Nonlinear Mixed-Effects Models with Dropouts & Missing Covariates

October, 2005

Lange Wu

Department of Statistics

University of British Columbia

Vancouver, BC, Canada

Motivation

- Nonlinear mixed-effects models (NLME) are popular in many longitudinal data analyses, such as HIV viral dynamics, pharmacokinetic analyses, and studies of growth and decay.
- In a NLME model, the within individual trajectory is modeled by a nonlinear model, while the between individual variation is incorporated by random effects.
- A NLME model incorporates correlation within repeated measurements, and allows for individual-specific inference.
- Covariates may be introduced to partially explain the between individual variation.

Motivation

Common problems in statistical analyses:

- Some subjects may drop out early, and the dropout may be informative.
- The covariates may contain missing data.

The above two problems may be present simultaneously. So standard complete-case methods are not applicable. We propose likelihood methods which incorporate dropouts and missing covariates *simultaneously*.

Notation

The following notation will be used in this talk:

- $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$: response measurements at times $t_{i1}, \ldots, t_{in_i}$ for individual i.
- $\mathbf{r}_i = (r_{i1}, \ldots, r_{in_i})^T$: missing data indicator for individual i, i.e., $r_{ij} = 1$ if y_{ij} is missing and 0 otherwise.
- $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})^T$: p baseline covariates
- We write $\mathbf{y}_i = (\mathbf{y}_{mis,i}, \mathbf{y}_{obs,i})$, $\mathbf{z}_i = (\mathbf{z}_{mis,i}, \mathbf{z}_{obs,i})$.

The Model and Likelihood

A general NLME model

$$\begin{aligned} y_{ij} &= g(\mathbf{x}_{ij}, \boldsymbol{\beta}_i) + e_{ij}, \qquad \mathbf{e}_i|\boldsymbol{\beta}_i \sim N(\mathbf{0}, \sigma^2 I), && (1) \\ \boldsymbol{\beta}_i &= \mathbf{d}(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i), \qquad \mathbf{b}_i \ \text{ i.i.d.} \sim N(\mathbf{0}, D), && (2) \end{aligned}$$

$$j = 1, \cdots, n_i,\ i = 1, \cdots, N, \qquad (3)$$

Marginal density of $\mathbf{y}_i$

$$f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{v}_i, \boldsymbol{\beta}, \sigma^2, D) = \int f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{v}_i, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) f(\mathbf{b}_i|D) d\,\mathbf{b}_i, \qquad (4)$$

which usually does not have a closed-form expression.

The Model and Likelihood

Monte-Carlo EM Algorithm: A modification of standard EM algorithm, with the E-step being approximated by Monte-Carlo methods (e.g., MCMC methods, rejection sampling methods, importance sampling methods).

We may view the random effects $\mathbf{b}_i$ as additional “missing data”, so we have

Missing data: $(\mathbf{b}_i, \mathbf{z}_{mis,i}, \mathbf{y}_{mis,i})$
Observed data: $(\mathbf{y}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i)$
Complete data: $(\mathbf{y}_i, \mathbf{z}_i, \mathbf{b}_i, \mathbf{r}_i)$

Let $\boldsymbol{\psi} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\eta}, \boldsymbol{\phi})$ denotes the collection of all parameters.

The Model and Likelihood

The observed data likelihood is given by

$$\begin{aligned} L_{obs}(\boldsymbol{\psi}) &= \int\int\int f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{v}_i, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) \\ &\quad \times f(\mathbf{z}_i|\boldsymbol{\alpha}) f(\mathbf{b}_i|D) \;\; d\mathbf{y}_{mis,i} d\mathbf{z}_{mis,i} d\,\mathbf{b}_i, \end{aligned}$$

The above integral is usually intractable.

We consider obtaining the MLEs of $\boldsymbol{\psi}$ by a Monte-Carlo EM algorithm.

A Monte-Carlo EM Algorithm

E-step: compute the conditional expectation of the complete-data log-likelihood given the observed data, i.e.,

$$\begin{aligned}
Q_i(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)}) \;=\; & \iiint \Big\{ \log f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{v}_i, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) \\
& + \log f(\mathbf{z}_i|\boldsymbol{\alpha}) + \log f(\mathbf{b}_i|D) \\
& + \log f(\mathbf{r}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{v}_i, \boldsymbol{\phi})\Big\} \\
& \times f(\mathbf{y}_{mis,i}, \mathbf{z}_{mis,i}, \mathbf{b}_i| \\
& \qquad \mathbf{y}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \boldsymbol{\psi}^{(t)})\; d\mathbf{b}_i\; d\mathbf{z}_{mis,i} d\mathbf{y}_{mis,i}
\end{aligned}$$

which can be evaluated by Monte-Carlo methods.

