UNIVERSIDADE ESTADUAL DE CAMPINAS
UNICAMP

IMECC

VIII Workshop on Probabilistic and Statistical Methods

A semiparametric mixed-effects model for censored longitudinal data

Larissa Avila Matos

Co-Authors: Thalita B. Mattos and Victor H. Lachos

February 13, 2020

Contents

Contents

Introduction

Linear and nonlinear mixed-effects (LME/NLME) models have been extensively studied in the literature and applied to analyze longitudinal data.

The classical LME model is often written in the following form:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i,$$

where $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$, $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$, $\quad i = 1, \ldots, n$, with $\mathbf{b}_i \perp \boldsymbol{\epsilon}_i$.

One difficulty that arises in longitudinal data analysis is when the response is censored for some of the observations.

- For example: HIV studies, where the detection of the viral load in the blood compartment is often limited by the sensitivity of a laboratory assay.

Introduction

Several statistical approaches have been developed to deal with longitudinal data with censored measurements in the LME framework:

- Hughes (1999): Monte Carlo EM (MCEM) for LME with censored responses (LMEC).
- Vaida and Liu (2009): EM algorithm for LME/NLME models with censored responses, which uses closed-form expressions at the E-step (LMEC/NLMEC).
- Matos et al. (2013): EM algorithm for LMEC/NLMEC based on the multivariate Student-t distribution, named t-LMEC/t-NLMEC.
- Lachos et al. (2019): a robust multivariate linear mixed model for multiple censored responses based on the class of SMN distributions.

Introduction

Semiparametric models:

- Assumption for LME models: the response variable is a known parametric function of both fixed-effects and random-effects.

- Nonparametric regression: no assumptions about the functional form, letting the data "speak for themselves" in determining the estimated trend.

- Nonparametric regression can also be combined with parametric models to form hybrid **semiparametric models**.

- In semiparametric models, the parametric components are often used to model important factors that affect the response and the nonparametric component is often used for nuisance factors.

Introduction

- Zeger and Diggle (1994) proposed a semiparametric model where a nonparametric function is used to model the time effect, and a random intercept together with a Gaussian stochastic process is used to account for the within-subject correlation.

- Vock et al. (2011) developed a mixed model framework for censored longitudinal data in which the random effects are represented by the flexible seminonparametric (SNP) density.

Goal: The aim of this work is to perform a study of statistical inference in the semiparametric mixed effects models for longitudinal irregularly observed censored data (SMEC).

Contents

Motivating examples

In this work we present two motivating examples from AIDS research.

1. ACTG 315 study; and
2. A5055 study.

ACTG 315 study

The case study:

- The AIDS Clinical Trials Group (ACTG) protocol 315 considers 46 HIV-1 infected patients treated with a potent antiretroviral regimen.
- Before initiating the antiretroviral regimen, all patients discontinued their own antiretroviral regimen for five weeks as a "washout" period.
- The aim of this antiretroviral regimen is to show that immunity can be partially restored in people with moderately advanced HIV disease.

ACTG 315 study

The dataset:

- The viral load was quantified irregularly on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 and 196 after start of treatment, generating 361 observations.
- $CD4^+$ cell counts were also measured along with viral loads.
- Measurements below the detectable threshold of 100 copies/mL (40 out of 361, 11%) were considered left-censored.
- The number of measurements per subject varied from 4 to 10.
- For a more detailed description of the HIV/AIDS study, see Kotzin et al. (2000).

ACTG 315 study

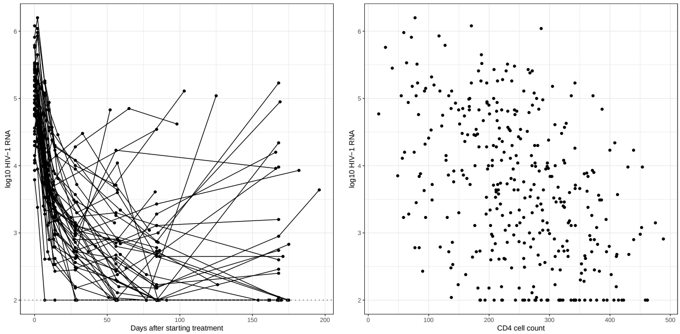


Figure: **ACTG 315 study**. Individual profiles for HIV viral load (in $\log_{10}$ scale) at different follow-up times.Scatter plot of the $CD4^{+}$ cell counts against viral loads (in $\log_{10}$ scale)

ACTG 315 study

This dataset was previously analyzed by Matos et al. (2016) using a biphasic nonlinear model adopting a DEC structure for the error term (DEC-NLMEC).

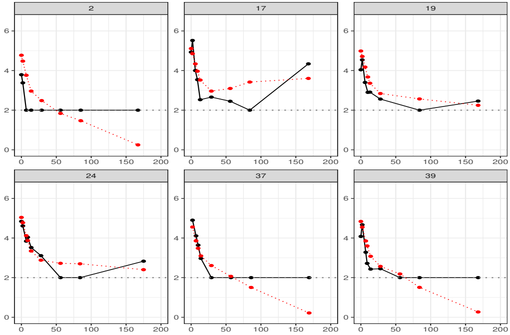


Figure: **ACTG 315 study**. Profiles for HIV viral load (in $\log_{10}$ scale) for 6 randomly chosen subjects and estimated trajectories (dotted line) in the DEC-NLMEC model.

A5055 study

The case study:

- The ACTG protocol A5055 was a phase I/II, randomized, open-label, 24-week comparative study of the pharmacokinetics, tolerability, safety and antiretroviral effects of two regimens of indinavir, ritonavir and two nucleoside analogue reverse transcriptase inhibitors on HIV-1 infected patients.
- ARV therapies:
  - Treatment 1: IDV 800 mg twice daily (q12h) plus RTV 200 mg q12h,
  - Treatment 2: IDV 400 mg q12h plus RTV 400 mg q12h.
- In AIDS research, the number of RNA copies (viral load) in blood plasma and its evolutionary trajectories play a prominent role in the diagnosis of HIV-1 disease progression after an ARV treatment regimen.

A5055 study

The dataset:

- 44 infected patients with the human immunodeficiency virus type 1 (HIV-1).
- These patients were treated with one of two potent ARV therapies.
- The viral load ($\log_{10}$(RNA)) was quantified irregulary on days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up.
- CD4 and CD8, two immunologic markers frequently used to monitor disease progression in AIDS studies, were also measured along with the viral load.

