i|\mathbf{y}_i, b_i)f(b_i).$$

Furthermore, in the selection model, we assume nonignorable dropout in longitudinal data (Molenberghs and Kenward, 2009) where 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 nonignorable 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. Such assumption on missingness depends on unobserved individual effect, and may be reasonable if we can assume that r_{i1}, \dots, r_{im} are identically distributed within subject i . Shared parameter models are convenient and also intuitively appealing in ways of joint modeling the disease measurement and missingness process, where we assume a set of random effects to introduce interdependence.

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

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

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

$$l_{com}(\gamma) = \log f(\mathbf{y}_i, b_i, \mathbf{r}_i|\gamma)$$

$$\begin{aligned}
&= \sum_{i=1}^n \log \left[\left\{ \prod_{j=1}^m f(y_{ij}|\beta, \sigma^2, b_i) f(r_{ij}|\phi, b_i) \right\} f(b_i|\tau^2) \right] \\
&= \sum_{i=1}^n \sum_{j=1}^m \log f(y_{ij}|\beta, \sigma^2, b_i) + \sum_{i=1}^n \sum_{j=1}^m \log f(r_{ij}|\phi, b_i) \\
&\quad + \sum_{i=1}^n \log f(b_i|\tau^2) \\
&= l_1(\beta, \sigma^2) + l_2(\phi) + l_3(\tau^2).
\end{aligned}$$

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 $(\mathbf{y}_{\text{obs}}, \mathbf{r})$, the observed density can be obtained by integrating out the unobserved random effects and missing values of the joint complete density

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

Then the observed log likelihood function of γ is specified by

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

where

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

As we can see, since y_{ij} depends on b_i and r_{ij} depends on b_i as well, and so (β, σ^2) , ϕ , and τ^2 cannot be separated in $l_{\text{obs}}(\gamma)$ as we do in $l_{\text{com}}(\gamma)$. Thus parameters γ need to be estimated simultaneously.

Maximum likelihood estimator $\hat{\gamma}$ can be obtained by maximizing $l_{\text{obs}}(\gamma)$. Instead of maximizing $l_{\text{obs}}(\gamma)$, one can also obtain the MLE by maximizing

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

Computing the MLE using (3) will be discussed in the next section.

2.2 Parametric fractional imputation maximum likelihood estimation

We develop an EM algorithm by the PFI method of Kim (2011) to linear mixed models with nonignorable missing. To apply the 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

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

The MLE can be obtained by the EM-type algorithm

$$(4) \quad \hat{\gamma}_{(t+1)} \leftarrow \text{argmax } Q(\gamma|\hat{\gamma}_{(t)}).$$

The Monte Carlo EM method (MCEM) computes $Q(\gamma|\hat{\gamma}_{(t)})$ by regenerating the imputed values of size M for each EM iteration and assigning equal weight $1/M$ to each imputed value. The computation is cumbersome because it often requires an iterative algorithm such as Metropolis-Hastings algorithm for each EM iteration. There is also no guarantee for the MCEM sequence convergence of fixed M . Alternatively, the 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 update the importance weights only by using the updated 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 nonignorable missing in linear mixed model setup. The M imputed values $b_i^{*(1)}, \dots, b_i^{*(M)} \sim h_1(\cdot)$, $y_{ij}^{*(k)} \sim h_2(\cdot|x_{ij}, b_i^{*(k)})$ are generated from initial densities $h_1(b_i)$ and $h_2(y_{ij}|x_{ij}, b_i)$ with the same support as $f(y_{ij})$. The choice of $h_1(b_i)$ is somewhat arbitrary, but a t-distribution with small degrees of freedom seems to work well in practice. Given the current parameter estimates $\hat{\gamma}_{(t)}$ and the M imputed values $b_i^{*(1)}, \dots, b_i^{*(M)}$ and $y_{ij}^{*(1)}, \dots, y_{ij}^{*(M)}$ generated above, the joint density of $(\mathbf{y}_{i, \text{obs}}, \mathbf{y}_{i, \text{mis}}^{*(\mathbf{k})}, b_i^{*(\mathbf{k})})$ for each individual, where $\mathbf{y}_{i, \text{mis}}^{*(\mathbf{k})}$ is a vector of imputed values for $\mathbf{y}_{i, \text{mis}}$ is

