The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model.

Adeline Samson, Marc Lavielle, France Mentré

The SAEM algorithm for group comparison tests

in longitudinal data analysis based on non-linear

mixed-effects model

Adeline Samson[1], Marc Lavielle[2], France Mentré[1]

[1] INSERM U738, Paris, France; University Paris 7, Paris, France

[2] Paris-Sud University, Laboratoire de Mathématiques, Orsay, France

Abstract

Non-linear mixed-effects models (NLMEMs) are used to improve information gathering from longitudinal studies and applied to treatment evaluation in disease evolution studies, such as HIV infection. The estimation of parameters and the statistical tests are critical issues in NLMEMs since the likelihood and the Fisher information matrix have no closed form. An alternative method to numerical integrations, in which convergence is slow, and to methods based on linearisation, in which asymptotic convergence has not been proved, is the Stochastic Approximation Expectation-Maximization (SAEM) algorithm. For the Wald test and the likelihood ratio test, we propose estimating the Fisher information matrix by stochastic approximation and the likelihood by importance sampling. We evaluate these SAEM-based tests in a simulation study in the context of HIV viral load decrease after initiation of an antiretroviral treatment. The results from this simulation illustrate the theoretical convergence properties of SAEM. We also propose a method based on the SAEM algorithm to compute the minimum sample size required to perform a Wald test of a given power for a covariate effect in NLMEMs. Lastly we illustrate these tests on the evaluation of the effect of ritonavir on the indinavir pharmacokinetics in HIV patients and compare the results with those obtained using the adaptative Gaussian quadrature method implemented in the SAS procedure NLMIXED.

KEYWORDS: Likelihood ratio test; Wald test; longitudinal data; Non-linear mixed effects models; SAEM algorithm; sample size.

1 Introduction

Most clinical trials aim at comparing the efficacy of two different treatments or studying the effect of co-medication or physiological covariates. To assess whether the effect of such covariates implies a better reduction of the disease than without the covariates, several biological endpoints are repeatedly measured along the trial extent. The statistical approaches commonly used to study the influence of the covariate are classically based on the final measurements of this longitudinal data. Alternative methods to improve information extraction from longitudinal studies are analyses based on linear or non-linear mixed-effects models (NLMEMs). Such models have been developed for disease evolution studies, to determine the efficacy of anti-viral treatments in human immunodeficiency virus (HIV) [1, 2, 3, 4] or hepatitis B virus [5] infections evaluated through measures of viral load evolution, or prostate cancer treatment assessed by prostate-specific antigen dosage [6]. NLMEMs are also used to model the evolution of functional markers, for instance, for the decay of functional capacity in patients with rheumatoid arthritis [7], or the evolution of the ventilation function in patients with asthma [8]. NLMEMs are also powerful tools to analyze the pharmacokinetic properties of a drug. They allow for decreasing the number of samples per subject, which is an important advantage for interaction studies of protease inhibitors in HIV infected patients, for example [9].

Analysis of the covariate effect based on longitudinal data is thus essential. The properties of the statistical tests used to perform this analysis are based on the maximum likelihood (ML) theory. However, because of the non-linearity of the

1

regression function in the random effects, the likelihood of NLMEMs cannot be expressed in a closed form and the estimation of parameters by ML theory is complex. This situation leads to the development of widely used estimation methods based on likelihood linearization. These algorithms realize a first-order linearization of the regression function, as in the First Order and First Order Conditional Estimate (FOCE) algorithms [10, 11] implemented in the NONMEM software and in the nlme function of Splus and R software [12]. However these approximate methods cannot be considered as fully established theoretically. Furthermore, Vonesh gives an example of a specific design resulting in inconsistent estimates obtained with linearization methods, such as when the number of observations per subject does not increase faster than the number of subjects [13]. Particularly, convergence assumptions, on which the statistical tests are based, are not fulfilled. For instance, several authors show an inflation of the type I error by simulation of the most widely used group comparison tests, the Wald test and the Likelihood Ratio Test (LRT) [14, 15, 16, 17]. Thus, methods with proven convergence and consistency for finding the maximum likelihood estimate in NLMEMs are required.

Several estimation methods of conventional ML theory have been proposed as alternatives to linearization algorithms. A common method to handle numerical integrations is the adaptative Gaussian quadrature (AGQ) method. Estimation algorithms of parameters in a generalized mixed model and in a NLMEM based on this classical AGQ method have been proposed by Pinheiro and Bates [18] and are implemented in the SAS procedures GLIMMIX and NLMIXED, respectively [19]. However, the AGQ method requires a sufficiently large number of quadrature points implying an often slow convergence, which is not very stable. Improvement upon this method is thus needed. A second common method to handle numerical integrations is importance sampling, which is a stochastic integration method. However, as em-

phasized by Ge el al. [20], to achieve satisfactory numerical stability, this method can be computationally intensive, and hence numerically less efficient than many other parametric methods. The tool most commonly used to estimate models with missing or non-observed data such as random effects is the Expectation-Maximization (EM) algorithm [21]. The widespread popularity of the EM is largely due to its monotonicity: the likelihood increases at every step. Furthermore, the convergence of the EM algorithm has been widely studied [21]. Because of the non-linearity of the model, stochastic versions of the EM algorithm are proposed. Wei et al. [22]; Walker [23] and Wu [24, 25] propose MCEM algorithms, with a Monte-Carlo approximation of the expectation of the sufficient statistics in the E-step. This Monte-Carlo implementation is based on independently distributed samples with the posterior density of the parameters conditional on the observations. However the MCEM algorithm may have computational problems, such as slow or even no convergence and the large sample simulations realised by Monte Carlo Markov Chain (MCMC) procedure at each iteration are time consuming. Furthermore, the replication choice of the Monte Carlo sample is a central issue to guarantee convergence, and this problem remains unsolved. As an alternative to address both the pointwise convergence and the computational problem, we propose stochastic approximation versions of EM (SAEM) [26, 27]. This algorithm requires the simulation of only one realization of the missing data at each iteration, which substantially reduces computation time. In addition, pointwise almost sure convergence of the estimate sequence to a local maximum of the likelihood has been proved under general conditions [26]. Kuhn and Lavielle [28] propose to combine the SAEM algorithm with a MCMC procedure adapted to NLMEMs.