  33.5% (106 out of 316) of measurements lies below the limits (50 copies/mL) of assay quantification (left-censored).
- A more detailed description of this study and data can be found in Acosta et al. (2004)

A5055 study

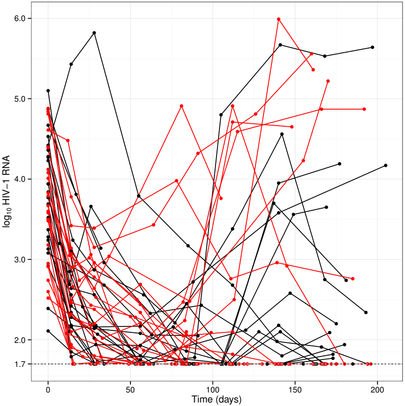


Figure: **A5055 study**. Individual profiles for HIV viral load (in log10 scale) at different follow-up times. Black lines indicate patients under treatment 1 and red lines indicate patients under treatment 2.

A5055 study

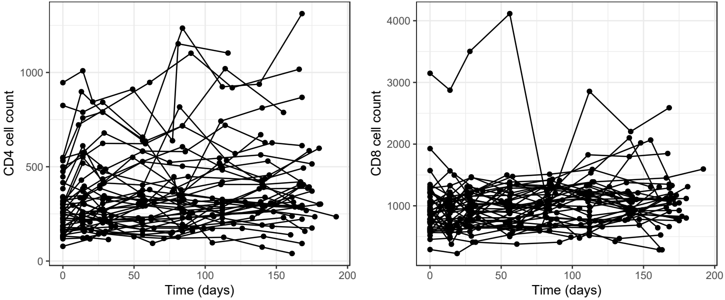


Figure: **A5055 study**. Individual profiles for CD4+ and CD8+ cell count at different follow-up times.

Contents

The EM Algorithm

Dempster et al. (1977)

Let $\boldsymbol{\theta}$ be the parameter vector and $\boldsymbol{y}_c = (\mathbf{y}^\top, \boldsymbol{q}^\top)$ be the vector of complete data, i.e., the observed data $\mathbf{y}^\top$ and the missing/censored data (or the latent variables, depending on the situation) $\boldsymbol{q}^\top$. The EM algorithm consists basically of two steps: the expectation (E-step) and the maximization (M-step).

- **E-Step:** Calculate the conditional expectation

$$Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}) = E\left[\ell_c(\boldsymbol{\theta} \mid \mathbf{y}_c) \mid \mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}\right],$$

  where $\widehat{\boldsymbol{\theta}}^{(k)}$ is the estimate of $\boldsymbol{\theta}$ at the k-th iteration.

- **M-Step:** Update $\boldsymbol{\theta}^{(k)}$ according to

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \arg\max \theta Q(\boldsymbol{\theta} \mid \widehat{\boldsymbol{\theta}}^{(k)}).$$

Correlation structures

DEC - Munoz et al. (1992)

Damped exponential correlation (DEC):

$$\mathbf{E}_i = \mathbf{E}_i(\phi, \mathbf{t}_i) = \left[\phi_1^{|t_{ij} - t_{ik}|^{\phi_2}} \right], \; i = 1, \ldots, n, \; j, k = 1, \ldots, n_i, \qquad (1)$$

For the DEC structure, we have that:

(a) if $\phi_2 = 0$, then $\mathbf{E}_i$ generates the compound symmetry correlation structure;

(b) when $0 < \phi_2 < 1$, then $\mathbf{E}_i$ presents a decay rate between the compound symmetry structure and the first-order AR (AR (1)) model;

(c) if $\phi_2 = 1$, then $\mathbf{E}_i$ generates an AR(1) structure;

(d) when $\phi_2 > 1$, $\mathbf{E}_i$ presents a decay rate faster than the AR(1) structure; and

(e) if $\phi_2 \to \infty$, then $\mathbf{E}_i$ represents the first-order moving average model, MA(1).

Contents

The model

The semiparametric mixed-effects model is specified as follows :

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{N}_i\mathbf{f} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n; \tag{2}$$

* $\mathbf{b}_i \overset{\text{iid.}}{\sim} \mathrm{N}_q(\mathbf{0}, \mathbf{D})$ is independent of $\boldsymbol{\epsilon}_i \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \boldsymbol{\Omega}_i)$, $i = 1, \dots, n$;
* $\mathbf{f} = (f(t_1^0), \dots, f(t_r^0))^\top$ is an $r \times 1$ vector with $t_1^0, \dots, t_r^0$ being the distinct and ordered values of t_{ij}, with $f(\cdot)$ a smooth function of time t_{ij};
* $\mathbf{N}_i$ is an $(n_i \times r)$ incidence matrix whose (j, s)-th element equals the indicator function $\mathbb{I}(t_{ij} = t_s^0)$ for $j = 1, \dots, n_i$ and $s = 1, \dots, r$;
* $\mathbf{D} = \mathbf{D}(\boldsymbol{\alpha})$ models between-subjects variability;
* $\boldsymbol{\Omega}_i = \sigma^2\mathbf{E}_i$ is the correlation structure of the error vector, where the $n_i \times n_i$ matrix $\mathbf{E}_i$ incorporates a time-dependence structure.

The model

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_n^\top)$, $\mathbf{N} = (\mathbf{N}_1^\top, \ldots, \mathbf{N}_n^\top)$, and $\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$.

Then, the model (2) can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{f} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \tag{3}$$

where

$$\begin{aligned} \mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top &\sim \mathrm{N}_{nq}(\mathbf{0}, \mathcal{D}(\boldsymbol{\alpha})) \quad \text{and} \\ \boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^\top, \ldots, \boldsymbol{\epsilon}_n^\top)^\top &\sim \mathrm{N}_N(\mathbf{0}, \boldsymbol{\Omega}), \end{aligned}$$

with $\mathcal{D}(\boldsymbol{\alpha}) = \mathrm{diag}(\mathbf{D}, \ldots, \mathbf{D})$ and $\boldsymbol{\Omega} = \mathrm{diag}(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_n)$.

The matrix $[\mathbf{X}, \mathbf{N}\mathbf{T}]$ is of full column rank, where $\mathbf{T} = [\mathbf{1}, \mathbf{t}^0]$ and $\mathbf{1}$ is an $r \times 1$ vector of 1's.