A Monte-Carlo EM Algorithm

We can generate samples from $f(\mathbf{y}_{mis,i}, \mathbf{z}_{mis,i}, \mathbf{b}_i | \mathbf{y}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{v}_i, \mathbf{r}_i, \boldsymbol{\psi}^{(t)})$ using MCMC methods. E.g., we may use Gibbs sampler to generate from the full conditionals

$$f(\mathbf{y}_{mis,i} | \mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{v}_i, \mathbf{b}_i, \mathbf{r}_i, \boldsymbol{\psi}^{(t)}),$$

$$f(\mathbf{z}_{mis,i} | \mathbf{z}_{obs,i}, \mathbf{y}_i, \mathbf{v}_i, \mathbf{b}_i, \mathbf{r}_i, \boldsymbol{\psi}^{(t)}),$$

$$f(\mathbf{b}_i | \mathbf{y}_i, \mathbf{z}_i, \mathbf{v}_i, \mathbf{r}_i, \boldsymbol{\psi}^{(t)}).$$

Then, we approximate $Q_i(\boldsymbol{\psi} | \boldsymbol{\psi}^{(t)})$ by its empirical mean, with missing data replaced by simulated values.

A Monte-Carlo EM Algorithm

Sampling methods for the E-step:

- Gibbs sampling
- adaptive rejection sampling
- multivariate rejection sampling
- importance sampling.

The M-step is like a complete-data maximization.

Multivariate Rejection Sampling

Consider sampling from $f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \boldsymbol{\psi}^{(t)})$. Let $f^*(\mathbf{b}_i) = f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{b}_i, \boldsymbol{\psi}^{(t)})$ and $\tau = \sup_{\mathbf{u}}\{f^*(\mathbf{u})\}$. A random sample from $f(\mathbf{b}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{r}_i, \boldsymbol{\psi}^{(t)})$ can be obtained as follows.

Step 1: sample $\mathbf{b}_i^*$ from $f(\mathbf{b}_i|D^{(t)})$, and independently, sample w from the uniform(0,1) distribution;

Step 2: if $w \leq f^*(\mathbf{b}_i^*)/\tau$, then accept $\mathbf{b}_i^*$, otherwise, go to step 1.

Importance Sampling

$$Q(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)}) \approx \sum_{i=1}^{N} \left\{ \frac{1}{m_t} \sum_{j=1}^{m_t} w_{ij}^{(t)} l(\boldsymbol{\psi}; \mathbf{y}_{obs,i}, \mathbf{y}_{mis,i}^{*(j)}, \mathbf{z}_{obs,i}, \mathbf{z}_{mis,i}^{*(j)}, \mathbf{b}_i^{*(j)}) \right\}, \quad (5)$$

where

$$w_{ij}^{(t)} = \frac{f(\mathbf{y}_{mis,i}^{*(j)}, \mathbf{z}_{mis,i}^{*(j)}, \mathbf{b}_i^{*(j)} | \mathbf{z}_{obs,i}, \mathbf{y}_{obs,i}, \boldsymbol{\psi}^{(t)})}{h^*(\mathbf{y}_{mis,i}^{*(j)}, \mathbf{z}_{mis,i}^{*(j)}, \mathbf{b}_i^{*(j)})}$$

are importance weights.

A Monte-Carlo EM Algorithm

Variance-covariance matrix of $\boldsymbol{\psi}$

$$I(\hat{\boldsymbol{\psi}}) \approx \sum_{i=1}^{N} \sum_{j=1}^{m_t} \frac{1}{m_i} S_{ij}(\hat{\boldsymbol{\psi}}) S_{ij}^T(\hat{\boldsymbol{\psi}})$$

where

$S_{ij}(\hat{\boldsymbol{\psi}}) = \partial l(\boldsymbol{\theta}; \mathbf{y}_{obs,i}, \tilde{\mathbf{y}}_{mis,i}^{(j)}, \mathbf{z}_{obs,i}, \tilde{\mathbf{z}}_{mis,i}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}) / \partial \boldsymbol{\psi} |_{\boldsymbol{\psi} = \hat{\boldsymbol{\psi}}}$ and $\{(\tilde{\mathbf{y}}_{mis,i}^{(j)}, \tilde{\mathbf{z}}_{mis,i}^{(j)}, \tilde{\mathbf{b}}_i^{(j)}), j = 1, \cdots, m_i\}$ denotes a sample generated from $f(\mathbf{y}_{mis,i}, \mathbf{z}_{mis,i}, \mathbf{b}_i | \mathbf{y}_{obs,i}, \mathbf{z}_{obs,i}, \boldsymbol{\psi}^{(t)})$.

The approximate asymptotic covariance matrix of $\hat{\boldsymbol{\psi}}$ is $I^{-1}(\hat{\boldsymbol{\psi}})$.

