Published in final edited form as:

Comput Stat Data Anal. 2014 April ; 72: 77–91. doi:10.1016/j.csda.2013.10.027.

Inference for longitudinal data with nonignorable nonmonotone missing responses

Sanjoy K. Sinha^{a,*}, Amit Kaushal^b, and Wenzhong Xiao^c

^aSchool of Mathematics and Statistics, Carleton University, Ottawa, ON, K1S 5B6, Canada

^bStanford Genome Technology Center, Stanford, CA 94305, USA

^cHarvard Medical School, Boston, MA 02114, USA

Abstract

For the analysis of longitudinal data with nonignorable and nonmonotone missing responses, a full likelihood method often requires intensive computation, especially when there are many follow-up times. The authors propose and explore a Monte Carlo method, based on importance sampling, for approximating the maximum likelihood estimators. The finite-sample properties of the proposed estimators are studied using simulations. An application of the proposed method is also provided using longitudinal data on peptide intensities obtained from a proteomics experiment of trauma patients.

Keywords

False discovery rate; Importance sampling; Incomplete data; Linear mixed model; Longitudinal study; Maximum likelihood; Proteomics experiment

1. Introduction

In many clinical experiments, we often investigate changes in a specific characteristic in the participating individuals being observed repeatedly over time. Longitudinal studies are commonly performed for the investigation of individual changes over time and for exploring the effects of aging and other factors that are likely to influence the change. The repeated measurements from individuals in a longitudinal study are correlated by nature. To model the correlations among observations and also to investigate the individual (or group) effects on the responses, linear and generalized linear mixed models are often used. For an overview of mixed models and their applications to longitudinal data analysis, see Diggle et al. (2002).

We often encounter missing data in longitudinal experiments. Missing data occur whenever one or more of the sequences of measurements from experimental units or individuals are incomplete, in the sense that the desired measurements are not available, are lost due to technical problems, or otherwise not taken. The missingness in the longitudinal data often

depends on the unobserved value of the outcome at a given assessment time, that is, the missing data are often nonignorable. When data are nonignorably missing, it is necessary to model the missing data mechanism for valid statistical inferences.

Analysis of missing data has been considered by many authors in the literature (e.g., (Diggle and Kenward, 1994; Ibrahim et al., 1999, 2001; Molenberghs and Verbeke, 2001; Sinha et al., 2010, 2011; Verbeke and Molenberghs, 2005; Wu et al., 2009; Xie, 2008; Yi and Cook, 2002); and many others). Little (1995) discusses techniques for modeling the data and the missing data mechanism simultaneously, and presents a number of examples to describe likelihood-based inferences via maximum likelihood or Bayesian approaches. Little and Rubin (2002) review methods for analyzing data with various types of missing data mechanisms. Note that for analyzing longitudinal data with nonignorable and nonmonotone missing responses, a full likelihood method often requires intensive computation, especially when there are many follow-up times. Our goal is to find a suitable method that is computationally feasible and is also almost as efficient as the exact maximum likelihood method for analyzing the data.

This research was motivated by a time course proteomics experiment in which longitudinal measurements on peptide intensities were obtained from monocytes of a group of trauma patients being studied at Stanford Genome Technology Center. A monocyte is a type of white blood cell and is a part of the human body's immune system. Trauma patients experience monocyte dysfunctions during the development of multiple organ dysfunction syndrome (MODS). The focus of the analysis is on the expression levels of proteins in monocytes of trauma patients, where the abundance of protein molecules is determined by peptide intensities measured repeatedly over a certain period after the injury to a trauma patient.

Note that peptides are short chains of amino acids, and polymers of amino acids make up protein molecules. Peptide intensities were measured by mass spectrometry (MS), an analytical technique for the determination of the elemental composition of a sample or molecule. Mass spectrometry can be useful for determining what proteins are expressed in cancer cells that are not expressed in healthy cells, possibly leading to further understanding of the disease and to the development of drugs targeting these proteins.

A goal of the aforementioned proteomics experiment is to identify the peptides and corresponding proteins that are differentially expressed over time. Fig. 1 exhibits plots of natural logarithms of peptide intensities for three representative peptides measured at seven observation times 0.5, 1, 4, 7, 14, 21, and 28 days after the injury to a patient, labeled as times 1, ..., 7, respectively. Measurements on peptide 1 were obtained from 97 patients, whereas measurements on peptides 2 and 3 were obtained from 88 and 30 patients, respectively. The peptide intensities for a given patient are shown with a distinct color in the three plots. The log-intensities for peptide 2 indicate a positive trend over time, whereas they show a curvilinear trend for peptide 3.

A feature of the proteomics experiment is that the repeated measurements from the patients are correlated. To model the correlations among these repeated observations, we consider

using a linear mixed effects model with a block-diagonal covariance structure, as introduced in the next section.

Another important and also challenging feature of the proteomics study is that the data contain nonmonotone missing values, as the measurements on the peptide intensity were intermittently missing at the seven observation times. For example, for peptide 1, although measurements were obtained from 97 patients, only a subset of 75 patients had measurements at the first time-point; a different subset of 74 patients had measurements at the second time-point; a further different subset of 69 patients had measurements at the third time-point; and so on. Even though the overall proportion of observed responses decreases over time, the missing data pattern is nonmonotone, i.e., some patients' responses are missing at one occasion and observed at the next occasion.

The missing data mechanism in the proteomics experiment is considered nonignorable or not missing at random (NMAR) in that the missingness is due to the low abundance of the peptide intensities. Fig. 2 presents boxplots of peptide intensities (on the log2 scale) against five replicates (repeated measurements) from the trauma patients. Boxplot 1 shows the peptide intensities for patients who had a single replicate over the time course; boxplot 2 presents the intensities for patients with only two replicates; and so on. It is clear from the plots that the location of the peptide intensities increases with the number of replicates obtained from the patients, which is consistent with the notion that the low peptide intensities are less likely to be observed or detected. Another possibility is that the measurements with low intensities occur earlier in time, when they are lower because of the trend. In other words, time is a potential confounder. A referee pointed out that the missingness in the peptide intensities could also be due to a "limit of detection". However, it is difficult to identify such a detection limit from the proteomics experiment, as the peptide intensities are randomly missing.

For the analysis of the proteomics data, we adopt a likelihood-based approach under the assumption that the data are missing due to a stochastic nonignorable missing data mechanism. When data are nonignorably missing, for the likelihood-based inference, it is necessary to incorporate the missing data mechanism into the observed data likelihood function. As the proteomics experiment involved many peptides, and repeated measurements on each peptide were obtained from the group of trauma patients over seven observation times, the exact computation of the likelihood function would require intensive computation involving irreducibly high-dimensional integrals. To avoid these computational difficulties, we propose and explore a Monte Carlo method based on importance sampling when analyzing the data.

More generally, the proposed Monte Carlo method can be used for approximate likelihood inference in longitudinal data with nonignorable nonresponse. The finite-sample properties of the approximate maximum likelihood estimators are investigated in some detail. The proposed method is also extended to multiple testing based on false discovery rates under the assumption of a nonignorable missing data model.

The paper is organized as follows. Section 2 introduces the model and notation, and describes the proposed method for analyzing nonignorable and nonmonotone missing responses. Section 3 studies empirical properties of the Monte Carlo estimators based on a series of simulations. Section 4 presents an application of the proposed method using the proteomics data introduced earlier. Section 5 discusses the use of false discovery rates for multiple comparisons with incomplete longitudinal data. Section 6 concludes the paper with some discussion.

2. Model, notation and method

2.1. Estimation with complete data

Suppose k individuals are observed at a fixed set of n time points, $t = 1, \dots, n$. Let $\mathbf{y}_j = (y_{j1}, \dots, y_{jn})^t$ represent the response vector of repeated measurements from the j th individual ($j = 1, \dots, k$) and $\mathbf{x}_{jt}$ represent a $p \times 1$ vector of covariates associated with the response y_{jt} from individual j at time t . The covariates may include demographic variables and are considered to be fully observed.