The first objective of this paper is to propose ML statistical tests for NLMEMs based on this SAEM algorithm. The Wald test statistic requires the computation

3

of the standard errors (SEs) of the estimated parameters. The inverse of the Fisher information matrix provides an upper bound of the estimated variance matrix but has no closed form because of the incomplete data structure. Therefore, we propose an estimate of this Fisher information matrix on the basis of Louis' principle [29] and the stochastic approximation procedure implemented in SAEM. To estimate the likelihood required for the LRT, we propose to use an importance sampling procedure. We then implement these methods and evaluate them in a simulation study of HIV infection dynamics. We simulate datasets from the bi-exponential model for HIV dynamics proposed by Ding and Wu [14], and evaluate the statistical properties of the SAEM parameter estimates, the standard error and the likelihood estimates. We also evaluate the type I error and the power of the tests for a comparison of a treatment effect on one parameter. Methods for minimum required sample size determination are needed for group comparison tests based on NLMEMs. Kang et al. [30] propose a method to compute sample sizes given a test hypothesis, based on a first-order linearization of the NLMEM. The second objective of this paper is to propose an alternative to this linearization-based approach, by using a SAEM-based approach. The sample size computation method is illustrated on the same HIV dynamics example.

After describing the model and notations (section 2), section 3 describes the SAEM algorithm and the statistical tests. Section 4 reports the simulation study and its results. Section 5 illustrates these tests on the evaluation of the effect of ritonavir on the indinavir pharmacokinetics, in patients with HIV infection. We compare the results with those obtained using the adaptative Gaussian quadrature method implemented in the SAS procedure NLMIXED. Section 6 concludes the article with some discussion.

2   Models and notations

Let us define $y_i = (y_{i1}, \ldots, y_{in_i})$ where y_{ij} is the response value for individual i at time t_{ij}, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, and let us define $y = (y_1, \ldots, y_N)$. We define an NLMEM as follows:

$$
\begin{aligned}
y_{ij} &= f(\phi_i, t_{ij}) + g(\phi_i, t_{ij})\, \varepsilon_{ij}, \\
\varepsilon_i &\sim \mathcal{N}(0, \sigma^2 I_{n_i}), \\
\phi_i &= X_i \mu + b_i, \text{with } \ b_i \sim \mathcal{N}(0, \Omega),
\end{aligned}
\tag{1}
$$

where $f(\cdot)$ and/or $g(\cdot)$ are non-linear functions of ϕ, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})$ represents the residual error, ϕ_i is a p-vector of individual regression parameters, μ is the $k \times p$-matrix of fixed effects, X_i is the k-vector of known covariates, b_i is a p-vector of random effects independent of ε_i, σ^2 is the residual variance, I_{n_i} the identity matrix of size n_i and Ω quantifies the variance matrix of the random effects. The maximum likelihood estimation in NLMEM is based on the log-likelihood function $L(y\,;\,\theta)$ of the response y, with $\theta = (\mu, \Omega, \sigma^2) \in \Theta$ the vector of all the parameters of the model. This function is equal to:

$$
L(y\,;\,\theta) = \sum_{i=1}^{N} L(y_i\,;\,\theta) = \sum_{i=1}^{N} \log\left(\int p(y_i, \phi_i; \theta)\, d\phi_i \right),
\tag{2}
$$

where $p(y_i, \phi_i;\,\theta)$ is the likelihood of the complete data (y_i, ϕ_i) of the i-th subject and is equal to $p(y_i, \phi_i; \theta) = \prod_{j=1}^{n_i} p(y_{ij}|\phi_i;\theta)p(\phi_i;\theta)$. As the random effects ϕ_i are unobservable and the regression functions are non-linear, the integral (2) has no closed form.

5

3   Estimation algorithm and group comparison tests

3.1   The SAEM algorithm

The EM algorithm is a classical approach to estimate parameters of models with non-observed or incomplete data [21]. For NLMEMs, the non-observed vector is the individual parameter vector $\phi = (\phi_1, \ldots, \phi_N)$ and the complete data of the model is (y, ϕ). Let us define the function $Q(\theta|\theta') = E(L_c(y, \phi; \theta)|y; \theta')$, where $L_c(y, \phi; \theta)$ is the log-likelihood of the complete data. At the m-th iteration of the EM algorithm, the E step is the evaluation of $Q_m(\theta) = Q(\theta \,|\, \widehat{\theta}_m)$, whereas the M step updates $\widehat{\theta}_m$ by maximizing $Q_m(\theta)$. For cases in which the E step has no analytic form, Delyon et al. [26] introduce a stochastic version of the EM algorithm that evaluates the integral $Q_m(\theta)$ by a stochastic approximation procedure. The authors prove the convergence of this SAEM algorithm under general conditions if $L_c(y, \phi; \theta)$ belongs to a regular curved exponential family:

$$L_c(y, \phi; \theta) = -\Lambda(\theta) + \langle S(y, \phi), \Phi(\theta) \rangle,$$

where $\langle .,. \rangle$ is the scalar product and $S(y, \phi)$ is the minimal sufficient statistic of the model. The E step is then divided into a simulation step (S step) of the non-observed data $\phi^{(m)}$ under the conditional distribution $p(\phi|y; \widehat{\theta}_m)$ and a stochastic approximation step (SA step) of $E\left(S(y, \phi)|\widehat{\theta}_m\right)$:

$$s_{m+1} = s_m + \gamma_m(S(y, \phi^{(m)}) - s_m), \tag{3}$$

6

using $(\gamma_m)_{m \geq 0}$ a sequence of positive numbers decreasing to 0. The M step is thus the update of the estimate $\widehat{\theta}_m$:

$$\widehat{\theta}_{m+1} = \arg \max_{\theta \in \Theta} \left(-\Lambda(\theta) + \langle s_{m+1}, \Phi(\theta) \rangle \right).$$