A Monte-Carlo EM Algorithm

Potential computational problems:

- very slow convergence
- non-convergence

These problems may occur when the dimension of the “missing data” $(\mathbf{b}_i, \mathbf{z}_{mis,i}, \mathbf{y}_{mis,i})$ is high, especially when the dimension of the random effects $\mathbf{b}_i$ is high.

An Approximate Method

To avoid the potential computational problems with the previous method, we consider an approximate method which may be computationally much more efficient:

- Take a Taylor expansion about $\hat{\mathbf{b}}_i$;
- Convert the NLME model to a linear mixed-effects (LME) model;
- Handle missing data in the LME model;
- Integrate out $\mathbf{b}_i$ in the E-step.

An Approximate Method

We rewrite the NLME model (1) and (3) as a single equation by combining the two stages

$$y_{ij} = g_{ij}(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i) + e_{ij}, \qquad j = 1, \ldots, n_i;\ i = 1, \ldots, N, \tag{6}$$

Iterative solve the following LME model

$$\tilde{\mathbf{y}}_i \;=\; X_i\boldsymbol{\beta} + T_i\mathbf{b}_i + \mathbf{e}_i, \tag{7}$$

where $\tilde{\mathbf{y}}_i = \mathbf{y}_i - \mathbf{g}_i(\mathbf{z}_i, \widehat{\boldsymbol{\beta}}, \widehat{\mathbf{b}}_i) + X_i\widehat{\boldsymbol{\beta}} + T_i\widehat{\mathbf{b}}_i,\ X_{ij} = \partial g_{ij}(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i)/\partial\boldsymbol{\beta}^T,\ T_{ij} = \partial g_{ij}(\mathbf{z}_i, \boldsymbol{\beta}, \mathbf{b}_i)/\partial\mathbf{b}_i^T$.

An Approximate Method

Note that we have

$$f(\widetilde{\mathbf{y}}_{mis,i}, \mathbf{z}_{mis,i}, \mathbf{b}_i \quad | \quad \widetilde{\mathbf{y}}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \widehat{\boldsymbol{\psi}}) = f(\mathbf{b}_i | \widetilde{\mathbf{y}}_i, \mathbf{z}_i, \widehat{\boldsymbol{\psi}}) \\ \times f(\widetilde{\mathbf{y}}_{mis,i}, \mathbf{z}_{mis,i} | \widetilde{\mathbf{y}}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \widehat{\boldsymbol{\psi}}),$$

where

$\widetilde{\mathbf{y}}_{mis,i} = \mathbf{y}_{mis,i} - \mathbf{g}_i(\mathbf{z}_{mis,i}, \widehat{\boldsymbol{\beta}}, \widehat{\mathbf{b}}_i) + X_i(\mathbf{z}_{mis,i})\widehat{\boldsymbol{\beta}} + T_i(\mathbf{z}_{mis,i})\widehat{\mathbf{b}}_i,$

$\widetilde{\mathbf{y}}_{obs,i}$ is defined similarly, and $\widetilde{\mathbf{y}}_i = (\widetilde{\mathbf{y}}_{mis,i}, \widetilde{\mathbf{y}}_{obs,i})$.

An Approximate Method

Under the LME model (7), it is straightforward to show that

$$[\mathbf{b}_i|\tilde{\mathbf{y}}_i, \mathbf{z}_i, \widehat{\boldsymbol{\psi}}] \sim N(\widetilde{\mathbf{b}}_i, \widetilde{\Sigma}_i)$$

where $\widetilde{\Sigma}_i = (\widehat{\sigma}^{-2} T_i^T T_i + \widehat{D}^{-1})^{-1}, \;\; \widetilde{\mathbf{b}}_i = \widetilde{\Sigma}_i T_i^T (\tilde{\mathbf{y}}_i - X_i \widehat{\boldsymbol{\beta}})/\widehat{\sigma}^2$.

An Approximate Method

Then, the E-step of the EM algorithm can be written as

$$\begin{aligned}
Q_i(\boldsymbol{\psi}|\boldsymbol{\psi}^{(t)}) \;=\; & \iiint \Big\{ \log f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{v}_i, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) \\
& \quad + \log f(\mathbf{z}_i|\boldsymbol{\alpha}) + \log f(\mathbf{b}_i|D) \\
& \quad + \log f(\mathbf{r}_i|\mathbf{y}_i, \mathbf{z}_i, \mathbf{v}_i, \boldsymbol{\phi})\Big\} \\
& \quad \times f(\mathbf{y}_{mis,i}, \mathbf{z}_{mis,i}, \mathbf{b}_i| \\
& \qquad \mathbf{y}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \boldsymbol{\psi}^{(t)})\; d\mathbf{b}_i\; d\mathbf{z}_{mis,i} d\mathbf{y}_{mis,i} \\
=\; & I_1 + I_2 + I_3 + I_4, \qquad (8)
\end{aligned}$$