Note that the repeated observations from the individuals are correlated by nature. To model this correlation structure and also to estimate the individual random effects and covariate effects, we consider using a linear mixed effects model with a block-diagonal covariance structure in the form

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{v}_j + \mathbf{e}_j, \quad (1)$$

for $j = 1, \dots, k$, with specified design matrices $\mathbf{X}_j$ and $\mathbf{Z}_j$ for the fixed effects parameters $\boldsymbol{\beta}$ and the vector of random effects $\mathbf{v}_j$, respectively. The random effects $\mathbf{v}_j$ and the random errors $\mathbf{e}_j$ are assumed to be independent and multivariate normally distributed with mean vectors $\mathbf{0}$ and covariance matrices $\mathbf{G}_j$ and $\mathbf{R}_j$, respectively. In this setting, the covariance matrix of $\mathbf{y} = (\mathbf{y}_1^t, \dots, \mathbf{y}_k^t)^t$ is block-diagonal with blocks

$\mathbf{V}_j = \mathbf{V}_j(\boldsymbol{\alpha}) = \mathbf{R}_j + \mathbf{Z}_j \mathbf{G}_j \mathbf{Z}_j^t$ ($j = 1, \dots, k$), depending on the vector $\boldsymbol{\alpha}$ of variance components. We may rewrite (1) in the matrix form

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z} \mathbf{v} + \mathbf{e}, \quad (2)$$

with $\mathbf{v} \sim N(\mathbf{0}, \mathbf{G})$, and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{R})$, where $\mathbf{G}$ and $\mathbf{R}$ are block-diagonal with blocks $\mathbf{G}_j$ and $\mathbf{R}_j$, respectively, for $j = 1, \dots, k$ (see Rao, 2003, p.107).

For given $\boldsymbol{\alpha}$, the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ is obtained as

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \left(\sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \mathbf{y}_j.$$

Under the normality of $\mathbf{v}$ and $\mathbf{e}$, the BLUE of $\boldsymbol{\beta}$ is also the maximum likelihood (ML) estimator. The ML estimators of the variance components $\boldsymbol{\alpha}$ cannot be obtained in an explicit form, and numerical methods can be used to find these estimators.

In the case of completely observed data, it is straightforward to find the ML estimators by numerically maximizing the likelihood function. But, in practice, missing data occur in many longitudinal studies where the measurements from the individuals are often missing due to a stochastic missing data mechanism. When this missingness mechanism is nonignorable, it is necessary to incorporate the missing data model into the observed data likelihood function for valid statistical inferences. In the next section, we discuss modeling the missing data pattern based on logistic regression.

2.2. Missing data model

Missing data models for longitudinal experiments have been extensively studied in the literature by many authors. Nonignorable missing data models are often considered when the missingness depends on the missing outcomes. Examples of such outcome-dependent missingness or dropouts are given in Little and Rubin (2002). Diggle and Kenward (1994) and Ibrahim et al. (2001) also review various types of nonignorable models with outcome-dependent dropouts.

To introduce our missing data model, we define n binary random variables, r_{jt} , ($t = 1, \dots, n$), with r_{jt} being equal to 1 if response y_{jt} is observed, and 0 if y_{jt} is missing. We consider a binary logistic regression model for the missing data indicators $\mathbf{r}_j = (r_{j1}, \dots, r_{jn})^t$ in the form

$$f(\mathbf{r}_j | \mathbf{y}_j, \mathbf{X}_j, \boldsymbol{\gamma}) = \prod_{t=1}^n \pi_{jt}^{r_{jt}} (1 - \pi_{jt})^{1-r_{jt}}, \quad (3)$$

where $\mathbf{X}_j$ is the design matrix for individual j and π_{jt} is the probability of response at time t , which is related to the “covariates” $\mathbf{y}_j$ and $\mathbf{x}_{jt}$ by the logistic regression model

$$\pi_{jt} = P(r_{jt}=1 | \mathbf{y}_j, \mathbf{x}_{jt}, \boldsymbol{\gamma}) = \frac{\exp(\gamma_0 + \gamma_1 y_{jt} + \gamma_2^t \mathbf{x}_{jk}^*)}{1 + \exp(\gamma_0 + \gamma_1 y_{jt} + \gamma_2^t \mathbf{x}_{jk}^*)}, \quad (4)$$

where the vector $\mathbf{x}_{jk}^*$ may include covariates $\mathbf{x}_{jt}$ and previous responses $y_{j,t-1}, y_{j,t-2}, \dots, y_{j1}$. Note that if $\gamma_1 = 0$, then the missing data mechanism is nonignorable, as the probability of nonresponse depends on possibly unobserved outcome y_{jt} .

The missing data model (3) assumes independence between the elements of $\mathbf{r}_j$, given the vector of responses $\mathbf{y}_j$ and design matrix $\mathbf{X}_j$. Note that model (3) provides a sensible way to specify the joint distribution of the missing-data indicators. Since each of the univariate distributions on the right side of (3) can be modeled as a logistic regression, each objective function is log-concave in the parameters. This property of log-concavity eases the computations of the maximum likelihood estimates. For a more general form of the missing data mechanism, see Little (1995) and Little and Rubin (2002). Ibrahim et al. (1999)

consider the missing data mechanism $f(\mathbf{r}_j|\mathbf{y}_j, \mathbf{X}_j, \boldsymbol{\gamma})$ as the product of a sequence of one-dimensional conditional binomial distributions for the missing data indicators.

The choice of appropriate covariates for the missing data model is also an important issue in incomplete longitudinal data analysis. To compare various models, one can use the likelihood ratio or Akaike information criterion for the fitted models. However, as pointed out by Baker and Laird (1988) and Ibrahim et al. (1999), one needs to be careful not to build a very large model for the missing data mechanism, since the model can easily become unidentifiable due to overparameterization. The issue of estimability can often arise in nonignorable missing data analysis and it is difficult to characterize the set of estimable parameters for a given class of models, as noted by Baker and Laird (1988).

2.3. Likelihood function for nonignorable missing data

Let $\{(\mathbf{y}_j, \mathbf{X}_j); j = 1, \dots, k\}$ denote the data that would occur in the absence of missing values. Also, let $\mathbf{y}_{\text{obs},j}$ denote the observed values and $\mathbf{y}_{\text{mis},j}$ the missing values of the response vector $\mathbf{y}_j$. Assuming arbitrary, nonmonotone patterns of missing data in $\mathbf{y}_j$, some permutation of the indices of $\mathbf{y}_j$ can be written as $\mathbf{y}_j = (\mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j})$, where $\mathbf{y}_{\text{mis},j}$ is the $n_j \times 1$ vector of missing values of $\mathbf{y}_j$.

For the j th individual, the observed data consist of the values $(\mathbf{y}_{\text{obs},j}, \mathbf{X}_j, \mathbf{r}_j)$. The distribution of these observed data may be obtained by integrating $\mathbf{y}_{\text{mis},j}$ out of the joint density of $(\mathbf{y}_j, \mathbf{r}_j)$ as

$$f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \int_{\mathbf{y}_{\text{mis},j}} f(\mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j} | \mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\alpha}) f(\mathbf{r}_j | \mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j}, \mathbf{X}_j, \boldsymbol{\gamma}) d\mathbf{y}_{\text{mis},j}. \quad (5)$$

The full likelihood of the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}', \boldsymbol{\gamma}')^t$ given the observed data $\{(\mathbf{y}_{\text{obs},j}, \mathbf{X}_j, \mathbf{r}_j); j = 1, \dots, k\}$ is any function of $\boldsymbol{\theta}$ proportional to the products of the densities in (5) for all k individuals. That is, the likelihood

$$L(\boldsymbol{\theta}) \equiv L(\boldsymbol{\theta} | \mathbf{y}_{\text{obs}}, \mathbf{X}, \mathbf{r}) \propto \prod_{j=1}^k f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \mathbf{X}_j, \boldsymbol{\theta}), \quad (6)$$

where $\mathbf{y}_{\text{obs}} = \{\mathbf{y}_{\text{obs},j}; j = 1, \dots, k\}$ and $\mathbf{r} = \{\mathbf{r}_j; j = 1, \dots, k\}$. This likelihood function cannot be expressed in a closed form, and we rely on numerical methods to evaluate the likelihood.