$$\begin{aligned}
(5) \quad &f_i^{*(\mathbf{k})}(\gamma) \\
&= \prod_{j=1}^m \{f(y_{ij}^{*(\mathbf{k})}|\beta, \sigma^2, b_i^{*(\mathbf{k})}) f(r_{ij}|\phi, b_i^{*(\mathbf{k})})\} f(b_i^{*(\mathbf{k})}|\tau^2).
\end{aligned}$$

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

(6)

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

The Monte Carlo approximate of $Q(\gamma|\hat{\gamma}^{(t)})$ in (4) is

(7)

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

where

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

and

$$\begin{aligned} Q_3^*(\tau^2|\gamma^{(t)}) &= \sum_{i=1}^n \sum_{k=1}^M w_i^{*(k)}(\gamma^{(t)}) \left\{ -\frac{1}{2} \log(2\pi\tau^2) - \frac{1}{2\tau^2} (b_i^{*(k)})^2 \right\}. \end{aligned}$$

Thus, the PFI method computes the E-step of the EM algorithm using fractional weights in (6). 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 $Q_1^*(\beta, \sigma^2|\gamma^{(t)})$, $Q_2^*(\phi|\gamma^{(t)})$, and $Q_3^*(\tau^2|\gamma^{(t)})$ for $\gamma = (\beta, \sigma^2, \phi, \tau^2)$.

Maximizing Q_1^* , Q_2^* , Q_3^* can be easily implemented by incorporating the fractional weights in the existing software, such as SAS or R. The EM sequence $\{\hat{\gamma}^{(t)}; t = 1, 2, \dots\}$ 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 sufficiently large M.

Now consider estimating general parameters, say η , which can be written as a solution to

$$(8) \quad E\{U(Y, \mathbf{b}; \eta)\} = 0.$$

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

Under complete response, a consistent estimator of η can be obtained by solving $\hat{U}(\eta) \equiv n^{-1} \sum_{i=1}^n U(\mathbf{y}_i, b_i; \eta) = 0$, for η . Under non-response, we can obtain a fractionally imputed estimating equation

$$(9) \quad \bar{U}^*(\eta) \equiv n^{-1} \sum_{i=1}^n \sum_{k=1}^M \{w_i^{*(k)} U(\mathbf{y}_i^{*(k)}, b_i; \eta)\} = 0,$$

where $w_i^{*(k)} = \lim_{t \rightarrow \infty} w_{i(t)}^{*(k)}$ and $w_{i(t)}^{*(k)}$ is defined in (6).

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 (6). By the law of large numbers

$$p \lim_{M \rightarrow \infty} \sum_{k=1}^M w_i^{*(k)} U(\mathbf{y}_i^{*(k)}, b_i; \eta) = E\{U(\mathbf{Y}_i, b_i; \eta)|\mathbf{r}_i, \hat{\gamma}\}$$

and $\bar{U}^*(\eta)$ converges to $\bar{U}(\eta|\hat{\gamma}) = E\{U(\mathbf{Y}, \mathbf{b}; \eta)|\mathbf{y}_{\text{obs}}, \mathbf{r}; \hat{\gamma}\}$ for sufficiently large M almost surely. The resulting estimator $\hat{\eta}^*$ obtained from (9) is asymptotically consistent and efficient.