where

An Approximate Method

$$
\begin{aligned}
\tilde{I}_1 &= -\frac{n_i}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\Bigg[\mathrm{tr}(T_i^T T_i \widetilde{\Sigma}_i) + \int \Big(\widetilde{\mathbf{y}}_i - X_i\boldsymbol{\beta} - T_i\widetilde{\mathbf{b}}_i\Big)^T (\,\cdot\,) \\
&\qquad \times f(\widetilde{\mathbf{y}}_{mis,i}, \mathbf{z}_{mis,i} | \widetilde{\mathbf{y}}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \widehat{\boldsymbol{\psi}})\; d\widetilde{\mathbf{y}}_{mis,i} d\mathbf{z}_{mis,i}\Bigg], \\
\tilde{I}_2 &= \int \log f(\mathbf{z}_i|\boldsymbol{\alpha}) f(\mathbf{z}_{mis,i}|\widetilde{\mathbf{y}}_i, \mathbf{z}_{obs,i}, \mathbf{r}_i, \widehat{\boldsymbol{\psi}}) d\mathbf{z}_{mis,i} \\
\tilde{I}_3 &= -\frac{1}{2}\log|D| - \frac{1}{2}\mathrm{tr}(D^{-1}\widetilde{\Sigma}_i) \\
&\qquad -\frac{1}{2}\int \Big(\widetilde{\mathbf{b}}_i^T D^{-1}\widetilde{\mathbf{b}}_i\Big) f(\widetilde{\mathbf{y}}_{mis,i}, \mathbf{z}_{mis,i} | \widetilde{\mathbf{y}}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \widehat{\boldsymbol{\psi}})\; d\widetilde{\mathbf{y}}_{mis,i} d\mathbf{z}_{mis,i}, \\
\tilde{I}_4 &= \int \log f(\mathbf{r}_i|\widetilde{\mathbf{y}}_i, \mathbf{z}_i, \boldsymbol{\phi}) f(\widetilde{\mathbf{y}}_{mis,i}, \mathbf{z}_{mis,i} | \widetilde{\mathbf{y}}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \widehat{\boldsymbol{\psi}})\; d\widetilde{\mathbf{y}}_{mis,i} d\mathbf{z}_{mis,i}.
\end{aligned}
$$

An Approximate Method

Characteristics of the approximate method:

- The integrals in the E-step do not involve the random effects $\mathbf{b}_i$;
- In the E-step we only need to simulate samples from $f(\tilde{\mathbf{y}}_{mis,i}, \mathbf{z}_{mis,i}|\tilde{\mathbf{y}}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \widehat{\boldsymbol{\psi}})$ rather than $f(\mathbf{y}_{mis,i}, \mathbf{z}_{mis,i}, \mathbf{b}_i|\, \mathbf{y}_{obs,i}, \mathbf{z}_{obs,i}, \mathbf{r}_i, \boldsymbol{\psi}^{(t)})$ in the likelihood method.
- The dimension of the “missing data” is reduced to $(\tilde{\mathbf{y}}_{mis,i}, \mathbf{z}_{mis,i})$ from $(\mathbf{y}_{mis,i}, \mathbf{z}_{mis,i}, \mathbf{b}_i)$. So the approximate method is computationally more efficient!

Time-Dependent Covariates

When the covariates are measured repeatedly over time, a multivariate repeated measurement model for the covariates with missing data should be specified.

We may consider a multivariate mixed-effects model for the incompletely observed time-varying covariates

Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \cdots, \mathbf{x}_{iv})$, $\mathbf{x}_i = Vec(\mathbf{X}_i)$ be a vector obtained by stacking all covariates $\{\mathbf{x}_{ij}, j = 1, \cdots, v\}$.

Time-Dependent Covariates

A multivariate LME model for individual i can be written as

$$\mathbf{x}_i = \mathbf{T}_i\boldsymbol{\alpha} + \mathbf{W}_i\mathbf{a}_i + \mathbf{e}_i, \qquad i = 1, 2, \cdots, N,$$

where $\mathbf{T}_i$ and $\mathbf{W}_i$ are block diagonal matrices with the v blocks containing design matrices for each of the v covariates, $\boldsymbol{\alpha}$ is a vector of fixed-effects parameters, $\mathbf{a}_i$ is a vector of random effects, and $\mathbf{e}_i = Vec(\mathbf{e}_{i1}, \mathbf{e}_{i2}, \cdots, \mathbf{e}_{iv})$.

We may assume that $\mathbf{a}_i \sim N(\mathbf{0}, A)$ and $\mathbf{e}_i \sim N(\mathbf{0}, R)$.