If the density function of the observed data $(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j)$ is in the form (5), then after some algebra we can show that the ML estimators of the regression parameters $\boldsymbol{\beta}$, variance components $\boldsymbol{\alpha}$, and nuisance parameters $\boldsymbol{\gamma}$ of the missing data model can be obtained by solving the estimating equations

$$\sum_{j=1}^k E \left\{ \frac{\partial \log f(\mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j} | \mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right\} = \mathbf{0}, \quad (7)$$

$$\sum_{j=1}^k E \left\{ \frac{\partial \log f(\mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j} | \mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right\} = \mathbf{0}, \quad (8)$$

and

$$\sum_{j=1}^k E \left\{ \frac{\partial \log f(\mathbf{r}_j | \mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j}, \mathbf{X}_j, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right\} = \mathbf{0}, \quad (9)$$

where the expectations are with respect to the conditional distribution of $\mathbf{y}_{\text{mis},j} | (\mathbf{y}_{\text{obs},j}, \mathbf{r}_j)$. The above score equations can be derived following McCulloch and Searle (2001, Chapter 8 p. 227–228). Details are given in the Appendix.

The estimating equations (7)–(9) are unbiased, meaning that $E[\mathbf{S}(\boldsymbol{\theta})] = \mathbf{0}$, where $\mathbf{S}(\boldsymbol{\theta})$ is the likelihood score function for $\boldsymbol{\theta}$. The solutions to these unbiased estimating equations give consistent estimators $\hat{\boldsymbol{\theta}}$. The asymptotic variance–covariance matrix of $\hat{\boldsymbol{\theta}}$ can be obtained from the observed Fisher information,

$$I_o(\boldsymbol{\theta}) = - \sum_{j=1}^k E \{ \partial \mathbf{U}_j / \partial \boldsymbol{\theta}^t | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \} - \sum_{j=1}^k E \{ \mathbf{U}_j \mathbf{U}_j^t | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \} + \sum_{j=1}^k E \{ \mathbf{U}_j | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \} E \{ \mathbf{U}_j^t | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \}, \quad (10)$$

where $\mathbf{U}_j = \mathbf{U}_j(\boldsymbol{\theta}) = \log f(\mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j}, \mathbf{r}_j | \mathbf{X}_j, \boldsymbol{\theta}) / \boldsymbol{\theta}$.

Note that the ML estimating equations (7)–(9) involve the calculation of conditional expectations with respect to the missing data $\mathbf{y}_{\text{mis},j}$, given the observed data $(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j)$. If the dimension of $\mathbf{y}_{\text{mis},j}$ is large, then the calculation of these expectations would require intensive computation. To avoid these computational difficulties, we suggest a Monte Carlo method based on importance sampling, as described in the next section.

2.4. Monte Carlo ML estimation

Recall the ML estimating equations (7) for the regression parameters $\boldsymbol{\beta}$. After some algebra, we can show that the ML estimators $\hat{\boldsymbol{\beta}}$ may be obtained iteratively from

$$\boldsymbol{\beta}^{(m+1)} = \left(\sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} E(\mathbf{y}_j | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta}), \quad (11)$$

where the expectation $E(\mathbf{y}_j | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta})$ is with respect to the conditional distribution of $\mathbf{y}_{\text{mis},j}$ given the observed data $(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j)$, and the parameters on the right side are evaluated at the current estimates $\boldsymbol{\theta}^{(m)}$ of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}^t, \boldsymbol{\gamma}^t)^t$, for $m = 0, 1, 2, \dots$

Similarly, the ML estimators $\hat{\boldsymbol{\alpha}}$ of the variance components $\boldsymbol{\alpha}$ can be obtained iteratively from the equations

$$\boldsymbol{\alpha}^{(m+1)} = \boldsymbol{\alpha}^{(m)} + \left\{ E \left(\frac{\partial^2 \log L}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^t} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta} \right) \right\}^{-1} E \left(\frac{\partial \log L}{\partial \boldsymbol{\alpha}} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta} \right), \quad (12)$$

for $m = 0, 1, 2, \dots$. As before, the second term on the right side are evaluated at the current estimates $\boldsymbol{\theta}^{(m)}$. For the above equations, we can show that

$$E \left(\frac{\partial \log L}{\partial \alpha_l} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta} \right) = -\frac{1}{2} \sum_{j=1}^k \text{tr} \left(\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \alpha_l} \right) + \frac{1}{2} \sum_{j=1}^k \text{tr} \left[\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \alpha_l} \mathbf{V}_j^{-1} E \{ \mathbf{z}_j \mathbf{z}_j^t | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta} \} \right] \quad (13)$$

and

$$\begin{aligned} E \left(\frac{\partial^2 \log L}{\partial \alpha_l \partial \alpha_s} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta} \right) &= -\frac{1}{2} \sum_{j=1}^k \text{tr} \left(-\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \alpha_l} \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \alpha_s} + \mathbf{V}_j^{-1} \frac{\partial^2 \mathbf{V}_j}{\partial \alpha_l \partial \alpha_s} \right) \\ &+ \frac{1}{2} \sum_{j=1}^k \text{tr} \left[\left(-2\mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \alpha_l} \mathbf{V}_j^{-1} \frac{\partial \mathbf{V}_j}{\partial \alpha_s} + \mathbf{V}_j^{-1} \frac{\partial^2 \mathbf{V}_j}{\partial \alpha_l \partial \alpha_s} \mathbf{V}_j^{-1} \right) E \{ \mathbf{z}_j \mathbf{z}_j^t | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta} \} \right], \end{aligned} \quad (14)$$

where $\mathbf{z}_j = \mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}$ and α_l is the l th component of the variance components $\boldsymbol{\alpha}$.

Clearly, Eqs. (11)–(14) involve the calculation of the expectations $E(\mathbf{y}_j | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta})$ and $E\{(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})(\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta})^t | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta}\}$ with respect to the conditional distribution of $(\mathbf{y}_{\text{mis},j} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j)$, which require intensive computation involving multidimensional integrals. Here we suggest a Monte Carlo method based on importance sampling for approximating these expectations. Importance sampling is widely used to approximate an expectation $E_f[h(\mathbf{y})]$ with respect to a density $f(\mathbf{y})$ by replacing $f(\mathbf{y})$ with a similar, but easily sampled, distribution $f^*(\mathbf{y})$, and then correcting for the bias introduced by sampling from the wrong distribution.

To approximate $E(\mathbf{y}_j | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta})$, we can write (suppressing suffix j)

$$E(\mathbf{y} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta}) \approx \frac{1}{S} \sum_{s=1}^S \mathbf{y}^{(s)} \frac{f(\mathbf{y}_{\text{mis}}^{(s)} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta})}{f^*(\mathbf{y}_{\text{mis}}^{(s)})} \equiv \frac{1}{S} \sum_{s=1}^S w^{(s)} \mathbf{y}^{(s)}, \quad (15)$$

where $\mathbf{y} = (\mathbf{y}_{\text{obs}}^t, \mathbf{y}_{\text{mis}}^t)^t$, $\mathbf{y}^{(s)} = (\mathbf{y}_{\text{obs}}^t, \mathbf{y}_{\text{mis}}^{(s)t})^t$, and the “importance weight” $w^{(s)}$ is obtained as

$$w^{(s)} = \frac{f(\mathbf{y}_{\text{mis}}^{(s)} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta})}{f^*(\mathbf{y}_{\text{mis}}^{(s)})}, \quad (16)$$

for given $\mathbf{y}_{\text{mis}}^{(s)}$ generated from a similar “sampling distribution” $f^*(\mathbf{y}_{\text{mis}}^{(s)})$ other than the computationally intractable distribution $f(\mathbf{y}_{\text{mis}} | \mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta})$. For simplicity, here we consider $f^*(\mathbf{y}_{\text{mis}}^{(s)})$ as the marginal distribution of $\mathbf{y}_{\text{mis}}$.