3. ADJUSTED PROFILE LIKELIHOOD FOR BIAS CORRECTION

We now consider approaches of reducing 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 for the inconsistency of the profile likelihood which automatically make “degrees of freedom” adjustments in normal theory cases. 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, the adjusted profile likelihood matches exactly the restricted maximum likelihood (REML) (Patterson and Thompson, 1971) using the marginal distribution of the error term $\mathbf{y} - X\hat{\beta}_\theta$, where $\theta = (\sigma^2, \tau^2)$. To see this, the data can be divided into two independent parts, the error term $\mathbf{y} - X\hat{\beta}_\theta = S\mathbf{y}$ and $Q\mathbf{y}$, $S = I - X(X^t \Sigma_\theta^{-1} X)^{-1} X^t \Sigma_\theta^{-1}$ and $Q = X^t \Sigma_\theta^{-1}$. The likelihood l_1 can be separated into two parts,

$$l_1(\beta, \theta) = P_\beta(l_1; \theta) + l_1''(\beta, \theta),$$

where

$$P_{\beta}(l_1; \theta) = l_p(\theta) - \frac{1}{2} \log |X^t \Sigma_{\theta}^{-1} X / (2\pi)|$$

with $l_p(\theta) = -\frac{1}{2} \log |2\pi \Sigma_{\theta}| - \frac{1}{2} (\mathbf{y} - X\hat{\beta}_{\theta})^t \Sigma_{\theta}^{-1} (\mathbf{y} - X\hat{\beta}_{\theta})$ the profile likelihood function and $\frac{1}{2} \log |X^t \Sigma_{\theta}^{-1} X / (2\pi)|$ the adjustment, and

$$(10) \quad \begin{aligned} l_1''(\beta, \theta) &= -\frac{1}{2} \log |X^t \Sigma_{\theta}^{-1} X| - \frac{1}{2} (\mathbf{y} - X\beta)^t \Sigma_{\theta}^{-1} X (X^t \Sigma_{\theta}^{-1} X)^{-1} \\ &\quad \times X \Sigma_{\theta}^{-1} (\mathbf{y} - X\beta). \end{aligned}$$

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

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

where $\hat{\Sigma} = \Sigma_{\hat{\theta}}$ with fixed $\hat{\theta}$.

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

$$(11) \quad \begin{aligned} Q(\gamma) &= E\{l_1(\beta, \theta) + l_2(\phi) \mid \mathbf{y}_{\text{obs}}, \mathbf{r}\} \\ &= E\{P_{\beta}(l_1; \theta) + l_1''(\beta, \theta) + l_2(\phi) \mid \mathbf{y}_{\text{obs}}, \mathbf{r}\} \end{aligned}$$

and further write function (11) as

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

where

$$\begin{aligned} Q_1'(\beta, \theta|\gamma) &= E\{P_{\beta}(l_1; \theta) \mid \mathbf{y}_{\text{obs}}, \mathbf{r}; \gamma\}, \\ Q_1''(\beta, \theta|\gamma) &= E\{l_1''(\beta, \theta) \mid \mathbf{y}_{\text{obs}}, \mathbf{r}; \gamma\}, \\ Q_2(\phi|\gamma) &= E\{l_2(\phi) \mid \mathbf{y}_{\text{obs}}, \mathbf{r}; \gamma\}. \end{aligned}$$

The imputed Q functions are given by

$$(12) \quad \begin{aligned} Q_1^*(\theta|\gamma) &= -\frac{1}{2} \log |2\pi \Sigma_{\theta}| - \frac{1}{2} \log |X^t \Sigma_{\theta}^{-1} X / (2\pi)| \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^M w_i^{*(k)}(\gamma) ((\mathbf{y}_i^{*(k)} - X_i \hat{\beta}_{\theta})^t \\ &\quad \times V_i^{-1} (\mathbf{y}_i^{*(k)} - X_i \hat{\beta}_{\theta})) \end{aligned}$$

and

$$(13) \quad \begin{aligned} Q_1^{''*}(\beta, \theta|\gamma) &= -\frac{1}{2} \log |X^t \Sigma_{\theta}^{-1} X| \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^M w_i^{*(k)}(\gamma) (\mathbf{y}_i^{*(k)} - X_i \beta)^t V_i^{-1} \\ &\quad \times X_i (X_i^t V_i^{-1} X_i)^{-1} X_i V_i^{-1} (\mathbf{y}_i^{*(k)} - X_i \beta), \end{aligned}$$