λ

The model

We assume that the response y_{ij} is not fully observed for all i, j.

Let the observed data for the i-th subject be $(\mathbf{V}_i, \mathbf{C}_i)$, where

- $\mathbf{V}_i$ represents the vector of uncensored readings or censoring level,
- $\mathbf{C}_i$ is the vector of left-censoring indicators,

such that

$$\begin{aligned} y_{ij} &\leq V_{ij} \quad \text{if} \quad C_{ij} = 1, \\ y_{ij} &= V_{ij} \quad \text{if} \quad C_{ij} = 0. \end{aligned} \tag{4}$$

The model defined in (2)-(4) is henceforth called the DEC-SMEC model.

The log-likelihood function

Following Vaida and Liu (2009), classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \mathbf{f}^\top, \sigma^2, \boldsymbol{\alpha}^\top, \boldsymbol{\phi}^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$.

For complete data, we have marginally that $\mathbf{y}_i \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where

$$\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{N}_i\mathbf{f} \quad \text{and} \quad \boldsymbol{\Sigma}_i = \boldsymbol{\Omega}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top.$$

For responses with censoring pattern as in (4), we have

$$\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i \sim \mathrm{TN}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i; \mathbb{A}),$$

where $\mathrm{TN}_{n_i}(.; \mathbb{A})$ denotes the truncated normal distribution on the interval $\mathbb{A}$, where $\mathbb{A}_i = A_{i1} \times \ldots \times A_{in_i}$, with

- $A_{ij} = (-\infty, \infty)$, if $C_{ij} = 0$;
- $A_{ij} = (-\infty, V_{ij}]$, if $C_{ij} = 1$.

The log-likelihood function

Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i with $(n_i = n_i^o + n_i^c)$ such that $C_{ij} = 0$ for all elements in $\mathbf{y}_i^o$, and 1 for all elements in $\mathbf{y}_i^c$.

The likelihood function for subject i (using conditional probability arguments) is given by:

$$
\begin{aligned}
L_i(\boldsymbol{\theta}) = f(\mathbf{y}_i|\boldsymbol{\theta}) &= P(\mathbf{V}_i|\mathbf{C}_i, \boldsymbol{\theta}) \\
&= f(\mathbf{y}_i^o|\boldsymbol{\theta})P(\mathbf{y}_i^c \leq \mathbf{V}_i^c|\mathbf{V}_i^o, \boldsymbol{\theta}) \\
&= \phi_{n_i^o}(\mathbf{y}_i^o; \boldsymbol{\mu}_i^o\boldsymbol{\beta}, \boldsymbol{\Sigma}_i^{oo})\Phi_{n_i^c}(\mathbf{V}_i^c; \boldsymbol{\mu}_{ico}, \mathbf{S}_i) = L_i. \qquad (5)
\end{aligned}
$$

The log-likelihood function for the observed data is thus given by $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}|\mathbf{y}) = \sum_{i=1}^{n}\{\log L_i\}$.

The log-likelihood function

However, maximization of $\ell(\boldsymbol{\theta})$ without imposing restrictions on the function $\mathbf{f}(\cdot)$ may cause over-fitting and non-identification of $\boldsymbol{\beta}$ (Green, 1987).

A well-known procedure that is based on the idea of log-likelihood penalization consists of incorporating a penalty function in the log-likelihood, such that:

$$\ell_p(\boldsymbol{\theta}, \lambda) = \ell(\boldsymbol{\theta}|\mathbf{y}) - \frac{\lambda}{2} J(\mathbf{f}), \tag{6}$$

where

- $J(\mathbf{f})$ denotes the penalty function over $\mathbf{f}(\cdot)$;
- λ is a smoothing parameter that controls the tradeoff between goodness of fit and the smoothness estimated function.

By maximizing (6), one obtains the MPL estimates.

Inference

The complete-data log-likelihood function

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{V} = \mathrm{vec}(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = \mathrm{vec}(\mathbf{C}_1, \ldots, \mathbf{C}_n)$, where $(\mathbf{V}_i, \mathbf{C}_i)$ is observed for the ith subject. So,

- missing data: $\mathbf{b}$ and $\mathbf{y}$;
- observed data: $\mathbf{V}$ and $\mathbf{C}$;
- complete data: $\mathbf{y}_{\mathrm{com}} = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top$

The complete-data log-likelihood function is given by

$$\ell_{\mathrm{c}}(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{com}}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{com}}),$$

where

$$\begin{aligned}\ell_i(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{com}}) \quad = \quad & -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{E}_i|) - \frac{1}{2\sigma^2}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i)^\top \mathbf{E}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i - \mathbf{Z}_i\mathbf{b}_i) \\ & -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i + C, \qquad (7)\end{aligned}$$

with C being a constant independent of the parameter vector $\boldsymbol{\theta}$.

The EM algorithm

Q-function

Given the complete-data log-likelihood function, the Q-function can be written as:

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= \mathbb{E}\left[\ell_c(\boldsymbol{\theta}|\mathbf{y}_{\mathrm{com}})|\mathbf{V},\mathbf{C},\widehat{\boldsymbol{\theta}}^{(k)}\right] \\
&= \sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) \\
&= \sum_{i=1}^{n} Q_{1i}(\boldsymbol{\beta},\mathbf{f},\sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^{n} Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}),
\end{aligned}
$$

where

$$
\begin{aligned}
Q_{1i}(\boldsymbol{\beta},\mathbf{f},\sigma^2|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{n_i}{2}\log\sigma^2 - \frac{1}{2}\log(|\mathbf{E}_i|) - \frac{1}{2\sigma^2}\left[\widehat{a}_i^{(k)} - 2\boldsymbol{\mu}_i^\top\mathbf{E}_i^{-1}\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\right. \\
&+ \left.\boldsymbol{\mu}_i^\top\mathbf{E}_i^{-1}\boldsymbol{\mu}_i\right]
\end{aligned}
$$

and

$$
Q_{2i}(\boldsymbol{\alpha}|\widehat{\boldsymbol{\theta}}^{(k)}) = -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{D}^{-1}\right).
$$