Similarly to (15), we approximate the conditional expectation of the cross-product term:

$$E\{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^t|\mathbf{y}_{\text{obs}}, \mathbf{r}, \boldsymbol{\theta}\} \approx \frac{1}{S} \sum_{s=1}^S w^{(s)} (\mathbf{y}^{(s)} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y}^{(s)} - \mathbf{X}\boldsymbol{\beta})^t, \quad (17)$$

where the weights $w^{(s)}$ are defined as in (16). We find the Monte Carlo ML estimators $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\alpha}}$ of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ from the iterative equations (11) and (12) by replacing the expectations $E(\mathbf{y}_j|\mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta})$ and $E\{(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\beta})^t|\mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta}\}$ with their approximations (15) and (17), respectively.

To estimate the nuisance parameters $\boldsymbol{\gamma}$ of the missing data model (3), we use a third set of iterative equations in the form

$$\boldsymbol{\gamma}^{(m+1)} = \boldsymbol{\gamma}^{(m)} + \left\{ \sum_{j=1}^k E(\mathbf{X}_j^* \mathbf{W}_j^{*-1} \mathbf{X}_j^* | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta}) \right\}^{-1} \sum_{j=1}^k E(\mathbf{X}_j^* (\mathbf{r}_j - \boldsymbol{\pi}_j) | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta}), \quad (18)$$

for $m = 0, 1, 2, \dots$; $\mathbf{X}_j^*$ is the design matrix associated with the missing data model and $\mathbf{W}_j^*$ is a diagonal matrix with diagonal elements $w_{jt}^* = \pi_{jt}(1 - \pi_{jt})$. Here also the conditional expectations on the right side of (18) are approximated by the Monte Carlo method based on importance sampling, as described earlier.

Note that the number of replications S required to get the stochastic Monte Carlo estimators to converge with three-or two-decimal accuracy would be very large. We studied the Monte Carlo estimators in simulations using a simple linear mixed effects model and found that these stochastic estimators could provide good approximations to the deterministic ones for $S = 500$ replications. However, for a complex model with a large proportion of missing data, we may need a much larger number of replications in order to obtain good approximations to the exact ML estimators. For analyzing the proteomics data introduced earlier, we used 10,000 replications of importance samples, for which the Monte Carlo ML estimators converged with three-decimal accuracy. An advantage of the importance sampling is that we can easily draw replications of random samples based on a known and easily sampled distribution, whereas it is difficult to generate random observations from the conditional distribution of the missing data given the observed data.

2.5. Variance of Monte Carlo estimator

Recall the ML estimating equations (11) for the regression parameters $\boldsymbol{\beta}$. The stochastic Monte Carlo approximations to the ML estimators are obtained as

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \bar{\mathbf{y}}_j^*, \quad (19)$$

where

$$\bar{\mathbf{y}}_j^* = \frac{1}{S} \sum_{s=1}^S w_j^{(s)} \mathbf{y}_j^{(s)} \equiv \frac{1}{S} \sum_{s=1}^S \mathbf{y}_j^{*(s)},$$

with random observations $\mathbf{y}_j^{(s)}$ and weights $w_j^{(s)}$, as defined in (15) and (16). An approximate variance–covariance matrix of the estimators $\tilde{\boldsymbol{\beta}}$ may be obtained from the sandwich-type variance estimator

$$V(\tilde{\boldsymbol{\beta}}) = \left(\sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1} \left(\sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \text{Var}(\bar{\mathbf{y}}_j^*) \mathbf{V}_j^{-1} \mathbf{X}_j \right) \left(\sum_{j=1}^k \mathbf{X}_j^t \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1}. \quad (20)$$

where $\text{Var}(\bar{\mathbf{y}}_j^*) = \mathbf{V}_j$ if the response vector $\mathbf{y}_j$ for subject j is completely observed; otherwise, it can be approximated as

$$\begin{aligned} \text{Var}(\bar{\mathbf{y}}_j^*) &= \text{Var} \left\{ E \left(\frac{1}{S} \sum_{s=1}^S \mathbf{y}_j^{*(s)} \mid \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right) \right\} + E \left\{ \left(\frac{1}{S} \sum_{s=1}^S \mathbf{y}_j^{*(s)} \mid \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right) \right\} \\ &\approx \left(\frac{1}{S} \sum_{s=1}^S \mathbf{y}_j^{*(s)} - \mathbf{X}_j \boldsymbol{\beta} \right) \left(\frac{1}{S} \sum_{s=1}^S \mathbf{y}_j^{*(s)} - \mathbf{X}_j \boldsymbol{\beta} \right)^t + \frac{1}{S} \left\{ \sum_{s=1}^S \left(\mathbf{y}_j^{*(s)} - \bar{\mathbf{y}}_j^* \right) \left(\mathbf{y}_j^{*(s)} - \bar{\mathbf{y}}_j^* \right)^t \right\}. \end{aligned} \quad (21)$$

We found in an empirical study (shown later) that (21) provides a good approximation to the sandwich-type variance estimator $V(\tilde{\boldsymbol{\beta}})$ for a reasonably large number of replications S under importance sampling.

3. Simulation study

To explore the performance of the proposed Monte Carlo method, we ran a series of simulations based on a linear mixed effects model with an intercept random effect term. We considered generating incomplete longitudinal data for k patients observed over a fixed set of three time-points $t = 1, 2, 3$ using the linear mixed model

$$y_{jt} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + v_j + e_{jt}, \quad j = 1, \dots, k, \quad (22)$$

where y_{jt} represents a continuous response from patient j at time t ; binary covariates x_{1j} and x_{2j} are indicators of the second and third time points; β_0 represents the overall response at baseline time $t = 1$; and β_1 and β_2 represent time effects for $t = 2$ and $t = 3$, respectively. The random effects v_j were assumed to be independent normal with mean 0 and variance τ^2 . The random errors e_{jt} were also assumed to be independent normal with mean 0 and variance σ^2 . Further, v_j and e_{jt} were assumed to be independent of each other.

Without loss of generality, we assumed that the data were all observed at the first time-point for ease of computation. However, incomplete data were obtained from the second and third time-points according to the missing data mechanism

$$P(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma}) = \prod_{t=2}^3 \pi_{jt}^{r_{jt}} (1 - \pi_{jt})^{1-r_{jt}}, \quad (23)$$

where $\mathbf{r}_j = (r_{j1}, r_{j2}, r_{j3})^t$ is the vector of missing data indicators, r_{jt} , with response probabilities

$$\pi_{jt} = P(r_{jt}=1 | y_{jt}, \boldsymbol{\gamma}) = \frac{\exp(\gamma_0 + \gamma_1 y_{jt})}{1 + \exp(\gamma_0 + \gamma_1 y_{jt})}, \quad t=1, 2. \quad (24)$$

Here our main interest is in the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^t$ and the variance components $\boldsymbol{\alpha} = (\tau^2, \sigma^2)^t$, with $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^t$ being viewed as the nuisance parameters. The regression parameters were fixed at $\boldsymbol{\beta} = (10, 1.5, -0.5)^t$, whereas the parameters of the missing data model were fixed at $\boldsymbol{\gamma} = (-4, 0.5)^t$. For each combination of the values $\{(0.5, 0.5), (0.5, 1.0), (1.0, 0.5), (1.0, 1.0)\}$ of the variance components (τ^2, σ^2) and the number of patients $k = 60$ and $k = 120$, we generated 500 data sets based on models (22)–(24). For a relatively simple model as we consider here, simulations with 500 runs were found sufficient to provide stable empirical results for the quantities of interest.