where the weights $w_i^{*(k)}(\gamma)$ are given by (6). The REML can be obtained by the EM-type algorithm:

$$\begin{aligned} \hat{\theta}^{(t+1)} &\leftarrow \operatorname{argmax} Q_1^*(\theta|\gamma^{(t)}) \\ \hat{\beta}^{(t+1)} &\leftarrow \operatorname{argmax} Q_1^{''*}(\beta, \hat{\theta}^{(t+1)}|\gamma^{(t)}). \end{aligned}$$

That is,

$$\begin{aligned} \hat{\beta}^{(t+1)} &= \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^M w_i^{*(k)}(\gamma^{(t)}) \\ &\quad (X_i^t (\hat{V}_i^{(t+1)})^{-1} X_i)^{-1} X_i^t (\hat{V}_i^{(t+1)})^{-1} \mathbf{y}_i^{*(k)} \end{aligned}$$

and

$$\hat{\phi}^{(t+1)} \leftarrow \operatorname{argmax} Q_2^*(\phi|\hat{\gamma}^{(t)}).$$

Thus, the EM algorithm using PFI method is directly applicable to REML by replacing the original likelihood with the adjusted profile likelihood.

4. SIMULATION STUDY OF LINEAR MIXED MODEL WITH NONIGNORABLE MISSING DATA

To test our theory, we performed a limited simulation study. In the simulation study, $B = 2,000$ Monte Carlo samples of sizes $n \times m = 10 \times 15 = 150$ were generated independently 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 $\beta_0 = 2, \beta_1 = 1, \sigma^2 = 0.2, \tau^2 = 0.2$. The response indicator variable r_{ij} for missing is distributed as Bernoulli(π_{ij}) where $\operatorname{logit}(\pi_{ij}) = \phi_0 + \phi_1 b_i$ with $\phi_0 = 0.5, \phi_1 = 1$. Note that this response mechanism follows the shared parameter model. Under this model setup, the average response rate is about 60%. The following parameters are computed.

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

For each parameter, we 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 $M = 50$,
4. PFI with adjusted profile likelihood estimation with imputed sample size of $M = 50$.

Table 1 presents Monte Carlo mean, variance and standardized variances (which is the ratio of variance and variance of complete sample estimator and times 100) 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

Table 1. Mean, variance and standardized variance of the point estimators, based on 2,000 Monte Carlo samples

Parameter	Method	Mean	Var	Std Var
$\beta_1 = 1$	Complete	1.00	0.01601	100
	Incomplete	1.00	0.02791	174
	PFI(MLE)	1.00	0.02735	171
	PFI(APL)	1.00	0.02613	163
$\mu_y = 2.53$	Complete	2.53	0.02157	100
	Incomplete	2.60	0.02281	106
	PFI(MLE)	2.53	0.02085	97
	PFI(APL)	2.53	0.02018	94
$Pr(y < 2) = 0.22$	Complete	0.22	0.004994	100
	Incomplete	0.19	0.004657	93
	PFI(MLE)	0.22	0.004837	97
	PFI(APL)	0.22	0.004662	93
$\tau^2 = 0.2$	Complete	0.19	0.009033	100
	Incomplete	0.19	0.010063	111
	PFI(MLE)	0.18	0.007450	82
	PFI(APL)	0.19	0.008795	97
$\sigma^2 = 0.2$	Complete	0.20	0.0005717	100
	Incomplete	0.20	0.0009732	170
	PFI(MLE)	0.20	0.0009501	166
	PFI(APL)	0.20	0.0009019	158

to respond; whereas individuals with small b_i values are likely to not respond. Thus the observed mean will tend to be larger than the true mean (in the simulation study, we know the true mean is 2.53) and the observed proportion of $y < 2$ will tend to be smaller than the true probability (the true probability is 0.22). On the other hand, the proposed PFI 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 for APL estimator (Appendix A.1). Variance estimators of the PFI estimators are computed using the Louis formula and the linearization method discussed in Appendix. 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 considered.