The EM algorithm

E-Step

- E-Step: Calculate the conditional expectation:

$$
\begin{aligned}
\widehat{a}_i^{(k)} &= \operatorname{tr}\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)}\mathbf{E}_i^{-1} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathbf{E}_i^{-1} + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top\mathbf{E}_i^{-1}\mathbf{Z}_i\right), \\
\widehat{\mathbf{b}_i}^{(k)} &= \mathbb{E}\left[\mathbf{b}_i\middle|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right] = \boldsymbol{\varphi}_i\left(\widehat{\mathbf{y}_i}^{(k)} - \boldsymbol{\mu}_i\right), \\
\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{b}_i\mathbf{b}_i^\top\middle|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right] = \boldsymbol{\Lambda}_i + \boldsymbol{\varphi}_i\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - 2\widehat{\mathbf{y}_i}^{(k)}\boldsymbol{\mu}_i + \boldsymbol{\mu}_i\boldsymbol{\mu}_i^\top\right)\boldsymbol{\varphi}_i^\top, \\
\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{y}_i\mathbf{b}_i^\top\middle|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right] = \left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}_i}^{(k)}\boldsymbol{\mu}_i^\top\right)\boldsymbol{\varphi}_i^\top, \\
\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} &= \mathbb{E}\left[\mathbf{y}_i\mathbf{y}_i^\top|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right], \\
\widehat{\mathbf{y}_i}^{(k)} &= \mathbb{E}\left[\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right],
\end{aligned}
$$

with $\boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top\mathbf{E}_i^{-1}\mathbf{Z}_i/\sigma^2)^{-1}$ and $\boldsymbol{\varphi}_i = \boldsymbol{\Lambda}_i\mathbf{Z}_i^\top\mathbf{E}_i^{-1}/\sigma^2$.

The EM algorithm

E-Step

- Following Green (1987), the MPL estimate of $\boldsymbol{\theta}$ is the value that maximizes the function

$$Q_p(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) - \frac{\lambda}{2}J(\mathbf{f}), \tag{8}$$

  where $J(\mathbf{f})$ and λ are as defined in (6) and $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ is the complete data log-likelihood function.

- Similarly to Ibacache-Pulgar et al. (2013), we will consider the following penalty function:

$$J(\mathbf{f}) = \int_a^b [f''(t)]^2 dt = \mathbf{f}^\top \mathbf{K}\mathbf{f},$$

  where $[f''(t)]$ denotes the second derivative of $f(t)$ with $[a, b]$ containing the values t_j^0, of $j = 1, \ldots, r$.

The EM algorithm

CM-Step

- CM-step: Update $\widehat{\boldsymbol{\theta}}^{(k)}$ by the maximization of $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$, which leads to the following expressions:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^{n} \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \left(\widehat{\mathbf{y}_i}^{(k)} - \mathbf{N}_i \widehat{\mathbf{f}}^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}\right), \\
\widehat{\mathbf{f}}^{(k+1)} &= \left(\sum_{i=1}^{n} \mathbf{N}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \mathbf{N}_i + \widehat{\sigma^2}^{(k)} \lambda \mathbf{K}\right)^{-1} \sum_{i=1}^{n} \mathbf{N}_i^\top \widehat{\mathbf{E}}_i^{-1(k)} \left(\widehat{\mathbf{y}_i}^{(k)} - \mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}\right), \\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N} \sum_{i=1}^{n} \Big[\widehat{a}_i^{(k)} - 2(\mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} + \mathbf{N}_i \widehat{\mathbf{f}}^{(k+1)})^\top \widehat{\mathbf{E}}_i^{-1(k)} (\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}) \\
&+ (\mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} + \mathbf{N}_i \widehat{\mathbf{f}}^{(k+1)})^\top \widehat{\mathbf{E}}_i^{-1(k)} (\mathbf{X}_i \widehat{\boldsymbol{\beta}}^{(k+1)} + \mathbf{N}_i \widehat{\mathbf{f}}^{(k+1)})\Big], \\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathbf{b}_i \mathbf{b}_i^\top}^{(k)}, \\
\widehat{\phi}^{(k+1)} &= \underset{\phi \in (0,1) \times \mathcal{R}+}{\arg\max} \left(-\frac{1}{2} \log(|\mathbf{E}_i|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}} \Big[\widehat{a}_i^{(k)} - 2\widehat{\boldsymbol{\mu}}_i^{(k+1)\top} \mathbf{E}_i^{-1} \left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i \widehat{\mathbf{b}}_i^{(k)}\right) \right. \\
&+ \left. \widehat{\boldsymbol{\mu}}_i^{(k+1)\top} \mathbf{E}_i^{-1} \widehat{\boldsymbol{\mu}}_i^{(k+1)}\Big]\right),
\end{aligned}
$$

where $N = \sum_{i=1}^{n} n_i$.

Approximate standard errors

In the context of nonparametric regression, the covariance matrix of the MPL estimates can be evaluated by inverting the observed information matrix obtained by treating the penalized likelihood as a usual likelihood (Segal et al., 1994).

Within the framework of censoring, the variance of the parameter estimates can be obtained using the missing information principle (Louis, 1982), according which:

$$observed\ information = complete\ information - missing\ information.$$

Approximate standard errors

Following Segal et al. (1994) and Louis (1982), we derive the covariance matrix of $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}})$ by using the inverse of the penalized observed information matrix.

Thus, the approximate covariance matrix of $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}})$ is given as:

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{f}}) \approx \mathcal{I}_p^{-1}(\boldsymbol{\beta}, \mathbf{f})\Big|_{\widehat{\boldsymbol{\theta}}},$$

where the penalized expected information matrix $\mathcal{I}_p(\boldsymbol{\beta}, \mathbf{f})$ takes the form:

$$\mathcal{I}_p(\boldsymbol{\beta}, \mathbf{f}) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}} \\ \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}^{\top} & \mathcal{I}_{\mathbf{f}\mathbf{f}} \end{pmatrix}. \tag{9}$$