For NLMEMs, the SA step reduces to:

$$
\begin{aligned}
s_{1,i,m+1} &= s_{1,i,m} + \gamma_m \left(\phi_i^{(m)} - s_{1,i,m} \right), \quad i = 1, \ldots, N, \\
s_{2,m+1} &= s_{2,m} + \gamma_m \left(\sum_{i=1}^{N} \phi_i^{(m)t} \phi_i^{(m)} - s_{2,m} \right), \\
s_{3,m+1} &= s_{3,m} + \gamma_m \left(\sum_{i,j} \left(\frac{y_{ij} - f(\phi_i^{(m)}, t_{ij})}{g(\phi_i^{(m)}, t_{ij})} \right)^2 - s_{3,m} \right),
\end{aligned}
$$

and $\widehat{\theta}_{m+1}$ is obtained in the maximization step as follows:

$$
\begin{aligned}
\widehat{\mu}_{m+1} &= \left(\sum_{i=1}^{N} X_i^t X_i \right)^{-1} \sum_{i=1}^{N} X_i^t s_{1,i,m+1}, \\
\widehat{\Omega}_{m+1} &= \frac{1}{N} \left(s_{2,m+1} - \sum_{i=1}^{N} (X_i \widehat{\mu}_{m+1}) s_{1,i,m+1}^t - \sum_{i=1}^{N} s_{1,i,m+1} (X_i \widehat{\mu}_{m+1})^t + \sum_{i=1}^{N} (X_i \widehat{\mu}_{m+1})(X_i \widehat{\mu}_{m+1})^t \right) \\
\widehat{\sigma^2}_{m+1} &= \frac{s_{3,m+1}}{\sum_{i=1}^{N} n_i}.
\end{aligned}
$$

However, the simulation step can be complex when the posterior distribution $p(\phi|y; \theta)$ has no analytical form, such as for NLMEMs. Therefore an MCMC procedure such as the Metropolis-Hastings algorithm can be used to simulate $\phi^{(m)}$. At the m-th iteration of the SAEM algorithm, the S step is thus the simulation of $\phi^{(m)}$ with use of a Metropolis-Hastings algorithm which constructs a Markov Chain, with $p(\phi|y; \widehat{\theta}_m)$ as the unique stationary distribution (see [27] for more details). Kuhn and Lavielle [27] present the details of the SAEM implementation and prove that under general hypotheses, the sequence $(\widehat{\theta}_m)_{m \geq 0}$ obtained by this algorithm converges

almost surely towards a (local) maximum of the likelihood $L(y; \cdot)$.

3.2 Estimation of the Fisher Information matrix with stochastic approximation

The computation of the standard errors (SEs) of the estimated parameters is needed to perform the Wald test and to compute the required minimum sample size. These SEs can be evaluated as the diagonal elements of the inverse of the Fisher information matrix estimate, of which evaluation is complex because it has no analytic form. We adapt the estimation of the Fisher information matrix, proposed by Delyon et al. [26] and based on the Louis' missing information principle [29]. The Hessian of $L(y; \theta)$ can then be expressed as:

$$\partial_\theta^2 L(y; \theta) = E\left[(\partial_\theta^2 L_c(y, \phi; \theta)) + \mathrm{Var}\left(\partial_\theta L_c(y, \phi; \theta)\right)\right].$$

The Jacobian of $L(y; \theta)$ is the conditional expectation of the complete data likelihood:

$$\partial_\theta L(y; \theta) = E\left(\partial_\theta L_c(y, \phi; \theta)|y, \theta\right).$$

For NLMEMs, the derivatives $\partial_\theta L_c(y, \phi; \theta)$ and $\partial_\theta^2 L_c(y, \phi; \theta)$ have analytical forms. Therefore we implement their estimation using the stochastic approximation procedure of the SAEM algorithm. At the m-th iteration of the algorithm, we evaluate the 3 following quantities:

$$
\begin{aligned}
\Delta_{m+1} &= \Delta_m + \gamma_m \left(\partial_\theta L_c(y, \phi^{(m+1)}; \widehat{\theta}_{m+1}) - \Delta_m\right), \\
G_{m+1} &= G_m + \gamma_m \left(\partial_\theta^2 L_c(y, \phi^{(m+1)}; \widehat{\theta}_{m+1}) + \partial_\theta L_c(y, \phi^{(m+1)}; \widehat{\theta}_{m+1})\partial_\theta L_c(y, \phi^{(m+1)}; \widehat{\theta}_{m+1})^t - G_m\right), \\
H_{m+1} &= G_{m+1} - \Delta_{m+1}\Delta_{m+1}^t.
\end{aligned}
$$

As the sequence $(\widehat{\theta}_m)_{m\geq 0}$ converges to the maximum of the likelihood, the sequence $(H_m)_{m\geq 0}$ converges to the Fisher information matrix.

3.3   Estimation of the likelihood with importance sampling

The computation of the likelihood is needed to perform the LRT. The likelihood can be computed by adaptative Gaussian quadrature (AGQ) method. However, the convergence of the AGQ method can be slow and not very stable, especially when the number p of random effects is large. Alternative methods are stochastic integrations, such as Monte Carlo or importance sampling methods. Kuhn and Lavielle [27] propose a simple Monte-Carlo procedure to estimate $L(y;\theta)$, the estimate of the log likelihood of the i-th subject is as follows:

$$L(y_i;\theta) = \log\left(\frac{1}{T}\sum_{t=1}^{T} p(y_i|\phi_i^{(t)};\theta)\right),$$

with $\phi_i^{(t)} \sim_{iid} \mathcal{N}(\mu,\Omega)$, for $t = 1,\ldots,T$. By the strong law of large numbers, this estimate $L(y;\theta)$ converges almost surely towards $E[L(y;\theta)]$. However, this Monte-Carlo estimate is susceptible to numerical instabilities and to computational precision issues [31]. To avoid these numerical problems, we propose to estimate the likelihood using an importance sampling procedure. The importance sampling estimates of the log likelihood of the i-th subject is as follows:

$$L(y_i;\theta) = \log\left(\frac{1}{T}\sum_{t=1}^{T} \frac{p(y_i|\phi_i^{(t)};\theta)p(\phi_i^{(t)};\theta)}{h_i(\phi_i^{(t)};\theta)}\right)$$

with $h_i(.;\theta)$ any instrumental distribution and $\phi_i^{(t)} \sim_{iid} h_i(.;\theta)$, for $t = 1,\ldots,T$. This estimate converges for the same reason as the regular Monte-Carlo estimate, whatever the choice of the distribution h_i. However, some choices of h_i are obviously