Table 2. Monte Carlo relative biases and t-statistics of the variance estimator for the imputation, based on 2,000 Monte Carlo samples

Parameter	Method	R.B. (%)	t-statistics
β_1	PFI(reml)	3.12	1.03
μ_y	PFI(reml)	2.40	0.87
$Pr(y < 2)$	PFI(reml)	1.31	0.43

5. GENERALIZED LINEAR MIXED MODEL

Parametric fractional imputation can be extended to generalized linear mixed models. Here we consider the data set on salamander mating, which could be modeled as generalized linear mixed model.

5.1 Data description

The salamander data came from the experiment conducted by S. Arnold and P. Verrell (1989), aimed to study the extent to which mountain dusky salamanders from different populations would interbreed. The data given here refer to two populations called Rough Butt (R) and Whiteside (W). Forty animals were used in each of three experiments, one conducted in the summer of 1986 and two in the Fall of the same year. The forty salamanders available in each of the three experiments were comprised of 10 Rough Butt males, 10 Rough Butt females, 10 Whiteside males and 10 Whiteside females. Although there were 400 possible crosses between the females and males in each experiment, only 120 of these were permitted by the design. So totally they observed 360 potential matings. The design of the experiment permits a comparison of the mating probabilities for the four possible crosses: RR, RW, WR and WW.

5.2 Generalized linear mixed model

For the total 360 observations in the data set, we consider models for the observed data conditionally on the actual animals used in the experiment. Denote y_{ij} to be a random variable representing the binary response indicator of a successful mating between the i th female and the j th male for $i, j = 1, 2, \dots, 60$ where only 360 of the (i, j) pairs are relevant (each i corresponds to six j 's). Let u_i^f denote the random effect that the i th female salamander has cross matings in which she is involved, and define u_j^m similarly for the j th male. Let \mathbf{x}_{ij} denote a 4 dimensional vector of covariates indicating the type of cross for the mating pair between female i and male j . We assume that the y_{ij} 's are all conditionally independent, and assume a Binomial regression model for the salamander data set, i.e.,

$$(14) \quad y_{ij} | u_i^f, u_j^m \sim \text{Bernoulli}(\pi_{ij}),$$

and

$$(15) \quad \eta_{ij} = g(\pi_{ij}) = \text{logit}(\pi_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i^f + u_j^m,$$

where $g(\cdot)$ is the link function, and we use the canonical link which is the logit link, $\boldsymbol{\beta} = (\beta_{RR}, \beta_{RW}, \beta_{WR}, \beta_{WW})^T$ is an unknown 4 dimensional regression parameter vector. The parameter vector $\boldsymbol{\beta}$ as fixed effects and u_i^f 's and u_j^m 's as random effects. Assume $u_i^f \sim N(0, \sigma_f^2)$ and $u_j^m \sim N(0, \sigma_m^2)$, so the resulting model has 6 unknown parameters $\beta_{RR}, \beta_{RW}, \beta_{WR}, \beta_{WW}, \sigma_f^2$ and σ_m^2 .

Let \mathbf{y} denote the full data vector, and let $\mathbf{u}^f, \mathbf{u}^m$ be two 60-variate random variables with parametric densities $g_1(\mathbf{u}^f|\sigma_f^2)$ and $g_2(\mathbf{u}^m|\sigma_m^2)$ respectively. The joint distribution of $(\mathbf{y}, \mathbf{u}^f, \mathbf{u}^m)$ is

$$(16) \quad \left(\prod_{i=1}^{60} \prod_{j=i+1}^{i6} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \right) g_1(\mathbf{u}^f|\sigma_f^2) g_2(\mathbf{u}^m|\sigma_m^2),$$

where $\pi_{ij} = g^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i^f + u_j^m) = \frac{\exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i^f + u_j^m)}{1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i^f + u_j^m)}$.