Approximate standard errors

Thus, we obtain the variance of $\hat{\beta}$ and $\hat{\mathbf{f}}$ estimated at convergence, respectively, as:

$$
\begin{aligned}
\widehat{\mathrm{Var}}_{\mathrm{approx}}(\widehat{\boldsymbol{\beta}}) &= \left(\mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}\mathcal{I}_{\mathbf{ff}}^{-1}\mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}^{\top}\right)\Big|_{\widehat{\boldsymbol{\theta}}}, \\
\widehat{\mathrm{Var}}_{\mathrm{approx}}(\widehat{\mathbf{f}}) &= \left(\mathcal{I}_{\mathbf{ff}} - \mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}^{\top}\mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\mathcal{I}_{\boldsymbol{\beta}\mathbf{f}}\right)\Big|_{\widehat{\boldsymbol{\theta}}},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \sum_{i=1}^{n}\left\{\mathbf{X}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i - \mathbf{X}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\mathrm{Var}\left[\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i\right]\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right\}, \\
\mathcal{I}_{\boldsymbol{\beta}\mathbf{f}} &= \sum_{i=1}^{n}\left\{\mathbf{X}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i - \mathbf{X}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\mathrm{Var}\left[\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i\right]\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i + \lambda\mathbf{X}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\left(\widehat{\mathbf{y}}_i - \boldsymbol{\mu}_i\right)\mathbf{f}^{\top}\mathbf{K}\right\}, \\
\mathcal{I}_{\mathbf{ff}} &= \sum_{i=1}^{n}\left\{\mathbf{N}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i + \lambda\mathbf{K} - \mathbf{N}_i^{\top}\boldsymbol{\Sigma}_i^{-1}\mathrm{Var}\left[\mathbf{y}_i|\mathbf{V}_i,\mathbf{C}_i\right]\boldsymbol{\Sigma}_i^{-1}\mathbf{N}_i\right. \\
&+ \left. 2\lambda\mathbf{N}_i^{\top}\boldsymbol{\Sigma}_i^{\top}\left(\widehat{\mathbf{y}}_i - \boldsymbol{\mu}_i\right)\mathbf{f}^{\top}\mathbf{K} + \lambda^2\mathbf{K}\mathbf{f}\mathbf{f}^{\top}\mathbf{K}\right\}.
\end{aligned}
$$

Note that when $\mathbf{f} = \mathbf{0}$, we obtain the variance of the fixed effects in the approximate ML estimation given by Vaida and Liu (2009) and Hughes (1999).

Contents

Estimation of the smoothing parameter

Several authors have shown the connection between a smoothing spline and a linear mixed-effects model for analysis of longitudinal data (see, for instance, Speed, 1991; Wang, 1998).

Zhang et al. (1998) formulated the semiparametric mixed model defined in (3) as a modified LME model and proposed to estimate the smoothing parameter λ and the variance component simultaneously using REML.

Following Green (1987) and Zhang et al. (1998), we can write $\mathbf{f}$ via a one-to-one linear transformation as:

$$\mathbf{f} = \mathbf{T}\delta + \mathbf{B}\mathbf{d}, \tag{10}$$

where $\boldsymbol{\delta}$ and $\mathbf{d}$ are vectors with dimensions 2 and $r-2$, $\mathbf{B} = \mathbf{L}(\mathbf{L}^\top\mathbf{L})^{-1}$ and $\mathbf{L}$ is an $r \times (r-2)$ full-rank matrix satisfying $\mathbf{K} = \mathbf{L}\mathbf{L}^\top$ and $\mathbf{L}^\top\mathbf{T} = 0$.

Estimation of the smoothing parameter

Given (10), Equation (3) can be reformulated as:

$$\mathbf{y} = \mathbf{X}_* \boldsymbol{\beta}_* + \mathbf{Z}_* \mathbf{b}_* + \boldsymbol{\epsilon},$$

where

- $\mathbf{X}_* = [\mathbf{X}, \mathbf{NT}]$;
- $\mathbf{Z}_* = [\mathbf{NB}, \mathbf{Z}]$;
- $\boldsymbol{\beta}_* = (\boldsymbol{\beta}^\top, \boldsymbol{\delta}^\top)^\top$ are the regression coefficients;
- $\mathbf{b}_* = (\mathbf{d}^\top, \mathbf{b}^\top)^\top$ are mutually independent random effects, with $\mathbf{d} \sim \mathrm{N}(\mathbf{0}, \frac{\sigma^2}{\lambda}\mathbf{I}_{r-2})$; and
- $\mathbf{b}$ and $\boldsymbol{\epsilon}$ have the same distributions as those given in (2).

Estimation of the smoothing parameter

Consider the following model:

$$\begin{aligned}\mathbf{y}|\mathbf{b}_* &\sim \mathrm{N}_N\left(\mathbf{X}_*\boldsymbol{\beta}_* + \mathbf{Z}_*\mathbf{b}_*, \boldsymbol{\Omega}\right),\\ \mathbf{b}_* &\sim \mathrm{N}_{(r-2+q)\times 1}\left(\mathbf{0}, \boldsymbol{\Psi}\right), \quad \text{where} \quad \boldsymbol{\Psi} = \begin{pmatrix} \frac{\sigma^2}{\lambda}\mathbf{I}_{r-2} & \mathbf{0} \\ \mathbf{0} & \mathcal{D}(\boldsymbol{\alpha}) \end{pmatrix}.\end{aligned}$$

In order to use the EM algorithm, we consider the augmented data vector $\mathbf{y}_{comp*} = (\mathbf{y}^\top, \mathbf{b}_*^\top)^\top$, where $\mathbf{b}_*$ is assumed to be the missing variable.

The complete-data log-likelihood function dropping all the terms that are not functions of λ, takes the form:

$$\ell(\lambda; \mathbf{y}_{comp*}) \propto -\frac{1}{2}\log|\boldsymbol{\Psi}| - \frac{1}{2}\mathbf{b}_*^\top\boldsymbol{\Psi}^{-1}\mathbf{b}_*.$$

Estimation of the smoothing parameter

The solution $\widehat{\lambda}$ can be obtained via the following joint iterative process:

Step 1: Obtain $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described previously;

Step 2: (E-step) Given the observed data, evaluate the expectation of $\ell(\lambda; \mathbf{y}_{comp*})$ and estimate in the kth iteration :

$$Q(\lambda|\widehat{\lambda}^{(k)}) \quad = \quad \mathbb{E}\left[\ell(\lambda; \mathbf{y}_{comp*})|\mathbf{y}, \widehat{\lambda}^{(k)}\right] = -\frac{1}{2}\log|\boldsymbol{\Psi}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}^{-1}\widehat{\mathbf{b}_*\mathbf{b}_*^\top}^{(k)}\right),$$

with $\widehat{\mathbf{b}_*\mathbf{b}_*^\top}^{(k)} = \mathbb{E}\left[\mathbf{b}_*\mathbf{b}_*^\top|\mathbf{y}, \widehat{\lambda}^{(k)}\right]$

Step 3: (M-step) Uptade λ by

$$\widehat{\lambda}^{(k+1)} = -\frac{r-2}{\mathrm{tr}\left(\boldsymbol{\Psi}^{-1}\frac{\partial\boldsymbol{\Psi}}{\partial\lambda}\boldsymbol{\Psi}^{-1}\widehat{\mathbf{b}_*\mathbf{b}_*^\top}^{(k)}\right)}.$$

Thus, by repeating Step 1, Step 2 and Step 3, this iterative process leads to the MPL estimates of $\boldsymbol{\theta}$ and the smoothing parameter λ.

