# Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: Application to HIV dynamics model

## Computational Statistics

Simon Queric and Vincent Herfeld

ENS Paris-Saclay, Télécom Paris

January 11, 2024

école normale supérieure paris—saclay

TELECOM Paris

# Table of content

# Introduction

- Understand HIV evolution : measure of viral load
- Sensor with a lower limit of quantification LOQ, data below LOQ are censored to LOQ
- Modelling decrease of HIV viral load with a longitudinal model : nonlinear mixed effects model

$$\begin{cases} y_{i,j} & = f(\phi_i, t_{i,j}) + \varepsilon_{i,j} \\ \varepsilon_{i,j} & \sim \mathcal{N}(0, \sigma^2 I_n) \\ \phi_i & = \mu + b_i, b_i \sim \mathcal{N}(0, \Omega) \end{cases}$$

where $\mu = (\log P_1, \log P_2, \log \lambda_1, \log \lambda_2)$.

- The parameters we want to find : $\theta = (\mu, \Omega, \sigma^2)$.
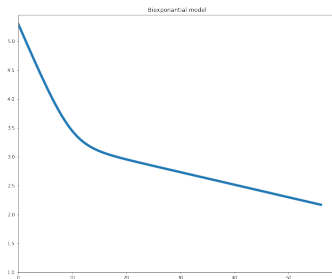
# Bi-exponential model



Figure: Function modelling decrease of log viral load

$$f(P_1, P_2, \lambda_1, \lambda_2, t) = \log_{10}\left(P_1 e^{-\lambda_1 t} + P_2 e^{-\lambda_2 t}\right)$$

introduced by [2]

# Previous method to find $\theta$

- ▶ Omit censored value or set value to LOQ or LOQ/2.
- ▶ Estimate $\theta$ with classical SAEM algorithm.
- ▶ Drawback : estimation has a strong bias
- ▶ Method proposed : Simulate $y^{cens}$
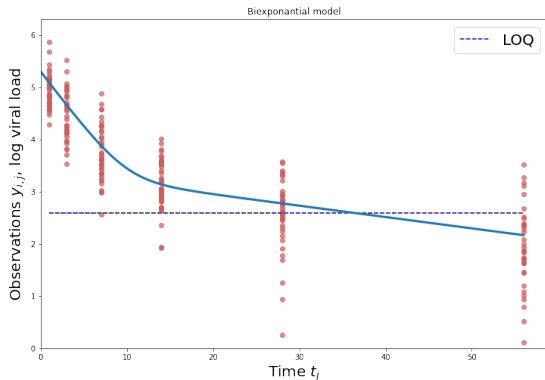
# Simulated data



Figure: Simulated data for $N = 40$ individuals, measurements on days $T = (1, 3, 7, 14, 28, 56)$

# Log-likelihood of the complete data

The log-likelihood of the data writes, up to an additive constant :

$$\log p_\theta(y, \phi) = \log p_\theta(y \mid \phi) + \log p_\theta(\phi)$$

$$= -NK \log \sigma + \sum_{i=1}^{N} \sum_{j=1}^{K} -\frac{(y_{i,j} - f(\phi_i, t_{i,j}))^2}{2\sigma^2}$$

$$+ \sum_{i=1}^{N} -\frac{(\phi_i - \mu)^T \Omega^{-1} (\phi_i - \mu)}{2} - N \log \det \Omega / 2$$

The missing data are $z = (\phi, y^{cens})$ and the parameters are
$\theta = (\log P_1, \log P_2, \log \lambda_1, \log \lambda_2, \Omega, \sigma^2) = (\mu, \Omega, \sigma^2)$.

where $\Omega = \begin{pmatrix} \omega_1^2 & 0 & 0 & 0 \\ 0 & \omega_2^2 & 0 & 0 \\ 0 & 0 & \omega_3^2 & 0 \\ 0 & 0 & 0 & \omega_4^2 \end{pmatrix}$

It's a curved exponential model :

$$\log p_\theta(y, \phi) = -\Lambda(\theta) + \langle S(y, z), \varphi(\theta) \rangle$$

# Table of content

# Method : SAEM algorithm

- ▶ EM algorithm with a stochastic approximation to generate missing data
- ▶ Iterative algorithm
- ▶ SAE step :
  1. Simulation de $z^{(m)} = (\phi^{(m)}, y_{cens}^{(m)})$
  2. $s_m = s_m + \gamma_m(S(y, z^{(m)}) - s_{m-1})$
- ▶ M step : $\widehat{\theta}_m = \underset{\theta \in \Theta}{\mathrm{argmax}} \left( -\Lambda(\theta) + \langle s_m, \varphi(\theta) \rangle \right)$