The likelihood function for $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \sigma_f^2, \sigma_m^2)$ is

$$(17) \quad L(\boldsymbol{\gamma}|\mathbf{y}) = \int \int \left(\prod_{i=1}^{60} \prod_{j=i+1}^{i6} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \right) g_1(\mathbf{u}^f|\sigma_f^2) g_2(\mathbf{u}^m|\sigma_m^2) d\mathbf{u}^f d\mathbf{u}^m.$$

The likelihood function $L(\boldsymbol{\gamma}|\mathbf{y})$ involves intractable integrals whose dimension depends on the structure of the random effects $(\mathbf{u}^f, \mathbf{u}^m)$ which is a 120-dimensional vector, so likelihood inference requires numerical evaluation of a high-dimensional integral.

5.3 Fractional imputation

The complete log-likelihood function of $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \sigma_f^2, \sigma_m^2)$ is given by

$$l_{com}(\boldsymbol{\gamma}) = \sum_{i=1}^{60} \sum_{j=i+1}^{i6} \{y_{ij} \log \pi_{ij} + (1 - y_{ij}) \log(1 - \pi_{ij})\} + \log g_1(\mathbf{u}^f|\sigma_f^2) + \log g_2(\mathbf{u}^m|\sigma_m^2).$$

We treat the random effects $(\mathbf{u}^f, \mathbf{u}^m)$ as missing data. The maximum likelihood estimator $\hat{\boldsymbol{\gamma}}$ can be obtained by maximizing

$$Q(\boldsymbol{\gamma}) = E\{l_{com}(\boldsymbol{\gamma}; \mathbf{y}, \mathbf{u}^f, \mathbf{u}^m) | \mathbf{y}\}.$$

In the above expectation, the reference distribution is the conditional distribution $\mathbf{u}^f, \mathbf{u}^m | \mathbf{y}$,

$$\mathbf{u}^f, \mathbf{u}^m | \mathbf{y} \propto \left(\prod_{i=1}^{60} \prod_{j=i+1}^{i6} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \right) g_1(\mathbf{u}^f|\sigma_f^2) g_2(\mathbf{u}^m|\sigma_m^2).$$

We consider a 120-dimensional multivariate Student t importance density (suggested by Booth and Hobert, 1998) with 3 degrees of freedom, whose mean and variance match

Table 3. Salamander Data set (observations = 360)

Method	β_{RR}	β_{RW}	β_{WR}	β_{WW}	σ_f^2	σ_m^2
Pseudo lik	0.78	0.24	-1.48	0.77	0.65	0.58
Imputation	0.97	0.33	-1.81	0.95	1.13	0.89
MLE	1.01	0.31	-1.90	0.99	1.17	1.04
MCEM	1.02	0.32	-1.94	0.99	1.39	1.23
Gibbs	1.03	0.34	-1.98	1.07	1.49	1.37

the mode and curvature of the target distribution $f(\mathbf{u}|\mathbf{y}; \boldsymbol{\gamma})$, $\mathbf{u} = (\mathbf{u}^f, \mathbf{u}^m)$. Write $f(\mathbf{u}|\mathbf{y}; \boldsymbol{\gamma}) = a \exp\{l(\mathbf{u})\}$, where a is the normalizing constant. Let $l^{(i)}(\mathbf{u})$ be the i th derivative of $l(\mathbf{u})$, and $\tilde{\mathbf{u}}$ denote the maximizer of $l(\mathbf{u})$ satisfying the equation $l^{(1)}(\mathbf{u}) = 0$. The Laplace approximations of the mean and variance are $\tilde{\mathbf{u}}$ and $-l^{(2)}(\tilde{\mathbf{u}})$. See Booth and Hobert (1998) for the formula. Denote $\mathbf{u}^{*(1)}, \dots, \mathbf{u}^{*(M)}$ as a random sample from $h(\mathbf{u}|\nu, \mu, \boldsymbol{\Sigma})$, which is a multivariate Student t distribution with $\nu = 3$, $\mu = \tilde{\mathbf{u}}$ and $\boldsymbol{\Sigma} = -l^{(2)}(\tilde{\mathbf{u}})$, the fractional weights are given by