Time-Dependent Covariates

The observed data log-likelihood can now be written as

$$\begin{aligned} l_o(\boldsymbol{\beta}, \boldsymbol{\alpha}, D, A, R, \boldsymbol{\psi}, \boldsymbol{\phi}) \; &= \; \sum_{i=1}^{N} \log \Big\{ \int\!\!\int\!\!\int \prod_{j=1}^{n_i} (f(y_{ij}|\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{x}_i) \\ &\quad \times f(\mathbf{x}_i|\mathbf{a}_i, \boldsymbol{\alpha}) f(\mathbf{b}_i|D) f(\mathbf{a}_i|A) f(\mathbf{r}_i|\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi}) \\ &\quad \times f(\mathbf{s}_i|\mathbf{r}_i, \mathbf{x}_i, \boldsymbol{\phi}) \;\; d\mathbf{b}_i d\mathbf{a}_i d\mathbf{x}_{mis,i} d\mathbf{y}_{mis,i} \Big\}. \end{aligned}$$

Time-Dependent Covariates

The “complete data” log-likelihood can now be written as

$$\begin{aligned} l_c^*(\boldsymbol{\gamma}) &= \sum_{i=1}^{N} \Big[\log\{f(\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{b}_i, \mathbf{x}_i)\} + \log\{f(\mathbf{x}_i|\mathbf{a}_i, \boldsymbol{\alpha})\} \\ &\quad + \log\{f(\mathbf{b}_i|D)\} + \log\{f(\mathbf{a}_i|A)\} + \log\{f(\mathbf{r}_i|\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\psi})\} \\ &\quad + \log\{f(\mathbf{s}_i|\mathbf{r}_i, \mathbf{x}_i, \boldsymbol{\phi})\} \Big], \end{aligned}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, A, D, \boldsymbol{\psi}, \boldsymbol{\phi})$. Then a Monte-Carlo EM algorithm can be used to obtain the MLE of $\boldsymbol{\gamma}$.

Time-Dependent Covariates

In this case, we need to generate random samples of $(\mathbf{y}_{mis,i}, \mathbf{x}_{mis,i}, \mathbf{a}_i, \mathbf{b}_i)$ from

$$f(\mathbf{y}_{mis,i}, \mathbf{x}_{mis,i}, \mathbf{a}_i, \mathbf{b}_i | \mathbf{y}_{obs,i}, \mathbf{x}_{obs,i}, \mathbf{r}_i, \mathbf{s}_i, \boldsymbol{\gamma}^{(t)}).$$

This can again be achieved using the Gibbs sampler combined with rejection sampling methods.

The computation here will be more intensive since the dimension of the “missing data” is higher. A more efficient approximate method, which only needs to sample $(\mathbf{y}_{mis,i}, \mathbf{x}_{mis,i})$, can be derived as before.

PX-EM Algorithm

The PX-EM algorithm of Liu, Rubin, and Wu (1998) can be used to speed up the EM algorithms.

The idea is to introduce working parameters to reduce the fraction of missing information thus speed up the algorithms.

The PX-EM introduces working parameters to the original model and then applies the standard EM to the expanded model rather than the original model.

PX-EM Algorithm

We first consider a PX-EM algorithm for the approximate method. The expanded model for the LME model can be written as

$$\tilde{\mathbf{y}}_i = X_i(\mathbf{z}_i)\boldsymbol{\beta} + T_i\,\gamma\,\mathbf{b}_i + \mathbf{e}_i, \qquad i = 1, \ldots, N, \tag{9}$$

where γ is an $(s \times s)$ matrix corresponding to the working parameters.

Implementation of PX-EM: the E-step is unchanged, while the M-step is a simple modification of the original M-step by including the maximization over γ as well.

PX-EM Algorithm

Implementation of a PX-EM for the likelihood method: we may introduce working parameters γ so that the second stage model of the NLME model becomes

$$\boldsymbol{\beta}_i = \mathbf{d}(\mathbf{z}_i, \boldsymbol{\beta}, \gamma \mathbf{b}_i).$$