For the above simulation setting, each data set contained roughly 16% and 34% missing values at the two time-points $t = 2$ and $t = 3$, respectively. The estimates of the model parameters were obtained by using the following three methods:

1. *Exact ML (ML)*: Estimates obtained from (11), (12) and (18) using numerical integration.
2. *Monte Carlo ML (MCML)*: Proposed estimates based on 500 Monte Carlo samples.
3. Estimates obtained under the missing at random (MAR) assumption.

Methods 1 and 2 are based on correctly specified nonignorable missing data model (24), that is, when data are not missing at random (NMAR). Method 3 is based on a misspecified MAR assumption for which the likelihood-based inference does not depend on the missing data model. Consequently, under the MAR model, the likelihood analysis of the data is performed based on the marginal distribution of the observed data $\{\mathbf{y}_{\text{obs},j}; j = 1, \dots, k\}$.

Note that in the simulation study, although the exact ML method involves only two-dimensional integration with respect to the missing data, it still requires much more computation as compared to the proposed Monte Carlo method. If the dimension of the vector of missing data becomes large, then it would be difficult to compute the exact ML estimates numerically, and the proposed Monte Carlo method would be a practical alternative to the exact method.

Table 1 presents the empirical biases and mean squared errors of the estimators of the regression parameters and the variance components under the three methods obtained for data with $k = 60$ patients. The exact ML and the proposed Monte Carlo ML methods appear to perform almost equally well in terms of biases and mean squared errors. The two methods provide approximately unbiased estimators of the regression parameters under all simulation

configurations. It is also clear from Table 1 that the Monte Carlo estimators are almost as efficient as the exact ML estimators. For example, at $(\tau^2, \sigma^2) = (2, 2)$, for the time effect β_1 , the Monte Carlo estimator is 97% as efficient as the exact ML estimator, and for β_2 , this estimator is 94% as efficient as the exact ML estimator.

Under the misspecified MAR assumption, it is clear that method 3 provides systematically large biases and mean squared errors of the estimators, as expected. For example, at $(\tau^2, \sigma^2) = (2, 2)$, the estimator of β_1 under MAR provides an absolute percentage relative bias of 15%, and that of β_2 provides a relative bias of 81%!

Table 2 presents the empirical biases and mean squared errors of the estimators obtained for data with $k = 120$ patients. Here also it is clear that the proposed Monte Carlo estimators are close to the exact ML estimators in terms of both biases and mean squared errors. On the other hand, the estimators under the misspecified MAR model provide large biases for all the regression parameters and the variance components. For example, at $(\tau^2, \sigma^2) = (2, 2)$, the ML estimator of σ^2 under MAR provides an absolute percentage relative bias of 11%. The mean squared errors of both the ML and approximate ML estimators are almost uniformly smaller than those of the estimators obtained under the MAR model. The large mean squared errors of the estimators are due to the large biases obtained under the misspecified MAR assumption.

It is important to point out that as the MCML is a simulation approximation to the ML, our focus is on how well it approximates the ML estimators, not on how it compares with the MAR method, which is, in fact, based on a wrong model here. From this standpoint, the MCML appears to work fine for the regression coefficients but not so good for the variances in some cases.

3.1. Coverage probabilities

We also investigated the extent to which the MCML estimators of the regression coefficients $(\beta_0, \beta_1, \beta_2)$ are adequate to allow for normal theory inferential procedures in samples of moderate size. The coverage probabilities were obtained for 95% confidence intervals, $\hat{\beta} \pm 1.96S.E.(\hat{\beta})$, where the estimated standard errors, $S.E.(\hat{\beta})$'s, of the MCML estimators, $\hat{\beta}$'s, were obtained from the sandwich-type variance estimator given in (20). The coverage probabilities were obtained for the same simulation setting as considered earlier with $(\beta_0, \beta_1, \beta_2) = (10, 1.5, -0.5)$ and $(\tau^2, \sigma^2) = (0.5, 0.5)$. Table 3 reports the coverages and mean lengths of the confidence intervals for the proposed MCML estimators. It is clear from the table that the empirical coverages are generally close to the nominal 95% confidence level. As expected, the lengths of the confidence intervals decrease with the larger sample size ($k = 120$).

Table 3 also reports the actual (simulated) and estimated standard errors of the MCML estimators. The estimated standard errors are generally close to the actual standard errors of the regression estimators. For the larger sample size with $k = 120$, the estimated standard errors are even closer to the actual standard errors. For example, with $k = 60$ the difference in actual and estimated standard errors for β_2 is 0.0198, whereas the difference is only 0.0054 for $k = 120$.

3.2. Hypotheses tests

To investigate whether the response y_{jt} varies across time, we can test the null hypothesis $H_0: \beta^* = 0$ with $\beta^* = (\beta_1, \beta_2)^t$ at a certain level of significance. As the calculation of the likelihood function requires intensive computation, we consider a simple Wald test based on the ML estimators of the regression coefficients and the corresponding variance matrix given in (20). If the normal theory of the approximate ML estimators is adequate, then the Wald statistic $(\hat{\beta}^* - \beta^*)^t V^{*-1}(\hat{\beta}^*)(\hat{\beta}^* - \beta^*)$ is asymptotically chi-square with 2 degrees of freedom.

To explore the finite-sample properties of the Wald test under the nonignorable missing data model as considered here, we ran a series of simulations based on 500 data sets using the previous setting but with $(\beta_0, \beta_1, \beta_2) = (10, 0, 0)$ and $(\tau^2, \sigma^2) = (1, 0.25)$. Here we set $(\beta_1, \beta_2) = (0, 0)$ to examine the level of significance of the Wald test using the proposed MCML estimators. Table 4 presents empirical levels of significance of the tests along with their approximate standard errors. We note from the table that the empirical levels of the tests are generally close to their nominal levels. In particular, at the nominal 0.10 level of significance, the empirical level appears to be closer to the nominal level with the larger sample size $k = 120$.

3.3. Sensitivity analysis

Note that we consider a nonignorable missing data model throughout the paper. Fitting such a nonignorable nonresponse model should be subject to a careful examination, as estimates of the model parameters can be sensitive to the missing data model being chosen. Sensitivity analysis for missing data has been considered by many authors in the literature (for example, (Diggle and Kenward, 1994; Rotnitzky et al., 1998; Molenberghs et al., 1999; Verbeke et al., 2001)). Such analysis may be performed in terms of studying how estimates of the model parameters change with changes in the missing data model parameters. Typically, more confidence can be placed in a model which is relatively stable under small modifications.

Here we perform a simple sensitivity analysis for the missing data model (24) as used in the simulations. This missing data mechanism will become MAR, and hence ignorable, corresponding to the setting $\gamma_1 = 0$. We conduct the sensitivity analysis by studying how the parameter estimates change with changing values of γ_1 . For this, we generated a data set with size $k = 200$ using the same simulation setting considered earlier. The joint ML estimates of the model parameters were obtained as $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = (10.07, 1.51, -0.52)$, $(\hat{\tau}^2, \hat{\sigma}^2) = (0.45, 0.48)$ and $(\hat{\gamma}_0, \hat{\gamma}_1) = (-3.22, 0.42)$. Fig. 3 exhibits estimates of the regression parameters and variance components against γ_1 . The values of γ_1 include both 0 and the ML estimate $\hat{\gamma}_1 = 0.42$. The plots do not indicate any extreme sensitivity of the ML estimates to the choice of γ_1 . In particular, the estimates of the variance components do not seem to be very sensitive to the choice of γ_1 . The regression estimates, however, tend to decrease at a small rate with increasing γ_1 , which is consistent with the fact that if the estimate $\hat{\gamma}_1 = 0.42$ is replaced by 0 under the MAR assumption, the regression estimators would be systematically biased.

4. Application: proteomics data

In this section, we analyze the three sets of proteomics data introduced in Section 1. Table 5 reports the counts of peptide intensities observed from the group of trauma patients over seven time-points. Clearly, data from each peptide contain nonmonotone missing values, and generally more missing data occur at latter time-points. These intermittent missing values are considered nonignorable in that the missingness is due to the low abundance of the peptide intensities. In this note, we analyzed the data by incorporating a missing data model into the observed data likelihood function.