9

better than others, especially to reduce the variance of the estimate. Among the distributions h_i, it is possible to exhibit the optimal distribution that minimizes the variance of the estimate [32]. For NLMEMs, this distribution is the individual posterior distribution $p(\phi|y_i;\theta)$. However, because this distribution has no closed form in NLMEMs, we propose for h_i a Gaussian approximation of the i-th individual posterior distribution (i.e. $\phi_i^{(t)} \sim \mathcal{N}(\mu_i^{post}, \Omega_i^{post})$). For each i, the posterior individual mean μ_i^{post} and the posterior individual variance Ω_i^{post} are estimated by the empirical mean and empirical variance of the $\phi_i^{(m)}$ simulated by the MCMC procedure.

3.4 Statistical tests

Let us assume that a scalar covariate effect β is tested on the k-th fixed effect. Let us denote by G_i the value of the covariate for subject i. The vector X_i is $X_i = (1, G_i)$ and the fixed effect matrix is:

$$\mu = \begin{pmatrix} \mu_1 & \cdots & \mu_{k-1} & \mu_k & \mu_{k+1} & \cdots & \mu_p \\ 0 & \ldots & 0 & \beta & 0 & \ldots & 0 \end{pmatrix}$$

The null hypothesis to test is H_0: $\{\beta = \beta_0\}$, with the alternative hypothesis H_1: $\{\beta \neq \beta_0\}$. Both the Wald test and the LRT can be performed to assess the covariate effect. For the Wald test, the parameter $\hat{\beta}$ and its variance $V(\beta) = SE^2(\beta)$ are estimated with the SAEM algorithm under H_1. The statistic $S_W = (\hat{\beta} - \beta_0)^2/V(\beta)$ follows a 1 degree of freedom χ_1^2 distribution under H_0. The rejection region of the Wald test for a nominal level α is therefore $\{S_W > \chi_{1;1-\alpha}^2\}$, where $\chi_{1;1-\alpha}^2$ is the critical value of the centered χ_1^2 distribution. For the LRT, the maximum likelihood estimates $\widehat{\theta_1}$ and $\widehat{\theta_0}$ of the models with and without the covariate effect respectively are computed with SAEM. The log-likelihoods $L_1 = L(y; \widehat{\theta_1})$ and $L_0 = L(y; \widehat{\theta_0})$ of the model with and without covariate effect respectively are estimated by importance

sampling. The statistic $S_{LRT} = 2(L_1 - L_0)$ is computed. Under H_0, S_{LRT} follows a χ_1^2 distribution. The rejection region of the LRT test for a nominal level α is therefore $\{S_{LRT} > \chi_{1;1-\alpha}^2\}$. These 2 tests can easily be extended to a vector β of covariate effects, the degrees of freedom for the χ^2 distribution being then the number of components of β.

3.5   Sample size computation

We propose a method to compute the power of a Wald test of a covariate effect based on NLMEMs. This computation requires proceeding through the following steps: specify the regression function and the NLMEM to be used; identify values for the parameter θ; specify an experimental design $(t_{ij})_{1 \le i \le N, 1 \le j \le n_i}$; identify the covariate effect to test, (i.e. the alternative hypothesis H_1); evaluate the standard errors SE_N, and finally, compute the power of the test. The minimum sample size required for a given power is then deduced from these last 2 steps, which are detailed below.

Let the tested parameter β be a scalar covariate effect on one fixed effect. For a clinical trial aiming at detecting a covariate effect of at least $(\beta_1 - \beta_0)$ on this fixed effect, the alternative hypothesis of the test is H_1: $\{\beta \ge \beta_1\}$. Under H_1, the statistic S_W is asymptotically distributed with a non-centered χ_1^2 distribution with a non-centrality parameter $(\beta - \beta_0)^2 / V(\beta)$. Therefore, the power of the Wald test is equal to:

$$p(\beta) = \int_{\chi_{1;1-\alpha}^2}^{\infty} \pi(x; 1, (\beta - \beta_0)^2 / V(\beta)) dx \qquad (4)$$

where $\pi(x; 1, c)$ is the probability density function of the non-centered χ_1^2 distribution, with a non-centrality parameter c. To compute the expected variance or standard error of β in an NLMEM, we propose to use the estimate of the Fisher

information matrix provided by the SAEM algorithm and detailed in section 3.2. When all patients have the same sampling design, a dataset with a covariate effect $\beta = \beta_1$ is generated with this sampling design and with a number N_{sim} of subjects large enough to ensure a fine approximation of the expected SE by the observed SE. The estimation of the Fisher information matrix is performed on this simulated dataset using the SAEM algorithm. Because the Fisher information matrix of the complete dataset is the sum of the individual Fisher information matrices, given the hypothesis of an identical sampling design for each subject, the SE for a dataset of N subjects, $SE_N(\beta)$, can be evaluated from the SE of the simulated dataset $SE_{N_{sim}}(\beta)$ using $SE_N(\beta) = SE_{N_{sim}}(\beta) \cdot \sqrt{N_{sim}/N}$. Therefore, for a given design and a given number of subjects N, the power of the Wald test can be evaluated from the equation (4). Finally, for a given power, the minimum sample size required is deduced from this power evaluation.

4   Simulation study

4.1   Simulation settings

The objective of this simulation study is to illustrate some statistical properties of the SAEM algorithm in the context of HIV viral dynamics. We evaluated the accuracy of the parameter estimates, the SE and the likelihood estimates. We performed the Wald test and the LRT in the context of group comparison tests, and lastly, we computed the power of the Wald test.