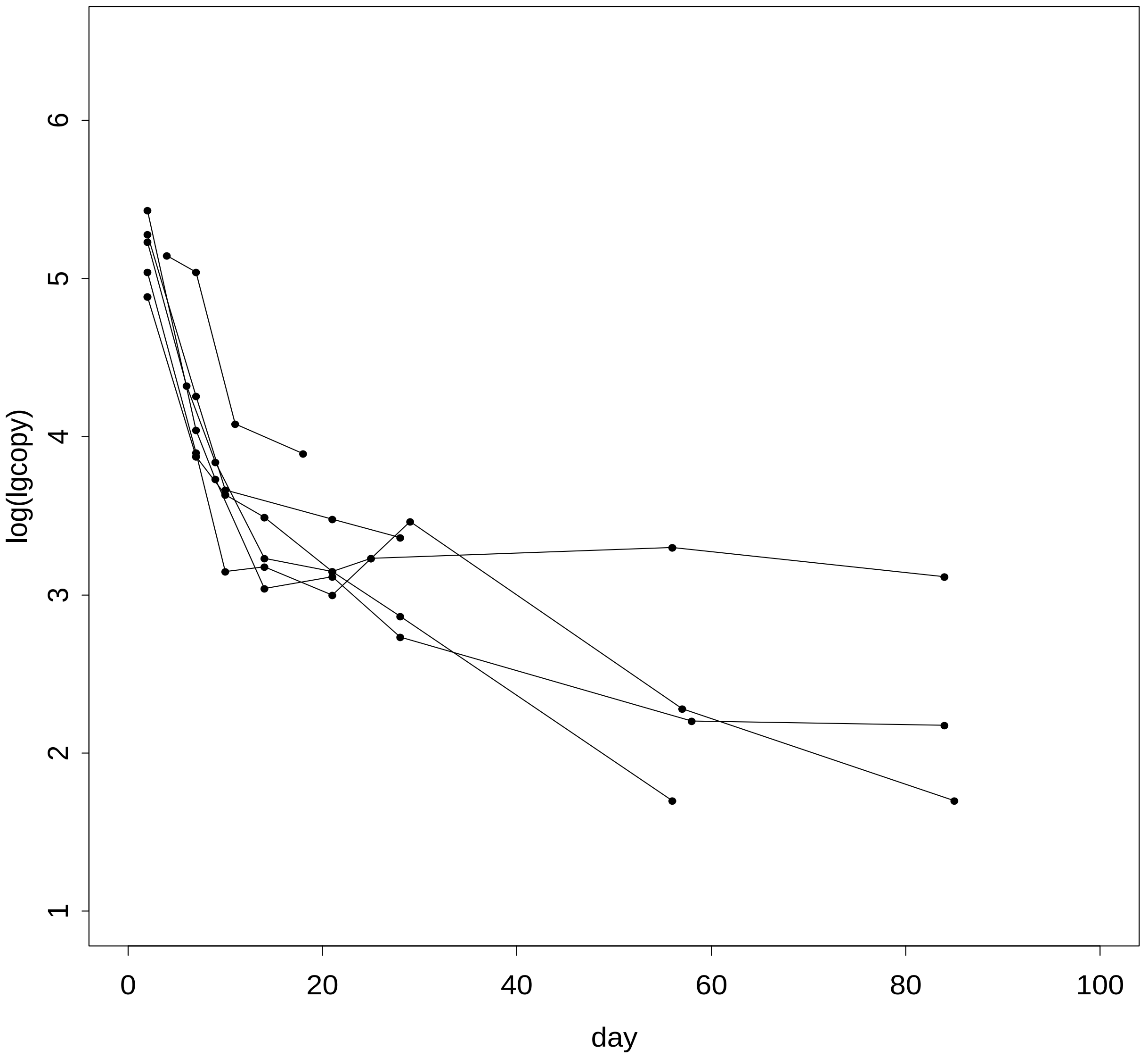
Then we apply the standard EM to the expanded model, which differs from the original model only by the working parameters γ.

An Example

We consider an AIDS dataset, and model the viral load trajectory after an anti-HIV treatment. There are fifty-three HIV-1 infected patients. Five patients dropped out of the study, and sixteen patients have missing viral loads at scheduled time points.

Visual inspection of the raw data seems to indicate that dropout patients appear to have slower viral decay, compared to the remaining patients. Thus, the dropouts are likely to be informative or nonignorable. The covariates CD4, CH50, and TNF contain 20.5%, 19.0%, and 16.7% missing data respectively.

HIV Viral Load



An Example

We consider the following HIV viral dynamic model (Wu and Ding, Biometrics, 1999)

$$\begin{aligned} y_{ij} &= \log_{10}(P_{1i}e^{-\lambda_{1i}t_{ij}} + P_{2i}e^{-\lambda_{2i}t_{ij}}) + e_{ij}, \qquad (10) \\ \log(P_{1i}) &= \beta_1 + \beta_2 TNF + b_{1i}, \qquad \lambda_{1i} = \beta_3 + \beta_4 TNF + \beta_5 CD4 + b_{2i}, \\ \log(P_{2i}) &= \beta_6 + \beta_7 TNF + b_{3i}, \qquad \lambda_{2i} = \beta_8 + \beta_9 CH50 + b_{4i}, \qquad (11) \end{aligned}$$

where y_{ij} is the $\log_{10}$-transformation of the viral load measurement, λ_{1i} and λ_{2i} represent two viral decay rates, and b_{ki}'s are random effects.