$$w^{*(k)}(\boldsymbol{\gamma}) = \frac{f(\mathbf{y}, \mathbf{u}^{*(k)}|\boldsymbol{\gamma})/h(\mathbf{u}^{*(k)}|\nu, \mu, \boldsymbol{\Sigma})}{\sum_{l=1}^M f(\mathbf{y}, \mathbf{u}^{*(l)}|\boldsymbol{\gamma})/h(\mathbf{u}^{*(l)}|\nu, \mu, \boldsymbol{\Sigma})}.$$

The Monte Carlo approximate of the observed likelihood function is given by

$$Q^*(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)}) = \sum_{k=1}^M w^{*(k)}(\boldsymbol{\gamma}^{(t)}) \log f(\mathbf{y}, \mathbf{u}^{*(k)}|\boldsymbol{\gamma}),$$

which is maximized in each M-step in the EM algorithm to update the parameter estimates $\boldsymbol{\gamma}^{(t)}$ to $\boldsymbol{\gamma}^{(t+1)}$.

Table 3 shows the results from PFI and various estimation methods, including Pseudo likelihood method (Arnold and Verrel), MLE from a modified EM algorithm with Laplace approximation (Steele, 1996), Gibbs sampling method (Karim and Zeger, 1992) and the Monte Carlo EM method (Vaida and Meng, 2005). Ver Hoef et al. (2010) suggested the Pseudo likelihood approach to create a linear mixed pseudo-model where the resulting estimator is called pseudo likelihood estimator. Pseudo-likelihood estimates can be implemented in SAS GLIMMIX procedure. As we can see, the estimates of parameters in the pseudo-likelihood approach are quite different from other approaches which is caused by the lack of efficiency of pseudo-likelihood approximation to the original likelihood when a large dimension of random effects are involved. Steele (1996) suggests replacing the conditional expectation in the E-step with a second order approximation and this modified EM algorithm produced accurate estimates of the fixed effects in generalized linear mixed models. Our imputation method gives estimates close to MLE. The Gibbs sampling approach and the Monte Carlo EM method tend to produce larger estimates than MLE. However, as we discussed previously, both methods involve heavy computation which is not desirable in practice. The PFI sam-

ples are created only once in the beginning of EM algorithm and thus largely reduce the burden of computation. This example shows that statistically efficient estimation is possible without requiring a computationally extensive method.

6. DISCUSSION REMARK

Parametric fraction imputation is proposed as a general tool for estimation with missing clustered 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 applicable to the restricted maximum likelihood method or the adjusted profile likelihood method. The variance estimator can be obtained from a Taylor linearization. PFI can also be easily extended to generalized linear mixed model, which allows statistically efficient estimation without requiring a computationally extensive method and can be more feasible in practice.

ACKNOWLEDGEMENT

We thank an anonymous referee for very helpful comments. The research was partially supported by Cooperative Agreement between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University.

APPENDIX A. APPENDIX SECTION

A.1 Variance estimation

Since β and θ are information orthogonal, we can use Louis's formula to construct the confidence intervals for β (18)