The bi-exponential model for initial HIV dynamics proposed by Ding and Wu [14] was used to simulate the datasets:

$$f(\phi_i, t_{ij}) = \log_{10}(P_{1i}e^{-\lambda_{1i}t_{ij}} + P_{2i}e^{-\lambda_{2i}t_{ij}}).$$

This function has p=4 individual parameters: P_{1i}, P_{2i} are the baseline values and λ_{1i}, λ_{2i} 2-phase viral decay rates. These parameters are positive and distributed according to a log-normal distribution. Thus, ϕ_i and μ take the following values: $\phi_i = (\ln P_{1i}, \ln P_{2i}, \ln \lambda_{1i}, \ln \lambda_{2i})$ and $\mu = (\ln P_1, \ln P_2, \ln \lambda_1, \ln \lambda_2)$. Identical sampling times are assumed for all subjects. Additive Gaussian random effects are assumed for each parameter with a diagonal covariance matrix Ω. Let $\omega^2 = (\omega_1^2, \omega_2^2, \omega_3^2, \omega_4^2)$ denote the vector of the variances of the random effects. Additive Gaussian error is assumed with a constant variance σ^2 (i.e. $g(\phi_i, t_j) = 1$ for all i, j). For the fixed effects, the values are those proposed by Ding and Wu [14]: $\ln P_1 = 12$, $\ln P_2 = 8$, $\ln \lambda_1 = \ln(0.5)$, $\ln \lambda_2 = \ln(0.05)$. The inter-subject variability is identical for the 4 parameters: $\omega_1^2 = \omega_2^2 = \omega_3^2 = \omega_4^2 = 0.3$ corresponding to a variation coefficient of 55%. The residual standard deviation is $\sigma = 0.065$, which corresponds to a variation coefficient of 15% for the viral load. With the Matlab software, we generated $N = 40$ or $N = 200$ total number of subjects with $n=6$ blood samples per patient, taken on days 1, 3, 7, 14, 28 and 56. A simulated dataset with $N = 40$ subjects is represented on Figure 1.
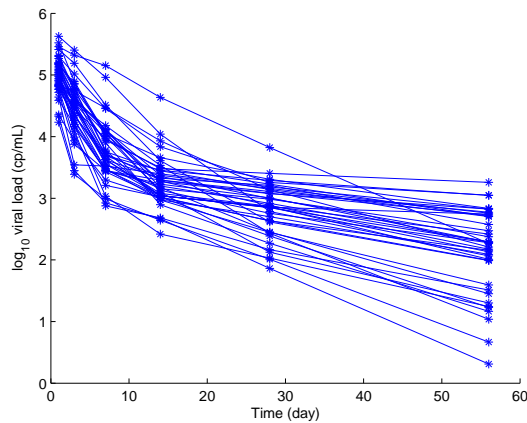

Figure 1: Simulated dataset with $N = 40$ subjects of the biexponential model describing the HIV viral load decrease under treatment.

13

4.2   Evaluation of estimates

Our aim is to evaluate the estimates produced by the SAEM algorithm. We fitted the datasets with the simulation model and computed the relative bias and relative root mean square error (RMSE) for each component of θ from 1000 replications of the trial described below for $N = 40$ and $N = 200$ subjects, respectively. The relative bias and RMSE on the 1000 data sets obtained for $N = 40$ and $N = 200$ subjects are presented in Table 1. For $N = 40$ subjects, the estimates have very

Table 1: Relative bias (%) and relative root mean square error (RMSE) (%) of the estimated parameters evaluated by the SAEM algorithm from 1000 simulated trials with $N = 40$ and $N = 200$ subjects.

Parameters	Bias (%)		RMSE (%)	
	$N = 40$	$N = 200$	$N = 40$	$N = 200$
$\ln P_1$	0.006	-0.003	0.78	0.35
$\ln P_2$	0.01	-0.003	1.23	0.55
$\ln \lambda_1$	0.48	-0.01	12.92	5.75
$\ln \lambda_2$	-0.04	0.01	3.03	1.36
ω_1^2	-2.45	-0.38	25.59	10.88
ω_2^2	-3.38	-1.21	29.02	12.19
ω_3^2	-1.75	-0.17	22.94	10.60
ω_4^2	-1.34	0.16	25.09	11.64
σ^2	0.10	0.15	15.82	6.91

low bias ($<1\%$ for the fixed effects, $<5\%$ for the variance parameters). The RMSE is satisfactory for the fixed effects ($<13\%$) as well as for the variance parameters ($<30\%$). As expected with $N = 200$ subjects, both the bias and the RMSE decrease with increasing subject number.

The SE estimated for each component of $\widehat{\theta}$ by the SAEM algorithm are compared with the "true" SE evaluated by the empirical standard deviation of the 1000 parameter estimates obtained for the simulated datasets. In Figure 2, for each component of θ, the 1000 estimated SEs and the true SEs with $N = 40$ datasets are plotted. For all parameters, the SEs estimated by SAEM are very close to the true
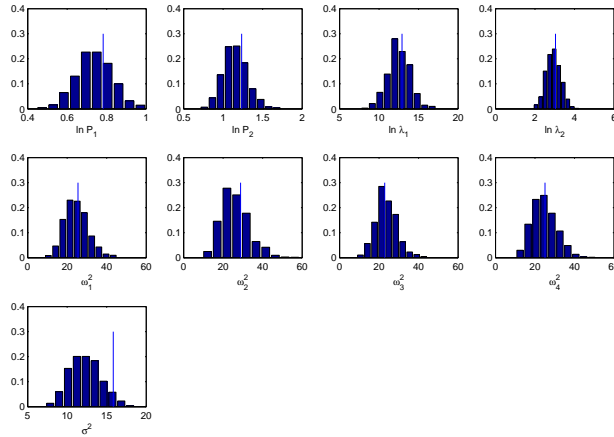
Figure 2: Histograms of the 1000 relative SEs (%) estimated by SAEM for datasets with N=40 subjects. The line represents an estimate of the "true" standard error estimated on 1000 replications.

SEs. Similar results are observed with $N = 200$ datasets.

The influence of the size T of the random samples used to evaluate the likelihood by importance sampling is studied for one dataset with $N = 200$ subjects. The log-likelihood is evaluated successively for different sample sizes $T = 1000$ or 5000 or 10 000 or 50 000, with 10 replications for each T, using the Gaussian approximation of the individual posterior distribution from the last 250 iterations of the SAEM algorithm. Results are reported in Figure 3 and show that the variability of the approximation is reduced by increasing the sample size. Therefore, the likelihood is evaluated by the importance sampling procedure with a sample size $T = 10\,000$, as a balance between estimate accuracy and time consumption.