For each peptide, we analyze the peptide intensities using a linear mixed effects model

$$y_{jt} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_3 x_{3j} + \beta_4 x_{4j} + \beta_5 x_{5j} + \beta_6 x_{6j} + v_j + e_{jt}, \quad (25)$$

where y_{jt} represents the natural logarithm of the peptide intensity for patient j at time t , $t = 1, \dots, 7$; the intercept term β_0 represents the overall log-intensity at the baseline time; and the regression coefficients $\beta_1, \dots, \beta_6$ represent the time effects at the six follow-up times. The binary covariate x_{lj} is defined to be 1 if response y is from patient j at time $l + 1$, and 0 otherwise, for $l = 1, \dots, 6$. The random effects v_j and the random errors e_{jt} are assumed to be independently and normally distributed with means 0, and variances τ^2 and σ^2 , respectively.

The binary random variable r_{jt} is defined to be 1 if y_{jt} is observed and 0, if y_{jt} is missing. Under the assumption of working independence, we use a simple logistic regression model for the missing data in the form

$$P(\mathbf{r}_j | \mathbf{y}_j, \gamma) = \prod_{t=1}^7 \pi_{jt}^{r_{jt}} (1 - \pi_{jt})^{1-r_{jt}}, \quad (26)$$

where the probability of response for individual j at time t is defined by

$$\pi_{jt} = P(r_{jt} = 1 | y_{jt}, \gamma) = \frac{\exp(\gamma_0 + \gamma_1 y_{jt} + \gamma_2 z_t)}{1 + \exp(\gamma_0 + \gamma_1 y_{jt} + \gamma_2 z_t)};$$

$z_t = (\text{Time}_t/5)$ with Time_t being the actual time in days at time-point t ; γ_1 represents the effects of (possibly unobserved) response y_{jt} on the response probability π_{jt} . Note that if $\gamma_1 = 0$, then the missing data mechanism is ignorable or missing at random (MAR).

We fit model (25) using the proposed approximate ML method under the NMAR model (26), and also using the observed data likelihood under the MAR model. The approximate ML method was based on 10,000 replicates of Monte Carlo samples drawn by importance sampling discussed earlier. Table 6 presents the estimates of the model parameters and their corresponding standard errors for the three peptides. It is clear from the table that the time effects are generally different under the two methods. For example, for peptide 1, the proposed method gives an estimate of 0.262 for the regression coefficient β_5 , whereas

method 3 under the MAR model gives a much larger estimate of 0.507. We also observe that the nonignorable missing data model (NMAR) provides generally smaller standard errors of the estimators, as compared to the MAR model. For example, for peptide 3, the two methods provide close estimates of 0.475 and 0.460 for the coefficient β_4 , but the proposed ML method provides a much smaller standard error of 0.262, as compared to the standard error of 0.535 under the MAR model.

From the estimates under the nonignorable model, we observe that for peptide 1, the level of intensity decreases after 24 h of injury, and then increases gradually, with the highest peak on day 14. We observe a similar pattern in peptide 2, with the highest peak on day 7. For peptide 3, we observe a somewhat more dramatic change to the level of intensity across the time-points; it appears to increase up to day 14, and then decreases gradually (with negative estimates) on days 21 and 28.

The estimates of the variance components τ^2 and σ^2 are somewhat similar under the two missing data models. The estimates for the nonignorable missing data model indicate that the covariate “time” has negative effects on the response probabilities, whereas the “covariate” y has positive effects. The positive effects of y on the response probabilities are consistent with our assumption that the missingness in the data are due to the low abundance of the peptide intensities.

An objective of this proteomics experiment is to investigate if the peptides are differentially expressed over time. For this, we set the null hypothesis that all coefficients representing the time effects are zero:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

against the alternative that at least one of the coefficients is non-zero. We use a simple Wald-type chi-square test of the form

$$(\hat{\beta} - \beta)^t V^{-1}(\hat{\beta})(\hat{\beta} - \beta) \sim \chi_6^2,$$

where $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_6)^t$ and $V(\hat{\beta}) = \text{Var}(\hat{\beta})$ is obtained from the approximate sandwich-type variance-covariance function (20). We test the null hypothesis H_0 under both the nonignorable and ignorable models considered. Table 7 presents the values of the test statistics with their corresponding p -values for each of the three peptides considered for the analysis. It is clear from the table that the two models produce somewhat similar results for peptides 1 and 2; the p -values are small under both models. But the results are very different for peptide 3; our proposed method gives a much larger value of the test statistic (81.7 vs. 6.6) as compared to the second method under the MAR assumption. This is perhaps due to the fact that the MAR model produced larger standard errors of the regression estimators as compared to the proposed approximate ML method, as indicated in Table 6.

5. Multiple comparisons

In the proteomics experiment, we conduct hypothesis tests simultaneously for many peptides. A common problem with the multiple-test is that the chances of obtaining a positive result can be high even if all null hypotheses are true. An approach to the multiple-testing problem is to control the *false discovery rate* (FDR) (Benjamini and Hochberg, 1995). The FDR is defined as the expected proportion of false positives among all rejected hypotheses. Let V denote the number of type I errors (false positive results) and R denote the total number of rejected hypotheses. Then the FDR is defined by

$$\text{FDR} = E \left[\frac{V}{R \vee 1} \right], \quad (27)$$

where $R \vee 1 = \max(R, 1)$.

Suppose there are M peptides in the proteomics experiment, and let p_i denote the p -value of the hypothesis test for peptide i , $i = 1, \dots, M$. Also, let $\text{FDR}(u)$ denote the FDR when rejecting the null hypotheses with $p_i \leq u$ for $u \in [0, 1]$. Then in terms of empirical values of V and R , the FDR can be defined as

$$\text{FDR}(u) = E \left[\frac{V(u)}{R(u) \vee 1} \right], \quad (28)$$

where $V(u)$ is the number of null hypotheses with $p_i \leq u$ (type I errors) and $R(u)$ is the total number of hypotheses with $p_i \leq u$, $i = 1, \dots, M$.

An estimate of $\text{FDR}(u)$ proposed by Storey (2002) is obtained from

$$\hat{\text{FDR}}_\delta(u) = \frac{\hat{\pi}_0(\delta)u}{\{R(u) \vee 1\}/M}, \quad (29)$$

where $\hat{\pi}_0(\delta)$ is an estimate of the proportion of true null hypotheses, $\pi_0 \equiv M_0/M$, with M_0 being the number of true null hypothesis. The estimate $\hat{\pi}_0(\delta)$ is defined as

$$\hat{\pi}_0(\delta) = \frac{W(\delta)}{(1-\delta)M}, \quad (30)$$

where $W(\delta) = M - R(\delta)$.

In the proteomics analysis, we investigated the FDRs for multiple-testing based on 1000 peptides randomly chosen from the experiment. Fig. 4 displays the empirical FDRs,

$\hat{\text{FDR}}_\delta(u)$ at $\delta = 0.99$, against u . The largest value of u for which $\hat{\text{FDR}}_\delta(u) \leq 0.01$ is 0.009. So to perform the multiple-test at the fixed FDR level 0.01, we considered rejecting the null hypotheses of no differential expression for p -values less than 0.009. Based on this criterion,

65% of the peptides were found to have rejected the null hypotheses, and so we may conclude that at FDR level 0.01 those peptides were differentially expressed over time.

6. Discussion

We have proposed and explored a Monte Carlo method for analyzing incomplete longitudinal data with nonignorable and nonmonotone missingness. We have demonstrated in simulations that the proposed Monte Carlo estimators are almost as efficient as the exact maximum likelihood estimators. The simulation study also indicated that one could attain considerable biases in the estimators when data are analyzed under the misspecified MAR assumption.

Note that we have considered implementing the rejection sampling through the Monte Carlo Newton–Raphson algorithm. Alternatively, one can use this rejection sampling through the stochastic EM algorithm for approximating the ML estimators of model parameters. We would expect similar results from the two algorithms.