An Example

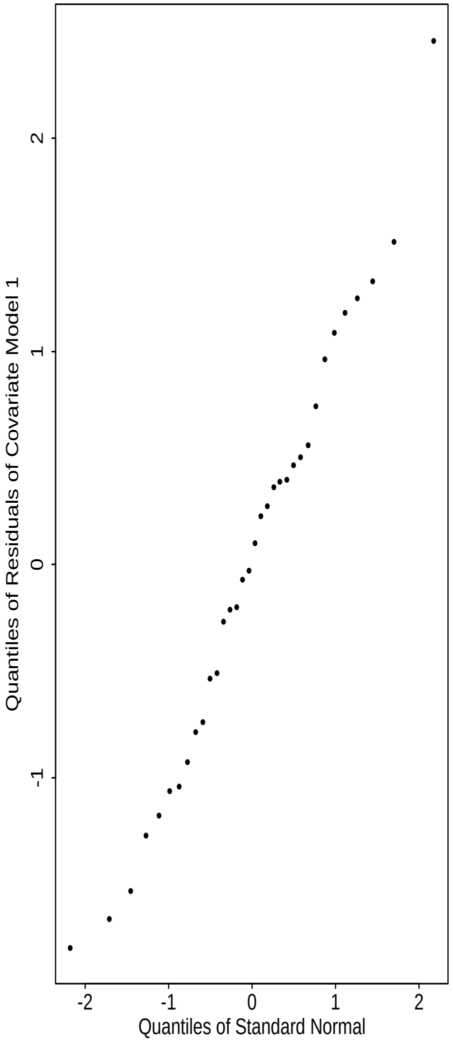
We focus on the following covariate models:

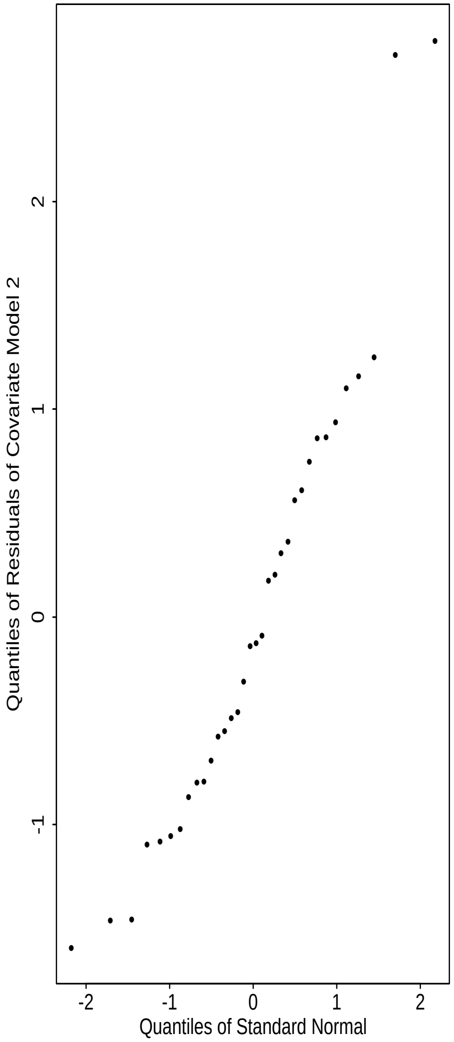
Model 1: $(z_{i1}|\boldsymbol{\alpha}_1) \sim N(\alpha_{10}, \alpha_{11})$,

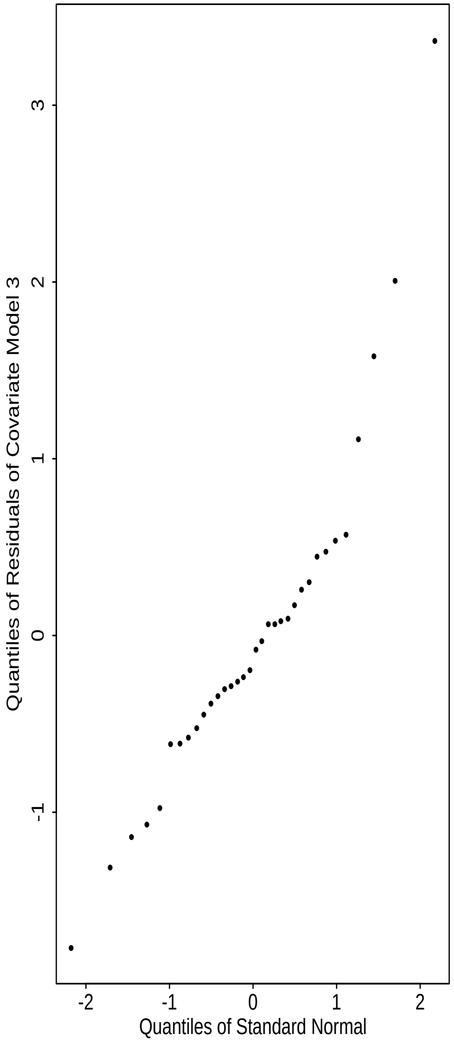
Model 2: $(z_{i2}|z_{i1}, \boldsymbol{\alpha}_2) \sim N(\alpha_{20} + \alpha_{21} z_{i1}, \alpha_{22})$,

Model 3: $(z_{i3}|z_{i1}, z_{i2}, \boldsymbol{\alpha}_3) \sim N(\alpha_{30} + \alpha_{31} z_{i1} + \alpha_{32} z_{i2}, \alpha_{33})$,

where $\boldsymbol{\alpha}_3 = (\alpha_{30}, \alpha_{31}, \alpha_{32}, \alpha_{33})$, $\boldsymbol{\alpha}_2 = (\alpha_{20}, \alpha_{21}, \alpha_{22})$, and $\boldsymbol{\alpha}_1 = (\alpha_{10}, \alpha_{11})$.