4.3   Evaluation of statistical tests

We performed a Wald test and an LRT to test the difference between two treatment groups on the viral load decrease, especially on the first viral decay rate, $\ln\lambda_1$, as proposed by Ding and Wu [14]. We considered that the two groups are of equal size (i.e. 20 and 100 subjects per group when $N = 40$ and $N = 200$ subjects, respec-
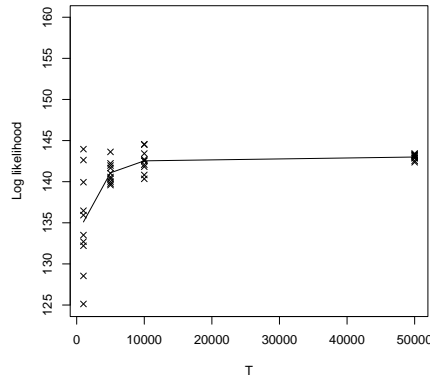
Figure 3: Log-likelihood estimates as a function of the sample size T used in the importance sampling procedure with 10 replications for each T, for one dataset with $N = 200$ subjects.

tively). The parameter vector θ is $\theta = (\mu, \omega^2, \sigma^2)$ under H_0 and $\theta = (\mu, \beta, \omega^2, \sigma^2)$ under H_1. We applied these tests using SAEM on the 1000 datasets simulated with $\beta = 0$ and evaluated the type I error of both the Wald test and the LRT by the proportion of trials for which H_0 is rejected, because these datasets are simulated without any treatment effect. We then evaluated the power of these tests for a treatment effect of almost 30% between the 2 treatment groups on the parameter $\ln \lambda_1$ (i.e. the alternative hypothesis H_1: $\{\beta \geq \beta_1\}$ with $\beta_1 = 0.262$). The power is estimated by the proportion of trials for which H_0 is rejected, within 1000 datasets simulated with a treatment effect $\beta_1 = 0.262$ on $\ln \lambda_1$.

The estimation of the type I error for a nominal value of 5% and the powers are given in Table 2 for datasets with 20 or 100 subjects per group. The estimated

Table 2: Evaluation on 1000 simulated datasets with 20 or 100 subjects per group of the type I error and the power of the Wald test and LRT for a treatment effect on the first decay rate.

	Type I error		Power	
Number of subjects per group	20	100	20	100
Wald test	4.0%	4.5%	37.5 %	90.4%
LRT	5.8%	5.6%	38.2 %	85.4%

type I errors are close to 5%, and given the dataset number of replications, do not differ significantly from the expected 5% value. The estimated powers are similar for the 2 tests and, as expected, are lower with $N = 40$ subjects than with $N = 200$ subjects.

4.4   Sample size computation example

The method proposed in section 3.5 to evaluate the power of the Wald test is applied to the model, the parameter values and the sampling design detailed above, for a difference of 30% between the 2 treatment groups in the parameter $\ln \lambda_1$ (i.e. $\beta_1 = 0.262$). A dataset is simulated with 2 groups of 5 000 subjects (i.e. $N = 10\,000$ subjects), and with a treatment effect β_1 on $\ln \lambda_1$. This dataset is analyzed using the SAEM algorithm to evaluate the Fisher information matrix. A $SE(\hat{\beta}) = 0.0112$ is obtained for 5 000 subjects per group. Applying equation (4), a sample size of 20 subjects per group ($N = 40$) provides a power of 32%, and a sample size of 100 subjects per group ($N = 200$) provides a power of 92%. These 2 values are close to the 2 estimated powers, obtained with the simulation study for the Wald test, of 37% and 90% respectively. Finally, the minimum sample size required to ensure a power of at least 80% is 70 subjects per group ($N = 140$). These results illustrate the ability of the SAEM approach to predict the SE of a fixed effect and the power of the Wald test to compute the minimum sample size required for a given power.

5 Effect of ritonavir on the indinavir pharmacokinetics in the COPHAR1-ANRS102 trial

5.1 Material and Methods

The COPHAR1-ANRS102 study was an open, multi-center, prospective trial including HIV-infected adults given an antiretroviral combination of at least 3 drugs, one being either indinavir or nelfinavir. Patients were required to have a baseline plasma HIV RNA below 200 copies/mL and to have maintained the same antiretroviral treatment for 6 months. Data on indinavir concentration were obtained for 45 patients who received different dosages of indinavir: 31 patients, indinavir alone three times a day (for most, 800 mg), and 14 patients indinavir twice a day (for most, 800 mg) with a booster dose of 100mg of ritonavir. From each patient, 5 blood samples were collected for the indinavir concentrations: a sample before indinavir administration and samples at 0.5, 1, 3 and 6 hours after indinavir administration. More details on this trial can be found in Goujard et al. [33]. The concentration data of the indinavir group were analyzed with an NLMEM and the FOCE algorithm implemented in the WinNonMix software by Brendel et al [34].

The aim of the present analysis is to evaluate the effect of the co-administration of ritonavir on the pharmacokinetic parameters of indinavir using the SAEM algorithm and the tests developed. The results are compared to those obtained with the adaptative Gaussian quadrature (AGQ) method implemented in the SAS procedure NLMIXED.