# Gibbs algorithm to simulate $z$

▶ At the $m$-th iteration :
  1. $\phi^{(m)} \sim p(\cdot \mid y_{obs}, y_{cens}^{(m-1)}, \widehat{\theta}_{m-1})$ with one iteration of Hasting-Metropolis (proposal : $\mathcal{N}(\phi^{(m-1)}, \lambda \widehat{\Omega}_{m-1})$
  2. $y_{cens}^{(m)} \sim p(\cdot \mid y_{obs}, \phi^{(m)}, \widehat{\theta}_{m-1})$ with a truncated gaussian distribution at the right of LOQ with a rejection method proposed by [3].
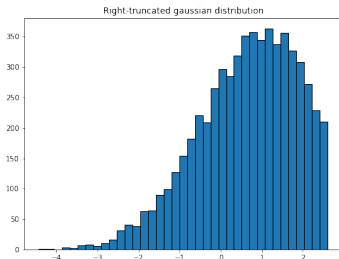


Figure: Histogram of right-truncated gaussian distribution with rejection sampling. It's the distribution of $y_{cens} = \mathbb{E}[y \mid y < LOQ]$

# Table of content

# Convergence criterion

(SAEM 1) For any $\theta \in \Theta$, the Gibbs algorithm generates a uniformly ergodic chain which invariant probability is $p(z|y; \theta)$.

(SAEM 2) For all $m \in \mathbb{N}^*$, $\gamma_m \in [0, 1]$, and verifies : $\sum_{m=1}^{\infty} \gamma_m = \infty$, $\sum_{m=1}^{\infty} \gamma_m^2 < \infty$

# Table of content

# Results on synthetic data

Empirical convergence of the SAEM algorithm
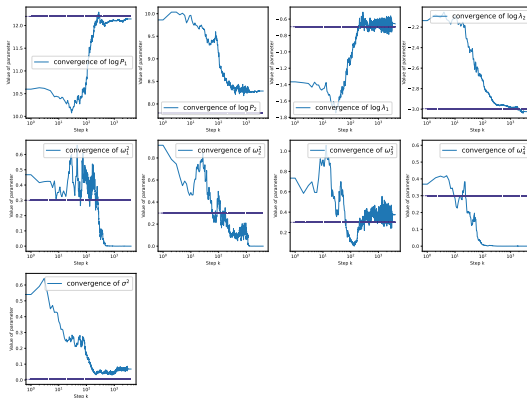


Figure: convergence of the parameters

# Results on synthetic data



Figure: Estimated censored values after SAEM
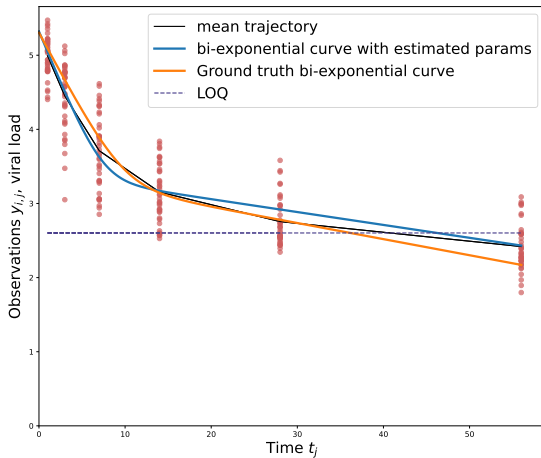
# Relative biases of estimates

| $\log P_1$ | $\log P_2$ | $\log \lambda_1$ | $\log \lambda_1$ | $\omega_1^2$ | $\omega_2^2$ | $\omega_3^2$ | $\omega_4^2$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| 1.5 | 2.7 | 21.7 | 6.4 | 100 | 45 | 33 | 100 | 740 |

Table: Relative biases of our estimates (in %)

| $\log P_1$ | $\log P_2$ | $\log \lambda_1$ | $\log \lambda_1$ | $\omega_1^2$ | $\omega_2^2$ | $\omega_3^2$ | $\omega_4^2$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|
| 0.03 | 0.23 | 0.57 | 0.62 | 4.26 | 6.21 | 1.67 | 6.59 | 0.63 |

Table: Relative biases obtained by A. Samson, M. Lavielle and F. Mentré (in %)

We have similar results to one of the previous method which consists to set the censored values to LOQ/2.

# Table of content

# Discussion

► Strengths
  – EM increases the likelihood at each step
  – SAEM allows to solve the E step when there is no closed-form for Q
  – Take into account variabilty of treatment response for each patient

► Weaknesses
  – No guarentee to be close to the MLE (no claim in this article)
  – Sensitive to initial parameters
  – Requires careful implementation

## Thanks to Adeline Samson who kindly answered some of our questions.

► Question : *[...] Toujours page 6, la mise à jour de $\omega_m^2$ nous pose problème. En effectuant les calculs, nous trouvons la variance empirique i.e $s^{(2)}/N - (s^{(1)})^2/N^2$.*

► Answer : *En effet, c'est une erreur. Comme pour la question précédente, si Omega est une matrice à estimer, ce serait $s^{(2)}/N - 1/N^2 s^{(1)} (s^{(1)})^T$*

# Table of content

# References

1. Samson, Adeline and Lavielle, Marc and Mentré, Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: Application to HIV dynamics model
2. Ding, A A, and H Wu. "Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models." Biostatistics (Oxford, England) vol. 2,1 (2001): 13-29. doi:10.1093/biostatistics/2.1.13
3. C. Robert, Simulation of truncated normal variables, Stat. Comput. 5(1995) 121–125