A referee also pointed out that as an alternative to the proposed approximate maximum likelihood inference, one could also consider approximate Bayesian inference using a Gibbs' sampler for analyzing the incomplete data. Bayesian inference is beyond the scope of the current research. We intend to investigate the Bayesian approach in future research.

The proposed Monte Carlo technique would be particularly useful when we are to analyze a large number of peptides in a proteomics experiment, where the exact maximum likelihood method would require intensive computation under a nonignorable missing data model. The proposed method can also be applied to microarray analysis in situations where time course gene expression data may be collected from arrays of genes and the data may be partially missing due to low gene expression levels.

We have briefly discussed the application of false discovery rates for multiple hypotheses tests using the proteomics experiment. The performance of the technique of false discovery rates has not been explored yet in detail in the context of missing data. It would be interesting to see how the empirical false discovery rates perform with various types of nonignorable and nonmonotone missing data mechanisms. More work remains to be done in this direction.

Acknowledgments

This research is partially supported by a grant from the Natural Sciences and Engineering Research Council of Canada. The authors thank Professor Wing H. Wong at Stanford University for useful comments on the manuscript. The authors also thank two referees, an associate editor, and the editor for helpful comments and suggestions that led to significant improvements in the presentation.

References

- Baker SG, Laird NM. Regression analysis for categorical variables with outcome subject to nonresponse. *J Amer Statist Assoc.* 1988; 83:62–69.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 1995; 57:289–300.

- Diggle P, Kenward MG. Informative dropout in longitudinal data analysis (with discussion). *Appl Stat*. 1994; 43:49–94.
- Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. 2. Oxford University Press; New York: 2002.
- Ibrahim JG, Chen MH, Lipsitz SR. Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika*. 2001; 88:551–564.
- Ibrahim JG, Lipsitz SR, Chen MH. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *J R Stat Soc Ser B*. 1999; 61:173–190.
- Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Amer Statist Assoc*. 1995; 90:1112–1121.
- Little, RJA.; Rubin, DB. *Statistical Analysis With Missing Data*. 2. John Wiley & Sons; New Jersey: 2002.
- McCulloch, CE.; Searle, SR. *Generalized, Linear, and Mixed Models*. Wiley; New York: 2001.
- Molenberghs G, Goetghebuer E, Lipsitz SR, Kenward MG. Non-random missingness in categorical data: strengths and limitations. *Amer Statist*. 1999; 53:110–118.
- Molenberghs G, Verbeke G. A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statist Model*. 2001; 1:235–269.
- Rao, JNK. *Small Area Estimation*. Wiley; New York: 2003.
- Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J Amer Statist Assoc*. 1998; 93:1321–1339.
- Sinha SK, Laird NM, Fitzmaurice GM. Multivariate logistic regression with incomplete covariate and auxiliary information. *J Multivariate Anal*. 2010; 101:2389–2397.
- Sinha SK, Troxel AB, Lipsitz SR, Sinha D, Fitzmaurice GM, Molenberghs G, Ibrahim JG. A bivariate pseudo-likelihood for incomplete longitudinal binary data with nonignorable non-monotone missingness. *Biometrics*. 2011; 67:1119–1126. [PubMed: 21155748]
- Storey JD. A direct approach to false discovery rates. *J R Stat Soc Ser B*. 2002; 64:479–498.
- Verbeke G, Molenberghs G. Longitudinal and incomplete clinical studies. *Metron*. 2005; 63:143–170.
- Verbeke G, Molenberghs G, Thijs H, Lesaffre E, Kenward MG. Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*. 2001; 57:7–14. [PubMed: 11252620]
- Wu L, Liu W, Liu J. A longitudinal study of children's aggressive behaviours based on multivariate mixed models with incomplete data. *Canad J Statist*. 2009; 37:435–452.
- Xie H. A local sensitivity analysis approach to longitudinal non-Gaussian data with non-ignorable dropout. *Stat Med*. 2008; 27:3155–3177. [PubMed: 17948917]
- Yi GY, Cook RJ. Marginal methods for incomplete longitudinal data arising in clusters. *J Amer Statist Assoc*. 2002; 97:1071–1080.

Appendix

Derivation of Eqs. (7)–(9)

Recall Eq. (5) for the joint distribution of the observed data for individual j . We have

$$f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta}) = \int_{\mathbf{y}_{\text{mis},j}} f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha}) f(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma}) d\mathbf{y}_{\text{mis},j}.$$

So for individual j , the score function of $\boldsymbol{\beta}$ can be expressed as

$$\begin{aligned}
\frac{\partial \log f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \frac{1}{f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta})} \int \mathbf{y}_{\text{mis},j} \frac{\partial f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} f(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma}) d\mathbf{y}_{\text{mis},j} \\
&= \int \mathbf{y}_{\text{mis},j} \frac{\frac{\partial \log f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} f(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma})}{f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta})} d\mathbf{y}_{\text{mis},j} \\
&= \int \mathbf{y}_{\text{mis},j} \frac{\partial \log f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} f(\mathbf{y}_{\text{mis},j} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta}) d\mathbf{y}_{\text{mis},j} \\
&= E \left\{ \frac{\partial \log f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right\}.
\end{aligned}$$

Then for k individuals, the score equation for $\boldsymbol{\beta}$ becomes

$$\sum_{j=1}^k E \left\{ \frac{\partial \log f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right\} = \mathbf{0}.$$

Similarly, we can derive the score Eqs. (8) for the variance components $\boldsymbol{\alpha}$. Also, the estimating equations (9) for the parameters $\boldsymbol{\gamma}$ of the missing data mechanism are obtained from the score function,

$$\begin{aligned}
\frac{\partial \log f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta})}{\partial \boldsymbol{\gamma}} &= \frac{1}{f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta})} \int \mathbf{y}_{\text{mis},j} f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha}) \frac{\partial f(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} d\mathbf{y}_{\text{mis},j} \\
&= \int \mathbf{y}_{\text{mis},j} \frac{\frac{\partial \log f(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta})} d\mathbf{y}_{\text{mis},j} \\
&= \int \mathbf{y}_{\text{mis},j} \frac{\partial \log f(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} f(\mathbf{y}_{\text{mis},j} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j, \boldsymbol{\theta}) d\mathbf{y}_{\text{mis},j} \\
&= E \left\{ \frac{\partial \log f(\mathbf{r}_j | \mathbf{y}_j, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right\}.
\end{aligned}$$

The estimating equations (7) are unbiased, since

$$\begin{aligned}
E \left[\sum_{j=1}^k \frac{\partial \log f(\mathbf{y}_{\text{obs},j}, \mathbf{r}_j | \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \right] &= E \left[\sum_{j=1}^k E \left\{ \frac{\partial \log f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} | \mathbf{y}_{\text{obs},j}, \mathbf{r}_j \right\} \right] \\
&= \sum_{j=1}^k E \left\{ \frac{\partial \log f(\mathbf{y}_j | \boldsymbol{\beta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \right\} \\
&= \sum_{j=1}^k E \left\{ \mathbf{X}_j^t \mathbf{V}_j^{-1} (\mathbf{y}_j - \mathbf{X}_j \boldsymbol{\beta}) \right\} = \mathbf{0}.
\end{aligned}$$

Using similar arguments, we can show that the estimating equations (8) and (9) are also unbiased.

Fig. 1.
Logarithm of peptide intensities at seven time points.

Fig. 2.
Boxplots of peptide intensities for five replicates.

Fig. 3. ML estimates versus γ_1 . Horizontal line represents ML estimates for $\hat{\gamma}_1 = 0.42$.

Fig. 4.
Estimated false discovery rates for $\delta = 0.99$.

Table 1

Percentage relative biases and mean squared errors of estimators for $k = 60$.