The pharmacokinetic statistical model proposed by Brendel et al [34] was used, which is a one-compartment model with first-order absorption and first-order elim-
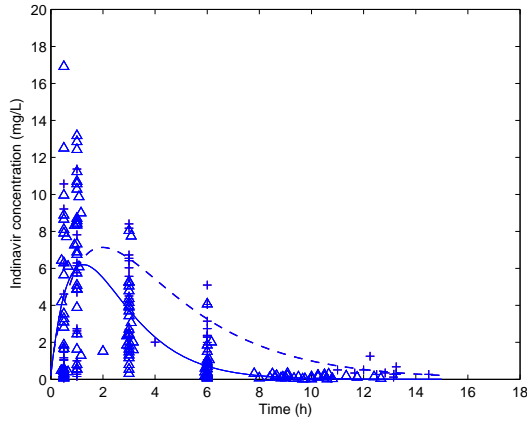
Figure 4: Observed individual indinavir concentrations: (+) and (Δ) for patients receiving ritonavir or not respectively; predicted mean curves obtained with SAEM: dotted line and plain line for patients receiving ritonavir or not respectively.

ination at steady-state:

$$f(\phi, t) = \frac{Dk_a}{Vk_a - Cl} \left(\frac{e^{-\frac{Cl}{V}t}}{(1 - e^{-\frac{Cl}{V}\tau})} - \frac{e^{-k_a t}}{(1 - e^{-k_a \tau})} \right)$$

where k_a is the first-order absorption rate constant, Cl the oral clearance, V the oral volume distribution and τ the delay between 2 drug administrations fixed to 12 and 8 h for patients receiving ritonavir or not, respectively. The individual parameters are $\phi = (\ln V, \ln k_a, \ln Cl)$. A diagonal variance matrix Ω and a homoscedastic error model were used. The Wald test was used to test the effect of ritonavir administration on the fixed effects k_a, Cl and V. The vector θ was estimated under H_1 with the SAEM algorithm and the procedure NLMIXED.

5.2   Results

Concentration data are displayed in Figure 4. The SAS procedure NLMIXED failed to estimate the variability on k_a which has to be fixed to 0. The SAEM algorithm succeeded in the estimation of all the parameters and estimated the k_a variability to 10^{-7}. Therefore, a model without a variability on k_a was considered with both

19

Table 3: Pharmacokinetic parameters of indinavir (estimate, SE (%) and p-value of the Wald test) estimated with the SAEM and the adaptive Gaussian quadrature (AGQ) algorithms

| | SAEM | | | AGQ | | |
Parameters	Estimate	SE (%)	p-value	Estimate	SE (%)	p-value
V (L)	46.70	32		43.80	22	
k_a (h−1)	0.76	22		0.78	14	
Cl (L/h)	37.80	7		42.14	7	
β_V	-0.59	89	0.154	-0.24	205	0.628
β_{k_a}	-0.90	32	0.002	-0.80	30	0.002
β_{Cl}	-0.66	21	<.001	-0.62	22	<.001
ω_V^2	1.22	55		1.22	40	
ω_{Cl}^2	0.05	50		0.07	51	
σ^2	2.40	12		2.77	12	

estimation methods. The parameters estimated by the SAEM and the AGQ methods for this model are presented in Table 3. To ensure convergence, 2000 iterations were used for the SAEM algorithm and 30 nodes were used with the SAS procedure NLMIXED. It takes about 90 s CPU time for the SAEM algorithm and 400 s CPU time for the procedure NLMIXED to run on a conventional Intel Pentium IV 2.8 GHz workstation.

A significant effect of co-medication with ritonavir was found with the 2 methods on ln k_a and ln Cl ($p < 0.01$), resulting in a decrease of 0.41/h and 0.44/h of k_a with the SAEM and the AGQ method, respectively, and a decrease of 0.51 L/h and 0.53 L/h of Cl with the SAEM and the AGQ method, respectively, when patients received ritonavir. The effect of ritonavir on V is not significant with both methods. The log-likelihood evaluated by the importance sampling procedure was equal to -469.6 at the estimates obtained with SAEM and equal to -472.1 at the estimates obtained with the AGQ method. The predicted concentrations of indinavir with and without ritonavir evaluated at the fixed effect values obtained with SAEM are overlayed on the plot of concentrations data of Figure 4. This graph illustrates the

slower decrease in indinavir concentration when it is co-administered with ritonavir.

6   Discussion

In this paper, we propose statistical tests for covariate effects in NLMEMs adapted to non-linear longitudinal data analysis. Because these tests take into account all the data, they are obviously more powerful than tests using only the final measurements. In the context of NLMEMs, several tests taking into account all the data have been proposed. However, those based on approximate methods have poor properties, especially with increase of the type I error. More recently, tests from numerical integration based methods, such as the adaptive Gaussian quadratures have also been proposed, even though they can be limited by slow convergence when the number of parameters is large. We propose here statistical tests based on the SAEM algorithm, which has good computational properties. The SAEM algorithm and the statistical tests proposed in this paper are implemented in a Matlab function called MONO-LIX and freely available on http://mahery.math.u-psud.fr/∼lavielle/monolix. The simulation study illustrates the accuracy of the SAEM algorithm to fit non-linear longitudinal data in the context of HIV viral load decrease, the parameter estimates being unbiased and with small RMSE. The SEs of the parameters are evaluated from the Fisher information matrix. We propose an estimation of this matrix with the stochastic approximation procedure of the SAEM algorithm and Louis' principle [29]. Results of the simulation study show that the SE estimates are very close to the "true" SEs evaluated on 1000 simulated datasets. Kuhn and Lavielle [27] propose to estimate the likelihood with a simple Monte Carlo procedure, but this method provides poor estimates and is prone to computational instabilities. To avoid this problem, we propose an importance sampling approach, with a Gaussian

approximation of the conditional posterior distribution used to sample the individual parameter. Hence, the Wald test and the LRT based on these Fisher information matrix and likelihood estimation procedures have accurate properties, especially the obtained type I errors are close to the expected threshold of 5%.

Another critical issue for NLMEMs is the computation of the minimum sample size required to observe a significant covariate effect on a fixed effect parameter with the Wald test. This issue requires evaluation of the expected SE for this covariate effect. Kang et al. [30] propose an analytic evaluation based on the linearization of the model, a method to be compared with the extension of the PFIM function for covariates proposed by Retout et al. [35]. However Kang et al. [30] show that this method underestimates the power of the test when the random effect variability increases. An alternative to this linearization method is to compute the expected Fisher information matrix with the SAEM estimate. A large dataset is simulated to be close to asymptotic results, from which the Fisher information matrix is estimated using SAEM. We show on the HIV example that the power predicted by this method is close to that evaluated on the simulation study. In this example, the same elementary design is used in every patient. This can easily be extended when the population design is composed of different groups of elementary designs [30].

