HAL

archives-ouvertes.fr

saemix, an R version of the SAEM algorithm for parameter estimation in nonlinear mixed effect models

Audrey Lavenu, Emmanuelle Comets, Marc Lavielle

▶ To cite this version:

Audrey Lavenu, Emmanuelle Comets, Marc Lavielle. saemix, an R version of the SAEM algorithm for parameter estimation in nonlinear mixed effect models. 1ères Rencontres R, Jul 2012, Bordeaux, France. <hal-00717539>

HAL Id: hal-00717539

https://hal.archives-ouvertes.fr/hal-00717539

Submitted on 13 Jul 2012

saemix, an R version of the SAEM algorithm
for parameter estimation in nonlinear mixed effect models

A. Lavenu[a], E. Comets[b] and M. Lavielle[c]

[a]University Rennes-I & INSERM CIC 0203
Rennes,France
audrey.lavenu@univ-rennes1.fr

[b]INSERM, UMR 738 & Univ Paris Diderot, Sorbonne Paris Cité
Paris, France
emmanuelle.comets@inserm.fr

[c]INRIA
Saclay, France
Marc.Lavielle@inria.fr

Mots clefs : Nonlinear mixed effect models, parameter estimation, SAEM algorithm, R, R package, pharmacokinetics, pharmacodynamics, longitudinal data.

Introduction: The use of modelling and simulation in clinical drug development is now well established. Regardless of whether a single outcome is considered at the end of the study, clinical trials often collect longitudinal data, with each subject providing several measurements throughout the study. Longitudinal data is a staple in particular of pharmacokinetic (PK) and pharmacodynamic (PD) studies, which are a required part of a new drug application file. Nonlinear mixed effect models can help to characterise and to understand many complex nonlinear biological processes, such as biomarkers or surrogate endpoints, and are crucial in describing and quantifying the mechanisms of drug action and the different sources of variation, e.g., the interindividual variability. Over the past decade, new and powerful estimation algorithms have been proposed to estimate the parameters of these models. The Stochastic Approximation Expectation Maximization (SAEM) algorithm has proven very efficient, quickly converging to the maximum likelihood estimators [1] and performing better than linearisation-based algorithms [2]. It has been implemented in the Monolix software [3] which has enjoyed increasingly widespread use over the last few years, more recently in the Statistics toolbox of Matlab (nlmefitsa.m), and is also available in NONMEM version 7 [4]. The objective of the present package was to implement SAEM in the R software [5].

Methods: Detailed and complete presentations of the nonlinear mixed effects model can be found in several reference textbooks, for instance [6]. We consider the following general nonlinear mixed effects model for continuous outputs:

$$y_{ij} = f(x_{ij}, \psi_i) + g(x_{ij}, \psi_i, \xi)\varepsilon_{ij} \quad , \quad 1 \le i \le N \quad , \quad 1 \le j \le n_i \tag{1}$$

where y_{ij} is the jth observation of subject i, N is the number of subjects, n_i is the number of observations of subject i, x_{ij} are known regression variables, and ψ_i is the vector of individual parameters. The SAEM algorithm is used to obtain maximum likelihood estimates of the parameters of nonlinear mixed effects models without any linearisation of the model. The log-likelihood for nonlinear mixed effect models is analytically intractable since it requires integration over the unknown individual parameters. The SAEM algorithm uses an EM algorithm

[7], where the unknown individual parameters are treated as missing data, and replaces the usual E-step with a stochastic approximation step [8]. The missing parameters are simulated at each iteration via a MCMC procedure, which can be used after the algorithm has converged to obtain the conditional modes, the conditional means and the conditional standard deviations of the individual parameters.

Results: The library uses the S4 class system of R to provide a user-friendly input and output system, with methods like `summary` or `plot` for fitted objects. The package provides summaries of the results, individual parameter estimates, standard errors (obtained using a linearised computation of the Fisher information matrix) Wald tests for fixed effects, and a number of diagnostic plots, including VPC plots and npde [9]. The log-likelihood can be computed by three methods: a linearisation of the model, an importance sampling procedure, or a Gaussian quadrature. The diagnostic graphs can be tailored to the user's individual preferences by setting a number of options, and are easily exported to a file.

We illustrate the use of the library with the well known PK dataset of theophylline. These data includes the concentration versus time data collected in 12 subjects given a single oral dose of theophylline, and for whom 11 blood samples were collected over a period of 24 h. We modelled this data using a one-compartment model with first-order absorption, parameterised as k_a, V, CL. The IIV was modelled using an exponential model with diagonal variance-covariance matrix, while the residual variability was modelled with a combined error model. Many diagnostic plots are available to evaluate convergence or model adequacy, such as individual plots, using a `plot` function through which user-specific options can be set.

Conclusion: The saemix package provides the SAEM algorithm for R users. The current version handles models in analytical form, with continuous or binary covariates.

Références

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

[2] Girard P., Mentré F. (2005). A comparison of estimation methods in nonlinear mixed effects models using a blind analysis (oral presentation). *Meeting of the Population Approach Group in Europe (PAGE)*, Pamplona

[3] Lavielle M. (2010). *MONOLIX (MOdèles NOn LInéaires á effets miXtes) User Guide.* MONOLIX group, Orsay, France. URL: http://software.monolix.org/

[4] Beal S., Sheiner L.B., Boeckmann A., Bauer R.J. (2009). *NONMEM User's Guides. (1989-2009)*, Icon Development Solutions, Ellicott City, MD, USA

[5] R Development Core Team (2006). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria

[6] Davidian M., Giltinan D (1995). *Nonlinear models for repeated measurement data.* Chapman & Hall, London

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

[8] Kuhn E., Lavielle M (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis* **49**, 1020–38

[9] Brendel K., Comets E., Laffont C., Laveille C., Mentré M.. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide (2006). *Pharmaceutical Research*, **23**, 2036–49