Contents

Simulation study

We simulated data from the model

$$y_{ij} = \beta_1 x_{1_{ij}} + \beta_2 x_{2_{ij}} + f(t_{ij}) + b_{0i} + b_{1i} t_{ij} + \epsilon_{ij},$$

with $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, $(b_{0i}, b_{1i}) \overset{\text{ind.}}{\sim} \mathrm{N}(\mathbf{0}, \mathbf{D})$, and $\epsilon_{ij} \overset{\text{ind.}}{\sim} \mathrm{N}_{n_i}(\mathbf{0}, \mathbf{\Omega}_i)$.

- The parameters were set at $\boldsymbol{\beta}^\top = (\beta_1, \beta_2) = (2, -1.5)$, $\sigma^2 = 0.55$, and $\mathbf{D}$ with elements $\alpha_{11} = 0.25$, $\alpha_{12} = 0.1$, and $\alpha_{22} = 0.2$.
- We chose a smoothing function $f(t_{ij}) = \cos(\pi\sqrt{t_{ij}})$, with $t_{ij} = (2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)$.
- For each sample size, we generated 500 samples of the DEC-SMEC model considering an AR(1) structure with parameter $\phi_1 = 0.6$.
- $x_1 \sim U(0, 1)$ and $x_2 \sim U(-1, 2)$, x_1 is independent of x_2.
- The censoring proportion was fixed at 15% and sample sizes at $n = 60, 100, 200$, and 400 were considered.

Evaluation of the parametric components

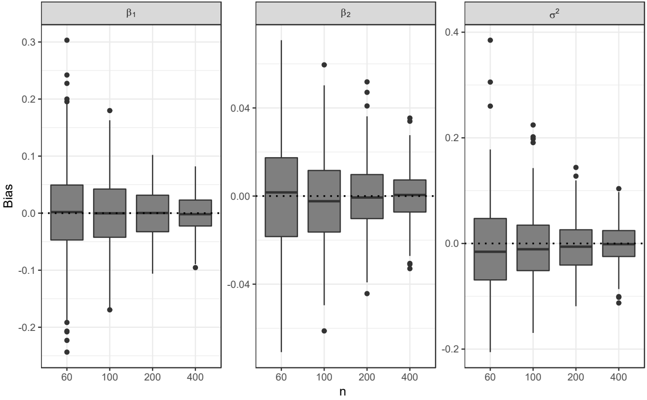


Figure: **Simulation study**. Box-plots of the biases of $\boldsymbol{\beta}$ and σ^2 estimates.

Evaluation of the parametric components

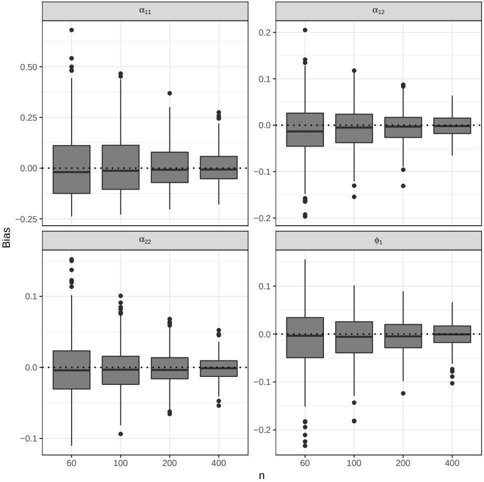


Figure: **Simulation study**. Box-plots of the biases of α and ϕ_1 estimates.

Evaluation of the parametric components

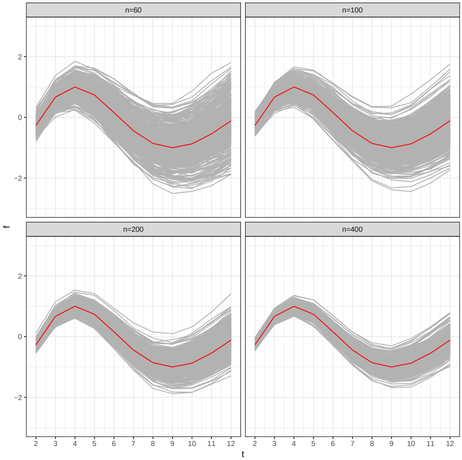


Figure: **Simulation study**. Graphs of the nonparametric components with 500 replications. Adjusted curves (gray lines) and true curves (red lines).

Contents

ACTG 315 study

We apply our proposed semiparametric linear mixed-effects model to the motivating ACTG 315 protocol HIV-1 RNA viral load dataset previously analyzed by Wu (2002).

We considered the following model:

$$y_{ij} = \mathrm{CD4}^{+}_{ij}\beta_1 + f(t_{ij}) + b_{0i} + b_{1i}t_{ij} + \epsilon_{ij}, \qquad (11)$$

where

- y_{ij} denotes the $\log_{10}$ transformation of the viral load for the ith subject at time t_{ij} $(i = 1, 2, \ldots, 46\ ;\ j = 1, 2, \ldots, n_i)$;
- $f(t_{ij})$ is an arbitrary smoothing function;
- b_{0i}, b_{1i} are the random intercept and random slope, respectively for the i-th patient;
- ϵ_{ij} are random errors.

ACTG 315 study

Table: **ACTG 315 study**. Parameter estimates of the SMEC model (11) for the ACTG 315 dataset. SE indicates the standard errors.