Parameter	True value	Relative bias (%)				Mean squared error			
		ML	MCML	MAR	ML	MCML	MAR	ML	MAR
β_0	10.0	-0.0610	-0.0610	-0.0610	0.0169	0.0169	0.0169	0.0169	0.0169
β_1	1.5	0.2718	0.1890	3.7375	0.0186	0.0187	0.0187	0.0211	0.0211
β_2	-0.5	1.0620	0.9085	-22.0140	0.0252	0.0264	0.0264	0.0351	0.0351
τ^2	0.5	-0.1235	-0.2981	-3.7311	0.0197	0.0197	0.0197	0.0195	0.0195
σ^2	0.5	-2.3392	-1.9935	-3.4976	0.0055	0.0055	0.0055	0.0055	0.0055
β_0	10.0	0.1363	0.1363	0.1363	0.0413	0.0413	0.0413	0.0413	0.0413
β_1	1.5	0.0383	-0.0675	11.7663	0.0878	0.0949	0.0949	0.1094	0.1094
β_2	-0.5	1.9790	3.8373	-69.2080	0.0904	0.1069	0.1069	0.1958	0.1958
τ^2	0.5	-5.1786	-4.0773	-11.8658	0.0638	0.0674	0.0674	0.0889	0.0889
σ^2	2.0	-2.2595	-0.8428	-5.9401	0.0872	0.0985	0.0985	0.0876	0.0876
β_0	10.0	-0.0996	-0.0996	-0.0996	0.0406	0.0406	0.0406	0.0406	0.0406
β_1	1.5	0.9253	0.9633	5.1913	0.0235	0.0235	0.0235	0.0289	0.0289
β_2	-0.5	-1.8708	-0.7357	-25.3252	0.0221	0.0225	0.0225	0.0381	0.0381
τ^2	2.0	-1.7182	-1.0911	-4.4405	0.1356	0.1412	0.1412	0.1425	0.1425
σ^2	0.5	-2.7753	-2.3199	-3.4604	0.0060	0.0060	0.0060	0.0059	0.0059
β_0	10.0	0.0687	0.0687	0.0687	0.0661	0.0661	0.0661	0.0661	0.0661
β_1	1.5	-0.6002	-0.5618	14.8637	0.0856	0.0887	0.0887	0.1275	0.1275
β_2	-0.5	3.2476	2.6283	-80.9223	0.1016	0.1082	0.1082	0.2521	0.2521
τ^2	2.0	0.6627	0.3870	-9.9281	0.2921	0.2929	0.2929	0.3062	0.3062
σ^2	2.0	-2.2298	-1.9268	-5.6733	0.0978	0.1046	0.1046	0.0986	0.0986

Table 2

Percentage relative biases and mean squared errors of estimators for $k = 120$.

Parameter	True value	Relative bias (%)			Mean squared error		
		ML	MCML	MAR	ML	MCML	MAR
β_0	10.0	-0.0012	-0.0012	-0.0012	0.0083	0.0083	0.0083
β_1	1.5	0.4208	0.4364	3.9129	0.0093	0.0094	0.0125
β_2	-0.5	-1.7311	-1.5339	-23.6074	0.0113	0.0125	0.0243
τ^2	0.5	-1.0590	-1.1909	-4.4583	0.0096	0.0097	0.0096
σ^2	0.5	-0.5671	-0.1535	-1.5933	0.0027	0.0028	0.0027
β_0	10.0	0.0072	0.0072	0.0072	0.0205	0.0205	0.0205
β_1	1.5	0.5644	0.1712	12.5157	0.0431	0.0451	0.0730
β_2	-0.5	2.6922	2.1282	-68.6275	0.0530	0.0598	0.1577
τ^2	0.5	-2.6772	-3.1427	-11.9506	0.0338	0.0339	0.0338
σ^2	2.0	-0.7001	-0.2308	-4.9868	0.0533	0.0598	0.0545
β_0	10.0	-0.0782	-0.0782	-0.0782	0.0203	0.0203	0.0203
β_1	1.5	0.5844	0.6235	4.9608	0.0103	0.0103	0.0155
β_2	-0.5	-0.2381	-0.2491	-24.4832	0.0126	0.0130	0.0265
τ^2	2.0	-0.4259	-0.5464	-3.8503	0.0898	0.0900	0.0901
σ^2	0.5	-1.1997	-0.8629	-1.9316	0.0030	0.0030	0.0030
β_0	10.0	0.0025	0.0025	0.0025	0.0328	0.0328	0.0328
β_1	1.5	-0.0262	-0.2098	15.5105	0.0401	0.0415	0.0907
β_2	-0.5	1.7727	2.0481	-80.6968	0.0495	0.0533	0.2080
τ^2	2.0	-0.8883	-0.9118	-11.2928	0.1519	0.1521	0.1823
σ^2	2.0	0.1706	0.4896	-3.2453	0.0514	0.0544	0.0481

Table 3

Actual (simulated) and estimated standard errors, and coverage probabilities of MCML estimators.

<i>k</i>	Parameter	Standard error		Coverage probability	
		Actual	Estimated	Coverage	Length
60	β_0	0.1293	0.1277	0.934	0.5006
	β_1	0.1423	0.1446	0.960	0.5667
	β_2	0.1574	0.1376	0.926	0.5393
120	β_0	0.0949	0.0909	0.940	0.3563
	β_1	0.0963	0.1000	0.948	0.3919
	β_2	0.1067	0.1013	0.924	0.3971

Table 4

Empirical levels of significance for Wald tests with MCML estimators. Standard errors (SE's) of the empirical levels are also shown.

<i>k</i>	Empirical level	Actual level (α)		
		0.02	0.05	0.10
60	Level	0.0180	0.0440	0.0780
	SE	0.0059	0.0092	0.0120
120	Level	0.0220	0.0580	0.0980
	SE	0.0066	0.0105	0.0133

Table 5

Counts of observations from patients at seven time-points.

Peptide	Patients	Time (in days)						
		0.5	1	4	7	14	21	28
1	97	75	74	69	62	39	28	29
2	88	48	40	50	49	37	24	28
3	30	14	14	19	16	12	7	9

Table 6

Estimators from proteomics data analysis. (Standard errors in parentheses.)

Parameter	NMAR			MAR		
	Peptide 1	Peptide 2	Peptide 3	Peptide 1	Peptide 2	Peptide 3
β_0	13.920 _(0.139)	14.236 _(0.222)	14.297 _(0.355)	14.008 _(0.110)	14.407 _(0.132)	14.862 _(0.427)
β_1	-0.099 _(0.113)	-0.163 _(0.170)	0.111 _(0.163)	-0.092 _(0.124)	-0.117 _(0.155)	0.086 _(0.473)
β_2	0.183 _(0.126)	0.174 _(0.180)	0.781 _(0.304)	0.211 _(0.128)	0.167 _(0.146)	0.478 _(0.456)
β_3	0.231 _(0.129)	0.368 _(0.154)	1.000 _(0.287)	0.299 _(0.132)	0.379 _(0.146)	0.853 _(0.482)
β_4	0.310 _(0.128)	0.275 _(0.159)	0.475 _(0.262)	0.530 _(0.153)	0.374 _(0.160)	0.460 _(0.535)
β_5	0.262 _(0.121)	0.205 _(0.149)	-0.306 _(0.238)	0.507 _(0.175)	0.425 _(0.185)	0.078 _(0.603)
β_6	0.030 _(0.116)	0.111 _(0.180)	-0.335 _(0.259)	0.124 _(0.173)	0.178 _(0.179)	-0.307 _(0.572)
τ^2	0.454	0.597	1.719	0.409	0.572	1.640
σ^2	0.579	0.491	1.680	0.553	0.454	1.494
γ_0	-5.732	-8.974	-5.594			
γ_1	0.496	0.644	0.392			
γ_2	-0.468	-0.240	-0.193			

Table 7

Inference on proteomics data under two missing data models.

Peptide	Patients	NMAR		MAR	
		Test statistic	p-value	Test statistic	p-value
1	97	24.397	0.00044	26.494	0.00018
2	88	20.265	0.00248	17.668	0.00712
3	30	81.651	0.00000	6.572	0.36227