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

which can be approximated by

$$(19) \quad - \sum_{i=1}^n \sum_{k=1}^M w_i^{*(k)} \dot{S}(\hat{\beta}; y_i^{*(k)}) - \sum_{i=1}^n \sum_{k=1}^M w_i^{*(k)} \{S(\hat{\beta}; y_i^{*(k)}) - \bar{S}_i(\hat{\beta})\}^{\otimes 2}$$

where $S(\beta; y) = \partial \log f(y; \beta) / \partial \beta$, $\dot{S}(\beta; y) = \partial S(\beta; y) / \partial \beta$ 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 obtained from $\bar{U}(\eta) = 0$ in (9), 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 $\text{Var}(\sum_{i=1}^n \tilde{\mathbf{u}}_i)$ is $\sum_{i=1}^n (\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}})(\hat{\mathbf{u}}_i - \bar{\hat{\mathbf{u}}})'$, where $\hat{\mathbf{u}}_i = \bar{\mathbf{u}}_i(\hat{\eta}; \hat{\gamma}) + K'\bar{\mathbf{s}}_i(\hat{\gamma})$. 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.

Received 6 November 2012

REFERENCES

- BARNDORFF-NIELSEN, O. E. (1986). Inference on full and partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307–322. [MR0855891](#)
- BOOTH, J. G. and HOBERT, J. P. (1999). Maximizing generalized linear models with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society: Series B* **61** 625–685.
- CHAN, J. S. K. and KUK, A. Y. C. (1997). Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics* **53** 86–87. [MR1450182](#)
- 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. [MR0893334](#)
- 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. [MR1341233](#)
- GREENLEES, J. S. and ZIESCHANG, K. D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association* **77** 251–261.
- IBRAHIM, J. G., LIPSITZ, S. R. and CHEN, M. (1999). Missing covariates in generalized linear models when the missing data mechanism is nonignorable. *Journal of the Royal Statistical Society: Series B* **61** 173–190. [MR1664045](#)
- KARIM, M. R. and ZEGER, S. L. (1992). Generalized linear models with random effects: Salamander mating revisited. *Biometrics* **48** 631–644.
- KIM, J. K. (2004). Finite sample properties of multiple imputation estimators. *The Annals of Statistics* **32** 766–783. [MR2060177](#)
- KIM, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika* **98** 119–132. [MR2804214](#)
- LITTLE, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90** 1112–1121. [MR1354029](#)
- MCCULLOCH, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **42** (437) 162–170. [MR1436105](#)
- MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. G Chapman and Hall, pp. 441–450. [MR0727836](#)
- MOLENBERGHS, G. and KENWARD, M. G. (2007). *Missing Data in Clinical Studies*. Wiley.
- PATTERSON, H. D. and THOMPSON, R. (1971). Recovery of inter-Block Information when Block sizes are unequal. *Biometrika* **58** 545–554. [MR0319325](#)
- STEELE, B. M. (1996). A modified EM algorithm for estimation in generalized mixed models. *Biometrics* **52** 1295–1310. [MR1422080](#)
- TEN HAVE, T., KUNSELMAN, A., PULKSTENIS, E. and LANDIS, J. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* **54** 367–383.

- 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.
- VAIDA, F. and MENG, X. L. (2005). Two slice-EM algorithms for fitting generalized linear mixed models with binary response. *Statistical Modelling* **5** 229–242. [MR2210735](#)
- VER HOEF, J. M., LONDON, J. M. and BOVENG, P. L. (2010). Fast computing of some generalized linear mixed pseudo-models with temporal autocorrelation. *Comput. Stat.* **25** 39–55. [MR2586723](#)
- WEI, G. C. G. and TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man data augmentation algorithm. *Journal of the American Statistical Association* **85** 699–704.
- WU, M. C. and CARROLL, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44** 175–188. [MR0931633](#)
- ZEGER, S. L. and KARIM, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86** 79–86. [MR1137101](#)

Jae-Kwang Kim
Department of Statistics
Iowa State University
1208 Snedecor Hall
Ames, IA, 50011
USA
E-mail address: jkim@iastate.edu

Zhengyuan Zhu
Department of Statistics
Iowa State University
1216 Snedecor Hall
Ames, IA, 50011
USA
E-mail address: zhuz@iastate.edu

Shu Yang
Department of Statistics
Iowa State University
1219 Snedecor Hall
Ames, IA, 50011
USA
E-mail address: shuyang@iastate.edu