Quantiles of Residuals of Covariate Model 1
-1
0
1
2
-2
-1
0
1
2
Quantiles of Standard Normal


Quantiles of Residuals of Covariate Model 2
-1
0
1
2
-2
-1
0
1
2
Quantiles of Standard Normal


Quantiles of Residuals of Covariate Model 3
-1
0
1
2
3
-2
-1
0
1
2
Quantiles of Standard Normal

An Example

We first consider the following dropout model

$$
\begin{aligned}
f(\mathbf{r}|\boldsymbol{\phi}) &\equiv \prod_{i=1}^{N} f(r_{i1}, \cdots, r_{in_i}|\boldsymbol{\phi}) \\
&= \prod_{i=1}^{N} \prod_{j=1}^{n_i} \left[P(r_{ij} = 1|\boldsymbol{\phi})\right]^{r_{ij}} \left[1 - P(r_{ij} = 1|\boldsymbol{\phi})\right]^{1-r_{ij}}, \\
\text{logit}[P(r_{ij} = 1|\boldsymbol{\phi})] &= \phi_0 + \phi_1 y_{i,j-1} + \phi_2 y_{ij} \text{ for } j > 1, \\
&= \phi_0 + \phi_2 y_{ij} \text{ for } j = 1,
\end{aligned}
$$

Table 1. Estimates for Viral Dynamic Parameters.

	Like. (nonig.)		Like. (ignor.)		Appr. (nonig.)		C-C	
Parameter	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
β_1	12.17	0.09	12.01	0.10	11.98	0.11	12.47	0.20
β_2	0.66	0.15	0.71	0.16	0.71	0.14	0.75	0.20
β_3	33.14	1.47	35.21	1.41	34.61	1.33	37.87	2.60
β_4	4.98	2.36	5.57	1.10	5.53	1.95	6.41	3.29
β_5	4.84	0.93	5.33	0.44	5.31	0.25	5.19	1.99
β_6	7.34	0.09	7.08	0.07	7.07	0.13	7.65	0.28
β_7	0.55	0.08	0.58	0.08	0.58	0.03	0.56	0.17
β_8	1.90	0.13	1.65	0.06	1.62	0.23	2.15	0.54
β_9	0.15	0.11	0.12	0.05	0.13	0.05	0.14	0.27

An Example

Data analysis results:

- The likelihood methods under nonignorable and ignorable missing data mechanisms and the complete-case method produce somewhat different results.
- The estimate of the initial viral decay rate β_3 is the smallest using the likelihood method under nonignorable missing data mechanism and is the largest using the complete-case method.
- The approximate method and the likelihood method produce similar estimates.

Sensitivity Analysis

It is important to check the sensitivity of parameter estimates to various plausible dropout models.

we consider the following plausible dropout models

Model I: logit$[P(r_{ij} = 1|\boldsymbol{\phi})] = \phi_0 + \phi_1 CD4 + \phi_2 TNF + \phi_3 CH50 + \phi_4 y_{ij},$

Model II: logit$[P(r_{ij} = 1|\boldsymbol{\phi})] = \phi_0 + \phi_1 y_{i,j-2} + \phi_2 y_{i,j-1} + \phi_3 y_{ij},$

Model III: logit$[P(r_{ij} = 1|\boldsymbol{\phi})] = \phi_0 + \phi_1 y_{ij} + \phi_2 y_{ij}^2.$

We find that the resulting parameter estimates for the viral dynamic parameters $\boldsymbol{\beta}$ are all similar.

Conclusions

- Statistical analyses ignoring dropouts or missing covariates may lead to misleading results, such as over-estimating the initial viral decay rate.
- The current or immediate previous viral load values may be most predictive for patients' dropout.
- The likelihood method may be preferable when it does not exhibit computational difficulties.
- When the likelihood method exhibits computational problems such as slow or non convergence, the approximate method may be preferable.

Conclusions

- The approximate method may avoid some computational difficulties such as slow or non convergence, and is also computationally more efficient.
- Estimates based on the approximate method may be excellent starting values for the likelihood method.
- We should avoid too complicated dropout models since these models may be non-identifiable.