	UNC		DEC		AR(1)		CS	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	-0.0703	0.0241	-0.0583	0.0292	-0.0617	0.0298	-0.0704	0.0241
f_1	4.9380	0.0918	4.9293	0.0912	4.9235	0.0952	4.9474	0.0926
f_2	4.9535	0.0778	4.9754	0.0996	4.9764	0.0945	4.9334	0.0834
f_3	4.1325	0.0842	4.1298	0.0870	4.1293	0.0898	4.1401	0.0829
f_4	3.7863	0.0833	3.7759	0.0867	3.7742	0.0900	3.7825	0.0821
f_5	3.4181	0.0893	3.4100	0.0904	3.4079	0.0928	3.4181	0.0875
f_6	3.0364	0.1009	3.0304	0.1017	3.0315	0.1022	3.0352	0.1005
f_7	2.7905	0.1269	2.7803	0.1294	2.7831	0.1286	2.7893	0.1268
f_8	2.4340	0.1647	2.4339	0.1666	2.4210	0.1666	2.4323	0.1647
f_9	2.9769	0.3025	2.8663	0.3008	2.8999	0.3034	2.9731	0.3024
f_{10}	3.4407	0.5585	3.3810	0.5995	3.3510	0.6102	3.4380	0.5531
σ^2	0.1449		0.2851		0.1991		0.2855	
α_{11}	0.2435		0.0507		0.1747		0.1034	
α_{12}	-0.0006		0.0008		-0.00003		-0.0006	
α_{22}	0.0001		0.0001		0.0001		0.0001	
ϕ_1			0.9		0.89		0.4914	
ϕ_2			0.6501		1		0	
λ	88.2971		63.7242		42.1648		174.4071	
loglikp	-275.481		-230.7881		-239.2762		-276.1796	
AIC	580.406		**495.4174**		510.4558		585.8651	

ACTG 315 study

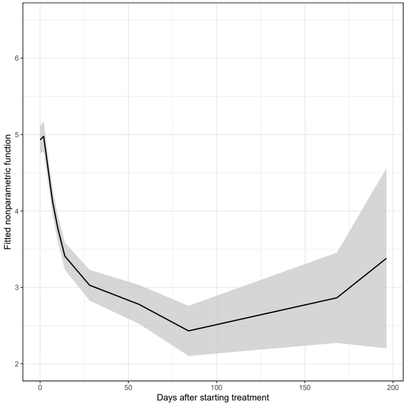
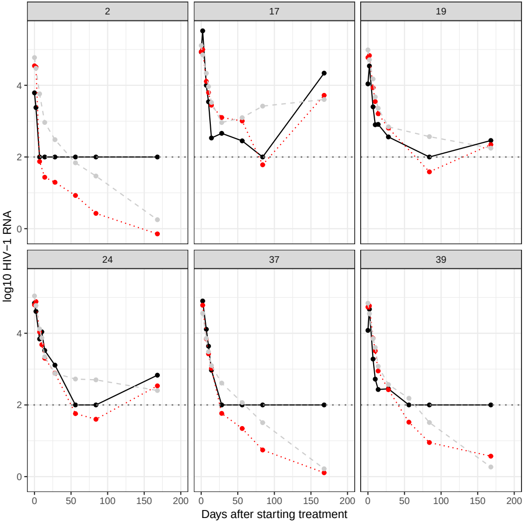


Figure: **ACTG 315 study**. Fitted curve of nonparametric part. The shaded regions denote the 95% confidence intervals obtained by $\hat{\mathbf{f}} \pm 1.96\sqrt{\hat{\mathrm{Var}}(\hat{\mathbf{f}})}$.

ACTG 315 study



A5055 study

Our purpose is to investigate the relationship between the viral load and the immunological markers in AIDS clinical trials. In order to avoid overly small estimates, which may be unstable, we standardized the covariates CD4+ and CD8+ cell counts. The predefined study day of viral load measurement (not the exact measured day) was used in our analysis.

We considered the following model:

$$y_{ij} = \mathrm{CD4}^{+}_{ij}\beta_1 + \mathrm{CD8}^{+}_{ij}\beta_2 + f(t_{ij}) + b_{0i} + b_{1i}t_{ij} + \epsilon_{ij}, \quad (12)$$

where

- y_{ij} denotes the $\log_{10}$ transformation of the viral load for the ith subject at time t_{ij} $(i = 1, 2, \ldots, 44\ ;\ j = 1, 2, \ldots, n_i)$;
- $t_{ij} = \mathrm{day}_{ij}/7$ (week);
- $f(t_{ij})$ is an arbitrary smoothing function;
- b_{0i}, b_{1i} are the random intercept and random slope, respectively for the i-th patient;
- ϵ_{ij} are random errors.

A5055 study

Table: **A5055 study**. Parameter estimates of the SMEC model for the A5055 dataset. SE indicates the standard errors.

	UNC		DEC		AR(1)		CS	
Parameter	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_1	-0.5315	0.0915	-0.5009	0.0917	-0.5261	0.0938	-0.5343	0.0004
β_2	0.1083	0.0715	0.1101	0.0674	0.1076	0.0696	0.1097	0.0002
f_1	3.5924	0.1107	3.6212	0.1319	3.6062	0.1332	3.5844	0.2257
f_2	3.0735	0.1184	3.0655	0.1207	3.0679	0.1231	3.0805	0.2214
f_3	2.6499	0.0883	2.6468	0.1137	2.6504	0.1163	2.6649	0.2247
f_4	2.2510	0.1129	2.2839	0.1207	2.2733	0.1224	2.2433	0.2292
f_5	1.7525	0.1319	1.7398	0.1424	1.7452	0.1427	1.7526	0.2430
f_6	1.6976	0.1548	1.6603	0.1674	1.6693	0.1662	1.6956	0.2571
f_7	1.8995	0.1902	1.8464	0.1974	1.8733	0.1949	1.9039	0.2871
f_8	2.3043	0.2135	2.2342	0.2285	2.2610	0.2249	2.2965	0.3188
f_9	2.0491	0.2859	1.9512	0.2901	1.9792	0.2840	2.0519	0.3551
σ^2	0.3914		0.7364		0.7639		0.5836	
α_{11}	0.4661		0.0190		0.0111		0.2756	
α_{12}	-0.0243		0.0005		-0.0013		-0.0243	
α_{22}	0.0054		0.0033		0.0033		0.0054	
ϕ_1			0.9		0.8628		0.3282	
ϕ_2			1.3498		1		0	
λ	21.6111		33.6207		36.6366		32.4563	
loglikp	-311.8647		-289.8267		-291.7714		-312.2359	
AIC	650.7973		**610.9712**		612.6926		655.2277	