Finally, the SAEM algorithm and the proposed tests are used to analyze the indinavir pharmacokinetics and to test the effect of the co-administration of ritonavir in HIV infected patients from the COPHAR 1-ANRS102 trial. As expected, a significant effect of ritonavir co-administration is found on the absorption and the elimination of indinavir [34]. Similar results were obtained with the two methods on this model. However models with more random effects may be difficult to analyse with the AGQ method.

22

The SAEM algorithm provides accurate estimates when working with NLMEMs and may be applied to even more difficult issues. For instance, when measuring a biological response such as a concentration or a viral load, the observations may be left-censored, due to the limits of quantification of the measuring equipment. We extended the SAEM algorithm to this case, the left-censored data being considered as non-observed data as well as the random effects [36].

Acknowledgments

References

[1] Wu H, Wu L. A multiple imputation method for missing covariates in non-linear mixed-effects models with application to HIV dynamics. *Statistics in Medicine* 2001; **20**:1755–1769

[2] Jacqmin-Gadda H, Thiébaut R, Chêne G, Commenges D. Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics* 2000; **1**:355–368

[3] Wu H, Wu L. Identification of significant host factors for HIV dynamics modeled by non-linear mixed-effects models. *Statistics in Medicine* 2002; **21**:753–771

[4] Wu H, Zhang JT. The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Statistics in Medicine* 2002; **21**:3655–3675

[5] Wolters L, Hansen B, Niesters H, de Man R. Viral dynamics in chronic hepatitis B patients treated with lamivudine, lamivudine-famciclovir or lamivudine-ganciclovir. *European Journal of Gastroenterology and Hepatology* 2002; **14**:1007–1011

[6] Law N, Taylor J, Sandler H. The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. *Biostatistics* 2002; **3**:547–563

[7] Welsing P, Van Gestel A, Swinkels H, Kiemeney L, Van Riel P. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. *Arthritis and Rheumatism* 2001; **44**:2009–2017

[8] Lange P, Parner J, Vestbo J, Schnohr P, Jensen G. A 15-year follow-up study of ventilatory function in adults with asthma. *New England Journal of Medicine* 1998; **17**:1194–200

[9] Barry M, Mulcahy F, Merry C, Gibbons S, Back D. Pharmacokinetics and potential interactions amongst antiretroviral agents used to treat patients with HIV infection. *Clinical Pharmacokinetics* 1999; **36**:289–304

[10] Beal S, Sheiner L. Estimating population kinetics. *Critical Reviews in Biomedical Engineering* 1982; **8**:195–222

[11] Lindstrom M, Bates D. Nonlinear mixed-effects models for repeated measures data. *Biometrics* 1990; **46**:673–687

[12] Pinheiro J, Bates D. *Mixed-effect models in S and Splus.* Springer-Verlag 2000

[13] Vonesh EF. A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* 1996; **83**:447–452

[14] Ding A, Wu H. Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics* 2001; **2**:13–29

[15] Wahlby U, Jonsson E, Karlsson M. Assessment of actual significance levels for covariate effects in NONMEM. *Journal of Pharmacokinetics and Pharmacodynamics* 2001; **28**:231–252

[16] Comets E, Mentré F. Evaluation of tests based on individual versus population modeling to compare dissolution curves. *Journal of Biopharmaceutical Statistics* 2001; **11**:107–123

[17] Panhard X, Mentré F. Evaluation by simulation of tests based on non-linear mixed-effects models in interaction and bioequivalence cross-over trials. *Statistics in Medicine* 2005; **24**:1509–1524

[18] Pinheiro J, Bates D. Approximations to the log-likelihood function in the non-linear mixed-effect models. *Journal of Computational and Graphical Statistics* 1995; **4**:12–35

[19] Wolfinger R. Laplace's approximations for non-linear mixed-effect models. *Biometrika* 1993; **80**:791–795

[20] Ge Z, Bickel P, Rice J. An approximate likelihood approach to nonlinear mixed effects models via spline approximation. *Computational Statistics and Data Analysis* 2004; **46**:747–776

[21] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B* 1977; **39**:1–38

[22] Wei G, Tanner MA. Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika* 1990; **77**:649–652

[23] Walker S. An EM algorithm for non-linear random effects models. *Biometrics* 1996; **52**:934–944

[24] Wu L. 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* 2002; **97**:955–964

[25] Wu L. Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. *Journal of the American Statistical Association* 2004; **99**:700–709

[26] Delyon B, Lavielle M, Moulines E. Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics* 1999; **27**:94–128

[27] Kuhn E, Lavielle M. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis* 2005; **49**:1020–1038

[28] Kuhn E, Lavielle M. Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM: Probability and Statistics* 2005; **8**:115–131

[29] Louis TA. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B* 1982; **44**:226–233

[30] Kang D, Schwartz J, Verotta D. A sample size computation method for

non-linear mixed effects models with applications to pharmacokinetics models. *Statistics in Medicine* 2004; **23**:2251–2566

[31] Robert C, Casella G. *Monte Carlo Statistical Methods*. Springer-Verlag 2002

[32] Geweke J. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 1989; **57**:1317–1339

[33] Goujard C, Legrand M, Panhard X, Diquet B, Duval X, Peytavin G, Vincent I, Katlama C, Leport C, Bonnet B, Salmon-Céron D, Mentré F, Taburet AM, Group COPHARANRSS. High variability of indinavir and nelfinavir pharmacokinetics in HIV-infected patients with a sustained virological response on highly active antiretroviral therapy. *Clinical Pharmacokinetics* 2005; **44**(12):1267–1278

[34] Brendel K, Legrand M, Taburet A, Baron G, Goujard C, Mentré F, Group CAT. Population pharmacokinetic analysis of indinavir in HIV-infected patient treated with a stable antiretroviral therapy. *Fundamental and Clinical Pharmacology* 2005; **19**:373–383

[35] Retout S, Mentré F. Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *Journal of Biopharmaceutical Statistics* 2003; **13**:209–227

[36] Samson A, Lavielle M, Mentré F. Extension of the saem algorithm to left-censored data in nonlinear mixed-effects model: application to hiv dynamics model. *Computational Statistics and Data Analysis* 2006; On line