A5055 study

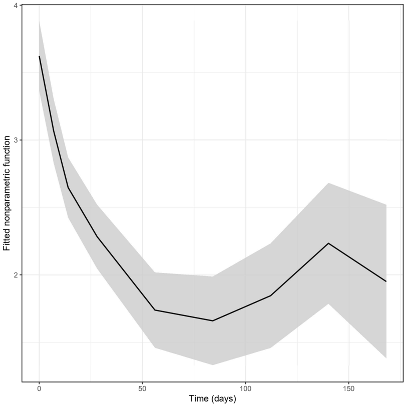


Figure: **A5055 study**. Fitted curve of nonparametric part. The shaded regions denote the 95% confidence intervals obtained by $\hat{\mathbf{f}} \pm 1.96\sqrt{\hat{\mathrm{Var}}(\hat{\mathbf{f}})}$.

A5055 study

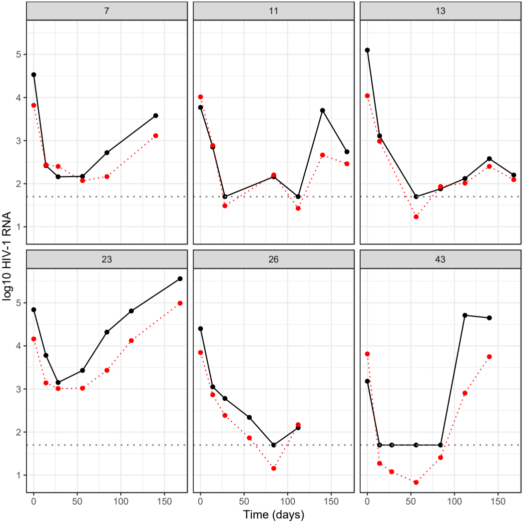


Figure: **A5055 study**. Viral loads in log estimated trajectories (red, dotted line) for the SMEC model in the DEC structure.

Contents

Conclusions

- This work provides a theoretical framework for a semiparametric mixed model for longitudinal censored data, which can be considered a generalization of the normal linear/nonlinear mixed-effects models for censored data proposed by Matos et al. (2016) and Vaida and Liu (2009).
- Simulation studies carried out suggest that the proposed method performs very well
- The approach was applied to analyze two HIV-AIDS studies, showing the advantage of the SMEC model to fit datasets with nonlinear subject-specific trajectories.
- It would thus also be interesting to consider a broader family of distributions such as the multivariate skew-normal distribution (Azzalini and Valle, 1996) and the multivariate skew-t distribution (Azzalini and Genton, 2008), which could be more realistic for the random effects and error terms.

Work on progress - Diagnostics analysis

Influence diagnostics are widely used in statistical modeling to identify and evaluate aberrant and influential points which may cause unwanted effects on estimation and goodness of fit.

This can be carried out by conducting local influence analyses, a general statistical technique used to assess the stability of the estimation outputs with respect to the model inputs, usually through the approach proposed in Cook (1986).

Additionally, Zhu and Lee (2001) proposed a method to assess the local influence in a minor perturbation of a statistical model with incomplete data.

Diagnostics analysis:

1. Case-deletion measures
2. Local Influence

References I

Acosta, E. P., H. Wu, S. M. Hammer, S. Yu, D. R. Kuritzkes, A. Walawander, J. J. Eron, C. J. Fichtenbaum, C. Pettinelli, D. Neath, et al. (2004). Comparison of two indinavir/ritonavir regimens in the treatment of hiv-infected individuals. *JAIDS Journal of Acquired Immune Deficiency Syndromes 37*(3), 1358–1366.

Azzalini, A. and M. Genton (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review 76*, 1490–1507.

Azzalini, A. and A. D. Valle (1996). The multivariate skew-normal distribution. *Biometrika 83*(4), 715–726.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *International Statistical Review/Revue Internationale de Statistique 55*, 245–259.

Hughes, J. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics 55*, 625–629.

Ibacache-Pulgar, G., G. A. Paula, and F. J. A. Cysneiros (2013, Mar). Semiparametric additive models under symmetric distributions. *TEST 22*(1), 103–121.

Kotzin, B. L., D. R. Kuritzkes, E. Connick, M. M. Lederman, A. D. Sevin, J. Spritzler, F. Rousseau, M. St. Clair, A. Martinez, J. D. Roe, L. Fox, J. M. Leonard, M. H. Chiozzi, A. Landay, and H. Kessler (2000, 01). Immune Reconstitution in the First Year of Potent Antiretroviral Therapy and Its Relationship to Virologic Response. *The Journal of Infectious Diseases 181*(1), 358–363.

Lachos, V. H., L. A. Matos, L. M. Castro, and M.-H. Chen (2019). Flexible longitudinal linear mixed models for multiple censored responses data. *Statistics in Medicine 38*(6), 1074–1102.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 44*(2), 226–233.

References II

Matos, L. A., L. M. Castro, and V. H. Lachos (2016, Dec). Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. *TEST 25*(4), 627–653.

Matos, L. A., M. O. Prates, M.-H. Chen, and V. H. Lachos (2013). Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. *Statistica Sinica 23*, 1323–1345.

Munoz, A., V. Carey, J. P. Schouten, M. Segal, and B. Rosner (1992). A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics 48*(3), 733–742.

Segal, M. R., P. Bacchetti, and N. P. Jewell (1994). Variances for maximum penalized likelihood estimates obtained via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 56*(2), 345–352.

Speed, T. (1991). [that blup is a good thing: The estimation of random effects]: Comment. *Statistical Science 6*(1), 42–44.

Vaida, F. and L. Liu (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics 18*, 797–817.

Vock, D. M., M. Davidian, A. A. Tsiatis, and A. J. Muir (2011). Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics 13*(1), 61–73.

Wang, Y. (1998). Mixed effects smoothing spline analysis of variance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(1), 159–174.

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association 97*(460), 955–964.

Zeger, S. L. and P. J. Diggle (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics 50*, 689–699.

Zhang, D., X. Lin, J. Raz, and M. Sowers (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association 93*(442), 710–719.

Thank you!

www.ime.unicamp.br/~larissam

larissam@unicamp.br